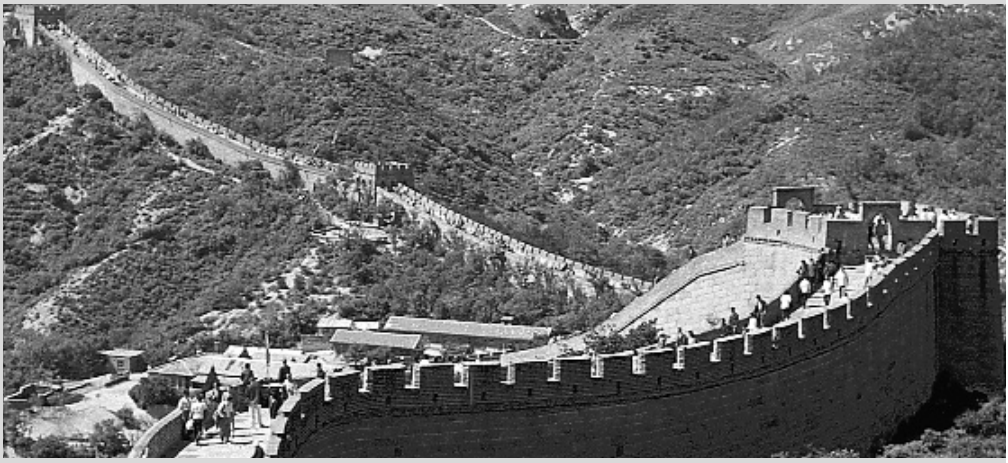
United Nations Development Programme Evaluation Office (UNDP/EO)
Chinese National Center for Science and Technology Evaluation (NCSTE)
The World Bank Operations Evaluation Department/World Bank Institute (OED/WBI)

EVALUATION CAPACITY DEVELOPMENT IN ASIA

EVALUATION CAPACITY DEVELOPMENT IN ASIA



Selected Proceedings from the International Conference Beijing, October 1999

Edited by Khalid Malik and Christine Roth

United Nations Development Programme Evaluation Office (UNDP/EO)
Chinese National Center for Science and Technology Evaluation (NCSTE)
The World Bank Operations Evaluation Department and World Bank Institute (WBI)

The views expressed in this report are those of the authors and do not necessarily represent those of the United Nations or the United Nations Developement Programme.

© Evaluation Office 2000

Evaluation Office
United Nations Developement Programme
One United Nations Plaza
New York, NY 10017, USA

Design: Julia Ptasznik, Colonial Communications Corp.
Cover Photograph: David Kinley/UNDP

CONTENTS

Foreword v

Acronyms vii

Executive Summary ix

Part I: The Evolving Context for ECD

1. **Opening Speech** 3
   Ms. Den Nan, Vice Minister of Science and Technology of PRC

2. **Opening Statement:**
   The New Dimensions of ECD in International Development 5
   Mr. Khalid Malik, Director, UNDP Evaluation Office

3. **The Asia Perspective in Evaluation Capacity Development:**
   The Challenges of the New Millennium 9
   Mr. Adil Khan

4. **Aid Evaluation:A Donor Perspective** 19
   Mr. Niels Dabelstein, Head of Evaluation Secretariat,
   DANIDA, OECD/DAC WP

Part II: Evaluation of Performance Management in the Public Sector

5. **Results Based Management in the Development**
   Cooperation Agencies: A review of Experience 23
   Ms. Annette Binnendijk, Consultant UNDP

6. **Evaluation Capacity Development in the**
   People's Republic of China: Trends and Prospects 39
   Mr. Ray Rist, Evaluation Advisor, World Bank

7. **Evolution of Evaluation in the People's Republic of China** 45
   Mr. Peter C. Darjes, Operations Manager, Asian Development Bank

8. **Linking Evaluation to Policy Formulation and**
   Budget Processes – Lessons Learned from Australia 49
   Mr. Leonard J. Early, Depatment of Finance – Australian Public Sector

9. **Public Performance Evaluation and Improvement** 55
   Mr. Marc Holzer, Executive Director,
   National Center for Public Productivity, NJ - USA

Part III:
A) Monitoring Evaluation in a Results-Based Management Approach
B) Evaluation Initiatives in Science and Technology

10.Evaluation Parternships **65**
Mr. Niels Dabelstein , Head of Evaluation Secretariat
DANIDA, OECD/DAC WP

11.Evaluation Tools and Methodologies:Case Studies **69**
Mr. Jacques Toulemonde, Director, Centre for European Evaluation Expertise

12.Evaluation Initiatives in Science and Technology Programs **77**
Mr. Rustam Lalkaka, President,
Business & Technology Development Strategies

13.Review of Draft of Science and Technology Evaluation Standard of China **91**
Ms. Chen Zhaoying, Vice President NCSTE

Part IV: ECD Opportunities for National Initiatives and International Cooperation

14. Guide to Diagnosing ECD Needs: Video Speech **97**
Mr. Keith Mackay, Senior Evaluation Officer, World Bank Institute

Part V: Meeting Wrap-Up

15.Meeting Wrap-Up **103**
Ms. Christine Roth, Evaluation Advisor, UNDP

Part VI: Annexes

Annex 1.List of Participants **107**

Annex 2.Conference Agenda **111**

FOREWORD

There is growing awareness in both developed and developing countries of the potential role that evaluation can play in improving the quality of public sector decision-making. Evaluation is becoming recognized as particularly important in promoting transparency and accountability; in some instances, it is even viewed as an integral part of responsible civic action and civil society development.

Concurrent with the increasing interest in evaluation as a component of public sector development and civil service reform, evaluation capacity development (ECD) is being placed on the agenda of cooperation between developing countries and development assistance agencies. The development of national evaluation capacities can be a critical component of efforts to strengthen the responsiveness of public management systems to both internal and external demands for transparency, effectiveness and a results orientation. Therefore, the World Bank, for example, organized a high-level seminar in Abidjan, Côte d'Ivoire, in 1998, during which senior officials from various African countries and international development assistance agencies looked at the status – and helped to plan the future – of monitoring and evaluation (M&E) capacity development in Africa. In the past few years, UNDP has also been particularly active in the field of ECD. A survey of existing national M&E systems was undertaken in more than fifteen programme countries. In collaboration with national entities, recommendations were made to strengthen the effectiveness of these systems.

In this context, the United Nations Development Programme and the Chinese National Center for Science and Technology Evaluation co-sponsored the Beijing Conference on Evaluation Capacity Development (ECD) with strong support from the World Bank. The conference, which was held 27-28 October 1999, focused particularly on issues of performance management and public sector reform.

Participants debated intensely on such issues as transparency and the independence of the evaluation function and, as a corollary, the challenges of introducing a culture of evaluation at the national level. Some of these challenges stem from the political environment and the need to develop an appreciation at decision-making level of the importance of and the need for evaluation. Other challenges are linked to the systemic problems of institutionalizing and integrating evaluation into the government machinery and the weakness of internal capacities, human and financial, necessary to support a professional cadre of evaluators.

The development of national capacities in evaluation requires strong alliances and partnership-building. Conference participants, aware of the benefits of such relationships, showed strong interest in launching an international development evaluation association to help forging links with evaluators engaged in development related activities in both donor and programme countries and support individual efforts of countries interested in creating such an association at the national level. UNDP and the World Bank emerged as the main promoters of this international association.

Another topic of considerable interest for the Conference was the work carried out by aid agencies on the interface between results-based management (RBM) and monitoring and evaluation in connection with public sector reform and good

governance. Despite the enormous methodological and technical challenges linked to this management approach, results orientation in the public sector is clearly the way of the future.

Given the high level of participation and interest in the topics debated during the Conference, it is clear that the relatively modest investment made to finance the Conference has the potential for yielding big dividends. In particular, the Confernce highligted the strong potential for maximizing synergies between the work of development agencies, development banks and national entities to promote the role of evaluation as part of good governance and public sector reform.

It is hoped that the insight contained in the formal presentations of this volume will stimulate further discussions and developments in the field of ECD.

ACKNOWLEDGMENTS

The sponsors of this conference were:

- The United Nations Development Programme Evaluation Office (UNDP/EO)
- The National Center for Science and Technology Evaluation (NCSTE)
- The World Bank Operations Evaluation Department and World Bank Institute (WBI)

Special thanks go to Professor Fang Yan and Christine Roth for their dedicated work in organizing the Conference. Thanks also go to others who provided substantive inputs in preparing the Agenda: Ray Rist, Keith McKay and Niels Dabelstein.
This event would have not been possible without the logistical support provided by Feng Bo, Bousso N'Diaye, Victoria Vieitez,John James and the whole administrative team of the NCSTE.
The report was produced by the UNDP's Evaluatin Office staff including Christine Roth (task manager) Habila Maiga, Barbara Brewka (editor) and Anish Pradhan.

ACRONYMS

ADB:	Asian Development Bank
CBC:	Construction Bank of China
CIECC:	China International Engineering Consulting Company
CMED:	Central Monitoring and Evaluation Division
CNAO:	China National Audit Office
DAC:	Development Assistance Committee
DANIDA:	Danish International Development Agency
ECB:	Evaluation Capacity Building
ECD:	Evaluation Capacity Development
EPU:	Economic Planning Unit
EXSL:	Exemplary State and Local Programs
GASB:	Governmental Accounting Standard Board
IFI:	International Finance Institutions
IMF:	International Monetary Fund
ITD:	Industrial Technology Development
KPIO:	Key Project Inspectorate Office
MEANS:	Methods for Evaluating Actions of a Structural Nature
MOF:	Ministry of Finance
ODA:	Overseas Development Administration
OECD:	Organization for Economic Co-operation Development
PEO:	Programme Evaluation Organizations
PPMS:	Project Performance Management System
PRC:	People's Republic of China
R&D:	Research and Development
RBM:	Results-Based Management
RTO:	Research and Technology Organizations
S&T:	Science and Technology
SAA:	State Audit Administration
SDB:	State Development Bank
SDPC:	State Development Planning Commission
SERU:	Socio-Economic Research Unit
SLEA:	Sri Lanka Evaluation Association
SME:	Small and Medium Enterprise
SPC:	State Planning Commission
SPO:	State Planning Office
SRF:	Strategic Result Framework
SURF:	Sub-regional resource facilities
TBI:	Technology Business Incubators
TEC:	Training and Enterprise Council
UNDP:	United Nations Development Programme
UNFSTD:	United Nations Fund for Science and Technology Development
UNICEF:	United Nations Children's Fund
USAID:	United States Agency for International Development
USPAP:	Uniform Standards for Professional appraisal Practice
WAITRO:	World Association of Industrial and Technology Research Organization

EXECUTIVE SUMMARY

In both developed and developing countries, there is a growing awareness of the potential role that the evaluation function can play in improving the quality of public sector decision-making. Evaluation is being recognized as a mechanism for providing accountability for the use of public resources as well as an instrument for organizational learning.

Concurrent with the increasing interest in evaluation as a component of public sector development and civil service reform, evaluation capacity development (ECD) is being placed on the agenda of cooperation between developing countries and development assistance agencies.

Against this backdrop, the United Nations Development Programme (UNDP) and the Chinese National Center for Science and Technology Evaluation (NCSTE) co-sponsored the Beijing Conference on Evaluation and Capacity Development, which took place from 27-28 October 1999. This initiative also received strong support from the Asian Development Bank (ADB), the Development Assistance Committee of the Organisation for Economic Co-operation and Development (OECD/DAC) and the World Bank.

The objectives of the Conference were fourfold: (a) to stimulate reflection on the role of evaluation in good governance and public sector reform, including sharing of experiences; (b) to explore the interface between results-based management (RBM) and monitoring and evaluation in connection with good governance; (c) to identify strategies and resources for building monitoring and evaluation (M&E) supply and demand in Asian countries;and (d) to encourage and support the creation of country and regional networks to facilitate follow-up actions.

To achieve these objectives, the Conference agenda was organized around two broad categories of evaluation issues:

(a) The policy and institutional **context** of evaluation:
- Policy frameworks for public sector results-orientation
- Organizational learning vs. management accountability
- Linkage between evaluation and policy formulation, resource planning, programme monitoring and management ;

(b) The **practice** of evaluation:
- Evaluation criteria and conceptual approaches
- Options for organization of the evaluation function
- Skills, resources and methodologies
- Challenges of evaluation in specific sectors, e.g., science and technology.

The Conference brought together high-level officials from 13 Asia and Pacific countries (China, Democratic Republic of Korea, Fiji,India,Indonesia, Kazakhstan, Kyrgyzstan,Malaysia,Mongolia, Nepal, Pakistan, Sri Lanka, Viet Nam), staff of the UNDP country offices in the region, members of EVALNET (the UNDP evaluation network), a large number of representatives of the international public sector and development agencies such as the World Bank,The National Center of Public Productivity (USA), The Centre for European Evaluation Expertise,

OECD/DAC, the Asian Development Bank and resource persons from the community of international experts.

The agenda comprised formal presentations by experts in M&E and RBM that explored general themes as well as specific experiences of donor agencies and national institutions. Participants shared their experiences and discussed issues surrounding the building of a national M&E system at the country level. Part of the agenda was also designed to start the process of reflection by participants on the design of their own ECD action plans.

Discussions during the Conference highlighted some important issues faced by national entities. First and foremost,there was unanimous agreement that in a world of globalization and shrinking resources, governments and public organizations are faced with growing scrutiny from their constituency to ensure increased responsiveness, transparency and accountability in the delivery of services. In this context, the role of both evaluation as part of good governance and ECD programmes that could help to support planned public sector reforms become extremely important. Second, the independence of the evaluation function was hotly debated as were the potential constraints to the establishment of a culture of evaluation. These constraints include the political environment and the weak demand for evaluation in many countries, the systemic problems that still exist in various countries that hinder the application of sound evaluation systems, and limited internal capacities in terms of financial and qualified human resources. Third, with the spread of results-based management and the realization that development outcomes require the forging of strategic alliances, discussions centered on the importance of partnerships in joint evaluations and the need to involve stakeholders and civil society.

Based on the experiences shared by the various participants, it became quite obvious that despite the commonality of issues,there is no standardized approach to strengthening evaluation capacity development and developing effective monitoring and evaluation systems.

In view of the short period allocated for group work, the participants were not requested to develop action plans for ECD in their countries but rather to identify a few concrete, realistic actions/strategies to promote and strengthen ECD in their respective countries in the short term.

The resulting action points isolated some important elements that were seen as critical to any plan to strengthen ECD:

- **Institutional support.** Institutional support for ECD is crucial. There should be an awareness and acceptance of M&E activities in key decision-making centres. Evaluation should be promoted as an independent function and standards and criteria should be developed.

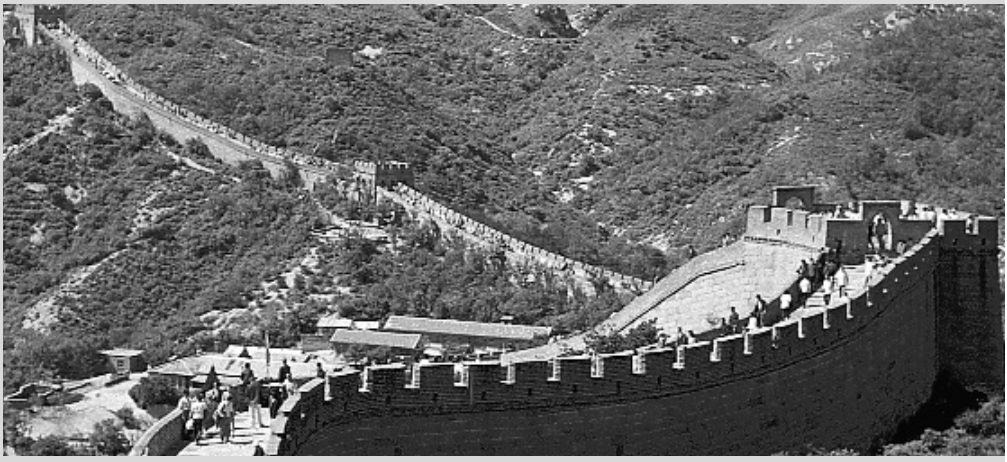
- **Training.** Training still represents an important element of any ECD programme: training of officials and technicians involved in M&E; participation of officials and civil society in joint evaluations with external aid agencies.

- **Networks.** The concept of networks emerged as a powerful instrument for facilitating the exchange of experiences, access to best practices and sharing of databases.

The establishment of evaluation networks at both the national and regional levels is the first concrete follow-up action to which participating countries made a commitment, with the support of the international community. China in particular and the National Center for Science and Technology Evaluation volunteered to establish and maintain such a network for the region.

At the international level, both UNDP and the World Bank, in agreement with several other development agencies, are now working together to promote the establishment of an international development evaluation association. The idea of the association was widely endorsed at the Beijing Conference. Such an association will help in forming links and support the individual efforts of the several countries that expressed interest in creating such an association at the national level.

PART I
The Evolving Context for ECD



1 OPENING SPEECH

Deng Nan, Vice Minister of Ministry of Science and Technology of PRC

Distinguish guests, Ladies and Gentlemen:

Good morning! First of all,please allow me on behalf of the Ministry of Science and Technology, as well as myself, to warmly welcome all the officials from UNDP and the World Bank, as well as the distinguish guests and experts,scholars on the evaluation from all the other countries participating at the International Conference on Evaluation Capacity Development.

We Chinese people have just celebrated the fiftieth anniversary of the founding of the great People's Republic of China, and together with other people in the world we are looking for the coming of the new century. Today, initiated jointly by UNDP and National Center for Science & Technology Evaluation, delegates from all over the world gathered here in Beijing, with the topic of improvement of capacity of evaluation to discuss and try to make improvement on the scientific decision making as well as the improvement of responsibility mechanism of government, and the utilization of public resources. This meeting has not only been a very important strategic impact on the course of the Chinese Open-door policy, but also a great impact on the development of all the other countries, so as to the progress of human being.

As everybody knows, humanity is facing a brand new century, with marks of information technology, and a focus on the competition of hi-tech. Technology innovation has greatly motivated the technological progress of the world. In order to grasp the future of our development, to manage the society more efficiently, for the best utilization of the resources, which are common to us, for steadily making a better living environment, the improvement of the quality of decision making and efficiency of management are vital.

Along with the development of Chinese reform and its open-door policy, the Chinese government, as early as 1986, has started to focus on the importance of scientific decision making according to the public opinions. In 1993, the Ministry of Science and Technology of China initated technological evaluation methods into the process of macro science and technological management, which is also regarded as an important breaking through of the reform of management on the Chinese National Key Science and Technological Plan. It has resulted in the foundation of the Center of the National Center of Science and Technological Evaluation. Studying international experience, coping with the Chinese realities, the National Center for Science ant Technology Evaluation (NCSTE) has made relevant and proper evaluation rules and indicators systems. The national Center for Science and Technology Evaluation has made a series of effective evaluations on the effectiveness of national key science and technological projects; the capacity of innovation, as well as the operational mechanism of the R&D insititutions and industrial bases.

Although, those activities are still explorative, people have seen the importance of the adoption of evaluation for the improvement of macro level management, the reinforcement of the responsibility of the government and public resource utilization, which are shown in the following aspects:

1. **Improve the process of decision making more scientifically:** The utilization of evaluation has optimized the process of decision making. The professional agencies on the evaluation, the utilization of standard evaluation programs

and technical ways have provided operational foundation for the scientific decision making. In the process of evaluation, the collection of information from multi-channels,different angles have provided more objective and comprehensive references for the scientific decision making. The technique based on the objective facts to get rid of personal bias, has provided warrantee on the justice of scientific decision making.

2. **Improved the capacity of adjustment of science and technical macro-management:** The introduction of technical evaluation methodology has improved the scope and depth of the viewing of management department. The third party role and standard evaluation methodology, make it possible for the management department to understand the real situation of the evaluated, in order to find the problem and key for the improvement. The evaluation analysis based on the facts, have provided references and reasons for the management department for the improvement of planning management, adjustement and allocation of resources.

3. **Promoting the innovation on the science and technological management system:** The introduction of evaluation methodology has separated the judgement from decision making, which has restrained various unhealthy trends, such as curtness, personal favor, the lack of information and poor data etc. It is also helpful to prevent corruption of the government departments. To understand the objective reality via the professinal agencies, the government departement may focus on the decision making process. This has changed the old system from point to point management to a new triangle balance system,which has provided guarantee for the correctness, effectiveness and safety of the investment from the government.

4. **Reinforcement of the authority of the making and enforcement of the National science plan:** The establishment of the process of science and technological evaluation, has helped to improve and attach importance to the plan contents for all levels and aspects, since it is from the third angle, based on the content of the plan,and verified with the reality. This practice helps to prevent people from making false for the approval from the government, thus to raise the quality of project approval. It is also helpful for promoting the management staff keeping on study the latest development, and undertaking investigations. And it is helpful for monitoring the enforcement, and for the self-conscienceness for the enforcement of the plan accordingly.

Generally speaking, in the process of setting up our socialist market economy with the Chinese characteristics, the professional evaluation, plays more and more important roles.

Recently, the central committee of CPC and the State Council clearly appointed out in the document of « Decisions on the Enforcement of Technological Innovations, Promotion of hi-tech, and industrialization of hi-tech, the evaluation agency system should be promoted in the near future. Since the capacity of evaluation is still far short from the requirement of development of the socialist economy. Thus, we are now making efforts on the building of the Chinese science and technological evaluation system.

For the promotion of the healthier development of the course of Chinese science and technology evaluation, a management rules and relevant counter measures have been under research and making, as well as the technique criterion and standards, thus to make effort on the perfection of the evaluation system, and improve the quality of the institutions professional and quality of their works.

We believe, experts and scholars from all over the world gathered together to exchange experience, which will greatly promote the building of the capacity of evaluation of China and delegates's countries. I appreciate very much the active spirit suggested in the handbook of UNDP Evaluation Office, that is « Study ». Evaluation is a process of study, a process of fighting for the realization and re-study on objective matter. Not only the management, but also the people undertaking evaluation, and the evaluated, all have the opportunity to increase their knowledge over the objectives in the process of evaluation, from which to acquire new knowledge, thus to improve the capacity to master the objective and reform the objective. The evaluation of science and technology has just started in China, thus we should even emphasized more on the spirit of study. One is to actively study the lessons internationally; the second is to cope with the characteristics and reality of the China. Thus to develop scientific and democratic principle. I believe it is the same way for the other countries.

Ladies,gentlemen, and friends, the autumn of Beijing is with a good fame of « Golden Time », the sky is blue and high. It is a season full of harvest and prospective future. For a long time, the Chinese government, the Ministry of Science and Technology of PRC enjoyed a broad,friendly and good cooperation with UNDP, the World Bank and the other countries. The opening of this meeting is not only a fruit of our past cooperation, but also a new page for the cooperation in the future in the field of evaluation. Now, please allow me to wish success to the symposium, and wish every delegates to enjoy a harvesting time in the Beijing « Golden Time» together.

Thank you.

2 The New Dimensions in Evaluation and Evaluation Capacity Development

Khalid Malik,Director EO, UNDP

When evaluators get together at a meeting, there is an inevitable interest in sharing information on concepts and good practices. And this meeting in no exception. However, this meeting is also an opportunity for us to collectively "push up" in the national and international agenda the relevance and value of evaluation concerns as an integral part of decision making and strategy setting. It should allow us to reinforce the need for evaluation capacity development at the country level, by building professional alliances and by broadening our understanding of the range of institutional capacities, strategies, resources and directions.

The dialogue started here should also help identify where genuine demand exists for evaluation to assess effectiveness and relevance of public action. It will also set the basis for exploring various modalities for the donor community to support ECD programmes in Asia.

THE CASE FOR EVALUATION

The 1999 UNDP Human Development Report is on globalization. While documenting the different forces promoting globalization and its likely consequences, it also makes a point that a lot of the momentum we see today in the world is the result of conscious decisions taken to promote an open trade and competitive environment. In a rapidly globalization and increasingly uncertain world,taking correct decisions and learning from experience become essential for survival of the firm or organization or indeed of the nation. Evaluation and the role of evaluators become more, not less important. But to grasp these new challenges, the field of evaluation itself has to develop and become more relevant to the need of the day.

As many speakers are going to talk about at this conference, the field of evaluation is changing rapidly, partly as a consequence of dealing with the new demands being placed on it. Increasingly, evaluation is being recognized as an essential part of good governance and public sector reform,and in particular in promoting transparency and accountability. Some go further and see it as part of responsible civic action and civil society development.

Let me in this paper focus on probably one of the most promising areas which has recently emerged, and, that is the interface between results-based management and evaluation. This is the future and the areas which represent enormous technical and methodological challenges for all of us.

Generally, it is fair statement to make that taxpayers, whether they finance domestic or international agencies, in donor or recipient countries, are ultimately

interested in results. Public sector performance has been an issue among citizens of industrialized countries for some time - more during the 1980's and 1990's, as taxpayers have challenged governments to demonstrate value for money in the provision of public services. And, increasingly, taxpayers are questioning the very relevance of institutions themselves and their mandates in a world for rapid change. Are police forces actually leading to better safety and reduce crime, how do related socio-economic forces affect the pictures; are R&D institutions actually leading to products and services in the market place, etc.

At the organizational level, similar questions can also be raised: what is the value of a project or programme or indeed the entire organization; and are the expected results being attained.

This enormous pressure on demonstrating results has led to the introduction of results-based management in many domestic agencies,mostly in the DAC countries, and increasingly international aid agencies. UNDP is no exception. This in turn has produced a sea of change in the methodologies within UNDP by which we evaluate and assess performance. Benchmarking, indicators,measurement are becoming familiar buzzwords in the evaluation community.

The challenge for international aid agencies has been to revisit their M & E arrangements, looking for ways to develop effective learning systems and to support their partner countries. Efforts to reshape the development assistance business are also putting pressure on evaluators to re-think and upgrade their work in capacity building in programme countries.

TOWARDS A COMMON FRAMEWORK

Evaluation is changing and is likely to change even more. A shared framework is necessary so that evaluation efforts can build on synergy and promote broad knowledge sharing. Drawing upon ongoing work, some thoughts for such a framework could be highlighted. The key building blocks may cover:

1. Moving towards the notion of outcomes so that the larger developmental purpose could be adequately addressed. In some ways, at least internationally, the global conferences supported by the UN provide a strong basis for follow-up and assessment at the country level.

   More broadly, the focus on outcomes implies that we not only look at the micro picture but also try to connect individual efforts with the macro picture, i.e.outcomes, the larger purpose of why we are doing things, whether it is to reduce poverty or promoting science and technology. This approach raised many methodological issues such as those concerning attribution, when assessing the contribution of different groups to the realization of a specific outcome.

   But the basic idea is to move away from a preoccupation with individual operations towards assessing the outcome and eventually the impact of development interventions. This approach puts the notion of results in the centre of planning and management efforts and therefore "managing for results" emerges as a consequence of this approach, with performance systems geared to delivering on results rather than an introverted risk-averse culture that stresses the control and management of inputs and activities.

2. Strategic Partnerships: Producing outcomes require the concerted effort of many actors in society and institutions - the government, the civil society and relevant aid agencies. If our focus is now on outcomes, then clearly we have of necessity to invest in partnerships which in turn requires us to have a much clearer idea of our own strength and the contribution we can make to the partnership. This approach fits in well with the idea that development is a broad-based multi-faceted exercise. Importantly, it provides a specific purpose to the need for partnerships.

   In some ways, we are really moving towards the notion of "shared accountability" which recognizes that valuable roles are played by many actors in society, from NGOs to local associations and so on. There are methodological challenges here, for instance how do we access "shared accountability" to the issue of how best we promote national capacity building in evaluation that includes the representatives of civil society, indeed as part of our aims of strengthening transparency and accountability in society.

3. Treating evaluation as a key part of good governance. I will not elaborate on this since several other speakers are talking about it. But, a brief comment may still be in order, evaluation is a natural analogue to strategic planning and management and the basis for sound decision-making.

4. And this brings me to my fourth point, we have to provide a better fit between evaluation and decision-making through what is being referred to in the literature as real time or just in time assessments. For too long, we have not adequately responded to the criticisms that evaluations are done late and are as such not of much relevance to decision making and strategy setting. A large part of this is due to the fact that the traditional focus of evaluation has been on projects. We have to move beyond this and perhaps venture into areas that hitherto have been seen as off limits to evaluators. To become relevant, evaluators have to take on additional, more difficult responsibilities, of tracking management responses and contributing to the conditions which may lead to the implementation of evaluation recommendations.

5. Refining evaluation methodologies in light of the emerging new role of evaluation in a "results-based environment":

   a) Asking the right questions:The questions that evaluators ask may well determine the answers- this may be obvious at the end, but not always at the beginning. So how do we construct an approach, which ensures that the important questions do get asked?

   b) Trying to find the right balance between M&E. There has been a tendency to separate the two. In the UN, we feel M goes with E, since the focus is very much on learning. Objective, discrete evaluations, are part of this search for learning benefiting from a structured approach to interacting with different stakeholders and trying to form a considered response to the famous evaluator's 'so what' question, you may have carried out a project well, but 'so what', what does it amount to.

   c) Link micro assessment with the macro questions. Traditional evaluation tries to form an image of effectiveness in terms of efficiency, sustainability and the achievement of objectives,essentially at the micro-level. Moving to next level may not be easy, but it is increasingly becoming necessary. We have to determine the best ways of collecting and assessing empirical evidence so that future policies and programmes draw upon the right lessons and that organizations are more focussed on doing the right things even as mandates change and we all try to adapt to a fast changing environment.

Since we are in China, let me close with a quotation from the sage Confucius: *"If concepts are not clear, words do not fit. If words do not fit, the day's work cannot be accomplished. If the day's work cannot be accomplished, morals and arts do not flourish, punishments are not just. If punishments are not just, the people do not know where to put hand or foot" (Analects XIII).*

3

ASIAN PERSPECTIVE IN EVALUATION CAPACITY DEVELOPMENT: The Challenges of the New Millennium

Adil Khan, Senior Advisor, M&E,UNOPS/UNDP, Sri Lanka

I. INTRODUCTION

The nexus between Evaluation Capacity Development (ECD) and good development management cannot be ignored.However one defines it, the emerging globalization process presents a new context for all activities including ECD. The Human Development Report 1999 stressed that as an outcome of globalization;"Shrinking space, shrinking time and disappearing boarders are linking people's lives more deeply, more intensely, more immediately than ever before".Therefore, the dictum "when a butterfly flutters its wings in Shanghai, it creates a typhoon in San Francisco" is probably more valid than ever before - meaning that the current situation imposes on each country the complex responsibility of tracking and monitoring changes, both global and local with utmost alertness and developing capacities that can help responding to these changes with equal promptness. The effects of globalization have been both good and bad for every country including that of the Asian economies. However, the recent upheavals in some of the Asian economies drive one home truth - that every country's ability to tackle the emerging challenges will very much depend on how well they can organize within themselves a good learning mechanism - a mechanism that is capable of tracking both progress as well as crises and the effective way in which the lessons learnt are feedback to the planning, policy and reform processes of a government.

The world today has more opportunities for people than ever before.Child death rates have fallen by half since 1965 and a child born today can expect to live a decade longer than a child born then. During the last three decades,school enrolment in developing countries has doubled. Adult literacy has increase from 48% in 1970 to 72% in 1997.

The world is definitely more prosperous, and a better place to live now than ever before. In the past 50 years, average per capita income tripled. The share of people enjoying medium development has risen more from 55% in 1995 to 66% in 1997 and the share of low human development fell,during the same period,from 20% to 10%.

THE CONTEXT

The current *globalization process* is indeed a very important context of evaluation. As has been explained above, with the promise of gains, globalization has also increased uncertainties,making it imperative the need for continuous monitoring and assessment of emerging situations. Likewise, the nagging problem of poverty which consistently challenges the gains of development presents the *development context* of evaluation. In this context, evaluation is seen as an important tool to learn lessons of experience and help ensure growth with equity. Similarly, post

implementation sustainability of development projects and weak *sustainability* management of invested resources present another important context of evaluation. It has been argued that post implementation monitoring and evaluation of completed projects has the potential to greatly enhance their sustainability. Problems of accountability, corruption etc. provide the *governance context* of evaluation.

It is evident that despite the phenomenal progress made during the last two or three decades, these benefits of globalization did not spread evenly. The 1999 HDR reports (UNDP, 1999):

- Nearly 1.3 billion people do not have access to clean water.
- One in seven children of primary school age are out of school. About 840 million are malnourished.
- An estimated 1.3 billion people live on incomes of less than $ 1 (1987 PPP $) a day.

In addition to the above, it is also becoming apparent that many developing countries in their quest for rapid economic development have also been bringing about, perhaps inadvertently, the most wanton destruction to their environment. These disappointing trends which are concurrent to some of the more positive outcomes of globalization are not only compromising the gains made so far, but even threatening their sustainability, which indeed indicate that serious efforts are needed to review constantly its dynamics.The concomitant changes, both positive as well as negative need to be analyzed constantly and that timely and appropriate feedback given.It is expected that such continuous monitoring of progress will help ensuring on the one hand the sustainability of the gains made and on the other, mitigate the losses incurred therein.

Against this backdrop one has also to recognize that while worldwide the investible resources are declining, demand for services are increasing implying, continuous need for prudent use of available resources and maximization of return on the invested ones. However, several recent studies suggest that large amount of resources invested in development projects did not always meet with the expected results. Some facts of life (World Bank 1993):

- a recent World Bank study indicates that, during 1981-1991, the proportion of successful projects has fallen from 85% to 63%, globally.
- the same study reports that the proportion of major problem projects had increased from 11% in 1982 to 20% in 1991.

The story is not very different with the other major regional development bank either- the Asian Development Bank's (ADB) recent studies confirm that despite experiencing more success with disbursements, successes with project sustainability are less satisfactory (Asian Development Bank, 1993).In fact, recent trend shows that less and less projects are being sustained. The implication of the latter is that while the debt burden of the recipient countries has been rising their capacity to earn income on these borrowings and service debts on borrowed resources have been falling, leaving a mounting debt burden on the future generation.

These and the lingering problems of poverty, falling governance standards and degradation of environment make it imperative that more attention be given to the aspects of: (i) improved planning; (ii) prudent and transparent use of existing resources; (iii) creation of a social infrastructure, allowing freer and wider participation of people in the decision making; and (iv) to complement the above, establishment of an institutional framework, an evaluation system, that enables achievement of good governance and sustainable development and allows continuous research, lessons learning and management feedback. In a recent paper, Mr. R. Picciotto (1999) of the World Bank echoes similar sentiments and presents the following as the contemporary context of evaluation - "Accountability, transparency, participation and the rule of law have become generally the standards of economic governance" and according to him, these elements form the basic framework for evaluation. Late Mahbub Ul Haq, the architect of the concept of "human governance" extends implicitly the context of evaluation to equity issues and states, "Every governing institution, every policy action should be judged by one critical test. How does it meet the genuine aspiration of the people" (Human Development Centre, 1999). It is only through evaluation that such a "critical test" can be installed and eventuated.

MONITORING TO EVALUATION: THE CHANGING PARADIGM OF ECD

Until recently, however, most donors as well as government attention on ECD seemed to have focussed more on monitoring than on evaluation. This has been due (i) firstly, to unsatisfactory experience with implementation;and (ii) secondly, the belief that early implementation of projects will automatically contribute to generation of expected results. However, all available evidences suggest that while monitoring did help speeding

up implementation, these did not automatically produce desired results. On the contrary, in some cases these actually acted predatorily - some development projects made things worse for the poor and in some cases, the environment got degraded.[1] In these countries, absence of an institution of evaluation deprived the decision-makers from learning the truth and correcting the situation before it was too late. Further with globalization and with integration of the national economies to the global system, many assumptions based on which projects and programmes are formulated tend also to become increasingly invalid - making it imperative the need for continuous information gathering and lessons learning and making adjustments to project conditions.

There are thus, a variety of reasons why more than ever before evaluation should be taken seriously and be treated as an important tool in development management. Such as:

I. With increasing democratisation of the societies, individuals are becoming more conscious about their rights and demanding better services and more value-for-money from public expenditure. There is only one way a government can know whether what they are doing, are in fact, measuring up to the expectations of the people - through evaluation.

II. Increase in population is creating increasing demand for more services, but resources are not increasing at the same rate. Evaluation helps to save money and ensure best use of existing resources.

III. With globalisation and liberalisation of national economies, institutions, relationships, pattern of production, pattern of economy and social behaviour etc. are constantly changing, making it extremely difficult for the governments to control and manage events to its own advantage. Constant vigilance through monitoring and evaluation and research of the changing events are seen as some of the important imperatives of good management in the new millennium.

IV. Evaluation ensures transparency and accountability in public finance, which are also key elements of good governance.

V. Evaluation can also be politically rewarding to a government which is benevolent - for, an evaluating government is also likely to be viewed by its citizens as a listening and a caring government.

In summary, concerns with efficient use of public resources,transparency and accountability, changing conditions with globalization, lingering problems with poverty and concerns with good governance etc,are making it increasingly obvious that efficient and effective evaluation of social and economic dynamics and of public sector programmes and policies are a sin qua non to good development management and an important tool to help facing the new challenges of the coming millennium.

EVALUATION CAPACITY DEVELOPMENT (ECD): THE CONCEPT

ECD can be defined as an activity or a set of activities that contribute to the establishment of evaluation capacities within the development administration structures of developing countries. Development of such capacities involves evolution of systems and methodologies that assist lessons learning from on-going or past projects and programmes and, through these lessons, adjust projects and programmes in such a manner that they achieve their planned objectives or improve the quality of design of similar projects in future.

The World Bank argues that ECD enables its member countries to develop suitable tools to ensure "quality of its loan portfolio" (World Bank, 1994: iv).The Bank thus sees ECD as a part of an overall exercise in public sector reform. It justifies the need for strong ECD by stating that:

"Effective evaluation - both "concurrent" (i.e.while a programme or project is in progress) and "ex-post" (i.e. after the programme or project is completed) - enhances the quality of public investments and of the Bank's portfolio. Well focused and properly timed evaluation can provide the information needed to bring about mid-course corrections in programmes and projects, analyze and resolve systematic policy issues, the design of future operations, and increase country ownership" (World Bank, 1994: iv).

Most other donors generally agree with the Bank's ECD strategy, although UNDP as well as some bilateral donors wish to focus more emphatically on the equity aspect of evaluation, and not limit the tool to assess financial and economic sustainability of projects and programmes only. In their view, it is important that each evaluation looks more candidly at the distribution aspects of project benefits and through lessons learnt consistently supports programmes and policies that help achieving

[1] *Alauddin and Tisdell (1989) explain how investment in shrimp culture in Bangladesh pushed small and marginal farmers out oftheir land, made rich richer and poor poorer and degraded the environment.*

human development objectives. To ensure that equity is incorporated as an important agenda in evaluation research, these donors also stress the need for incorporating the beneficiaries into the evaluation process.Thus, according to this school, participatory evaluation (mainly as a sequel to participatory development) should form the basis of all future ECD work. However, as corruption and inefficient use of public funds are also becoming concerns of many donor agencies, the aspects of transparency and accountability are being regarded as important subjects of evaluation as well.

II. THE ASIAN EXPERIENCE AND THE EMERGING ISSUES

As in many things in development, Asia also played a pioneering role in setting up formal evaluation practices in many of its countries. Interestingly enough, some of these initiatives came from within and not from outside. Presented below are some of the key lessons of ECD in Asia.

THE PIONEERS: THE INDIAN AND THE MALAYSIAN EXPERIENCE

Perhaps the most radical of all evaluation capacity building initiatives came from India (Khan,1989). Inspired by the post-colonial idealism,the political leadership of the day, led by Pandit Nehru set about to form a benevolent governance structure in the body politic of the country. Search for truth and anxiety to learn from facts to help improving the conditions of its citizens, especially that of the poor and the disadvantaged came as a natural choice. As early as 1950, a Programme Evaluation Organization (PEO) was created within the Planning Commission of India. Its mandate was to undertake post-evaluation of completed projects and give feedback to the planning process. Among other things, the most impressive aspect of this initiative has been the way the arrangements were made to ensure feedback. To guarantee maximum attention, transparency and response to evaluation findings, provisions were made to circulate the evaluation reports to the widest possible audience, including the Public Accounts Committee, the Parliament Library and the Press.It was envisaged that wider circulation of evaluation findings of government programmes would not only ensure maximum response, but by making provision to circulate these reports to the Press, a key element of good governance – "the right to know" by people might also be guaranteed. These arrangements functioned well for a while, but with the changing political and moral environment of the country such noble innovation of the bygone leadership could not be sustained. In recent years, PEO though continues to undertake evaluation work, weak political commitment to its works seemed to have made some of its initiatives somewhat less tenable. Nevertheless, the PEO model leaves one important guidance which is worth paying attention to, the provision of wider dissemination of evaluation information has indeed the potential to assure maximum feedback, an aspect which continues to remain a problem in many evaluation systems built in Asia. However, around this time another important trend in India to which a parallel similar to that of growth of evaluation could also be drawn.

While with the eroding idealism and falling standards in governance many institutions of checks and balance started to become weak in India, the institution of Judiciary emerged as the only institution that continued to serve the interest of the people, independently and fairly. But what made it possible? Many believe that two things worked in favor of Judiciary - one, the democratic environment within which the institution operated allowed transparent expression of its views, and secondly, the political and administrative immunity that the institution received from or been endowed with, by the constitution also helped the institution to operate freely and openly. By drawing lessons from this, many tend to argue that evaluation being a vital tool of check and balance and indeed, of developmental justice, as well as that of public sector accountability, its institutional arrangement cannot simply be left to the monopoly of governmental whims. Rather, evaluation should be seen as an integral part of good governance if not, of human development and thus, its institutional arrangement be hoisted in such a manner that it enjoys legislative immunity.[2] However, prevalence of a democratic environment in the country is equally important to ensure free and fair dissemination of information.

Malaysia's monitoring and evaluation capacity building, which also came from within, was started in 1960, under the direct patronage of the then Prime Minister Tun Abdur Razak (Khan, 1999). Established mainly to speed up delivery of services in the rural areas, the Malaysian M&E became attractive to many Asian governments, who like their Malaysian counterpart were also grappling with the problems of implementation. In course of

[2] *Reportedly, the evaluation organisation of Brazil has been established with constitutional backing and statutory funding which is helping it to operate independently and consistently, regardless of changes to governments and their attitudes.*

time, government also developed a separate agency for evaluation (the Socio-Economic Research Unit - SERU), but this has since been disbanded and evaluation as an activity has now been re-established as a small unit within the Economic Planning Unit (EPU), the central planning agency of the government.

Presently, not much is known about Malaysia's M&E. As far as monitoring is concerned most times are spent, almost on a daily basis, on monitoring those projects deemed important by the Prime Minister. There is, however, a Computerized reporting system to which all projects regularly report physical and financial progress. But, it is not known, whether the current system goes beyond physical and financial monitoring and assesses the results.

DONOR SUPPORT TO EVALUATION INSTITUTION BUILDING

In addition to these two pioneering initiatives, other Asian countries that established fairly permanent institutions of monitoring and evaluation are -Indonesia, the Philippines, Sri Lanka, Nepal, Thailand, China, Bangladesh and South Korea. Some of these countries also received a variety of ECD inputs from a range of donor agencies. For example UNDP intervention in 1991 saw the creation of the Central Monitoring and Evaluation Division (CMED) in Nepal in the National Planning Commission. The ADB intervention in 1992 in Sri Lanka saw the establishment of the Performance Evaluation Unit within the Ministry of Plan Implementation and Parliamentary Affairs. In 1993 World Bank supported the Ministry of Finance in China to establish a broad based institutional framework for undertaking evaluation in the country. During early to mid 1990s both UNDP and the Asian Development Bank continued to provide ECD support to various Asian countries.[3] While the UNDP support focused mainly on results based monitoring and evaluation and pursued introduction of participatory methodology in evaluation, ADB introduced what it calls the Project Performance Management System (PPMS).[4] In those cases where ADB felt that the countries were making significant progress in ECD, it followed up in those countries with second and third generation assistance. The Philippines, Sri Lanka, Thailand, Nepal and China seemed to have fallen under these categories of countries.

However, results of these ECD initiatives have been somewhat mixed. In many Asian countries evaluation continues to receive low priority. In some, evaluation has a stand-alone position with little effect on the planning process. Lessons learnt are not fed into formulation of new policies and programmes. Nor, the evaluation and the planning units are institutionally linked to enable formalization of an evaluation feedback.

CIVIL SOCIETY INVOLVEMENT IN EVALUATION: THE SRI LANKAN INITIATIVE

Despite these disappointments, in some Asian countries, for example the civil society in Sri Lanka, seemed to have taken initiatives to broaden the institutional framework of evaluation beyond the government and took initiatives to establish the Sri Lanka Evaluation Association (SLEA) in September this year.

With support from UNICEF and UNDP, several University teachers, government researchers and private sector consultants took initiatives to establish the Association. Dedicated to developing more of a culture than a mere system of evaluation, the SLEVA among others, gave itself the following key objectives: (i) to function as an advocacy group for growth of evaluation culture in the country; (ii) to help establishing ethics and standards in evaluation; and (iii) to establish national and international networking for information exchange and capacity development.

Establishment and nurturing of groups such as SLEA has to be the most exciting innovation in promotion of an evaluation culture in a society. Appearing as an advocacy group such an association has the potential to provide vital link between the civil society and the state and pursue evaluation on demand basis. Such groupings which enlist memberships of individuals and institutions of both the government as well as the non-government and academic institutions may equally help in establishing a collegial link between the government and the civil society and in many cases, help removing the misgivings associated with evaluation.

TRAINING: BROADENING THE OPTION

Another useful lesson learnt in ECD in Asia is that while these initiatives helped training a large

[3] *UNDP initiatives in preparation and publication of Country Monographs on National Monitoring and Evaluation Systems equally helped in sensitising and informing the key stakeholders about the situation of M&E in their countries. This analytical exercise provided useful foundation for future ECD.*

[4] *PPMS basically means management of the Project Cycle, with inputs from evaluation.*

number of officials in evaluation methods and techniques,these gains could not be sustained fully over time. Three main reasons are given for this lapse – (i) frequent turnover of staff from M&E institutions to other agencies; and (ii) secondly, poor staff quality; and (iii) thirdly, absence of facilities conducive to application of evaluation skills learnt through ECD. Moreover, training in evaluation continues to be an ad hoc, oneoff activity. Further, most government training institutes that received ECD inputs, tend generally to operate in isolation from the other non-government academic institutions and the Universities, particularly from those (overseas institutes) that are engaged in innovative research in evaluation (e.g. participatory evaluation etc.). Absence of this institutional linkage tends to make it difficult for the government M&E training institutes to maintain required standards. It is, therefore important that future ECD work in training consider on the one hand, creating better linkage between the government and the non-government training institutes and, on the other, involve the national level universities and non-government research institutions to incorporate courses within their curricular, and build research capacities in evaluation.Many believe that incorporation of Universities in ECD is likely to yield following important gains:(i) Firstly, the Universities, unlike the government training institutes will have built-in incentives to run and improve courses as well as undertake research and publications in evaluation; and thus ensure better the aspect of sustainability of ECD inputs; and (ii) secondly, continuous training through its graduate and post-graduate courses shall guarantee stream of graduates trained in evaluation, thus consistently filling the stock of qualified evaluators in a country.

METHODOLOGIES AND TECHNIQUES: MOVING TO NEW DIRECTIONS

Suitable methodologies to measure substantive aspects of a project - that is, project effects and impacts, remain a matter of highest concern in all evaluation work. First of all,there are complaints that evaluation research is too costly and that it takes too long to be of any benefit to management. Secondly, debates also occur with regard to sampling and research designs that are capable of making "before-after" and "with-without" comparisons.There are also contentious arguments with regard to "indicators" of success. It has been stated in a recent study that in the Philippines many evaluation studies - mostly undertaken by academics - are rejected by program managers due to disagreements with one or more of the above factors (Callanta in Khan,1989). Callanta suggests that the absence of a wider pre-evaluation stakeholder analysis and lack of adequate consultation with regard to the need, objectives and methodologies of evaluation risk both the credibility as well as the ownership and the use of evaluation.[5]

In recent years the introduction of *rapid and participatory* methods in evaluation has caught the attention of many and indicates an increasing willingness to apply rapid methods to conduct evaluation. Evaluators of developing countries require extensive training in these methods. But, at the same time, they also need to bear in mind that the *rapid method* is not a substitute for the *rigorous method*, rather, one complements the other and the choice of one method against the other would very much depend upon the evaluation questions that one would need to answer.[6]

In the matter of methodology training in ECD, some donors tend to introduce their own in-house jargons which seem to create some confusions among the developing country practitioners.

Recently, in the arena of evaluation methodology there are discussions about involving the beneficiaries into the evaluation process. This undoubtedly is a positive initiative and has the potential to enhance ownership of evaluation at all levels. Such emphasis will also enable gathering of more qualitative information. However, in societies where beneficiaries are divided by unequal power structure and where socio-economic relationships are unevenly structured, the ability of the evaluators to access the targeted beneficiaries may prove somewhat difficult. In such situations some non-threatening interactive methods of data gathering may become the obvious choice. Training in interactive and non-threatening methods of data gathering can be another aspect of future ECD activities.

With regard to participatory evaluation, Chambers (1997) talks about "empowerment evaluation" which involves incorporation of the beneficiaries right

[5] *Yee (1999) talks of similar difficulties when his team wanted to evaluate the impact of Asian Development Bank assistance in China (The study was called, Country Assistance Programme Evaluation -CARE) and concludes that wider participation ofprogramme stakeholders and adoption of qualitative assessment principles help assessing impacts in a more "plausible"manner. The study also introduced a agreed-upon simple development effectiveness index to assess success.*

[6] *A recent concept paper, Impact assessment ofUNDP Country Interventions,argues that qualitative studies (quite often the, rapid and participatory studies) are useful pre-requisite to quantitative studies. However, the paper cautions that, "by relying upon qualitative analysis alone, one risks taking perception for reality".*

WHAT IT TAKES TO MAKE IT WORK

Monitoring and evaluation (M&E) systems - especially evaluation systems are neither easy to introduce nor are they easy to sustain, especially a good quality system.

A recent United Nations Development Programme (UNDP) study - *Generic Issues in Monitoring and Evaluation. What Works and Does Not* - reflect that success and sustainability of an M&E system depends on a variety of factors, including (Khan, 1993):

- **political support and its continuity.**
- **suitable institutional framework with adequate feedback linkage.**
- **staff training and logistics support.**
- **demonstrated benefits of evaluation.**
- **democratic governance environment.**
- **develop a "culture" rather than a system - Link civil societies.**

The study argues that all of these factors are equally important and that weakening of any one of these factors may has threaten the success of the entire system. For example,the study reports that while political support at the introduction may assist in introduction and institutionalization of a system, if that support is not maintained (by way of use and backing), its operational efficiency and effectiveness will also ultimately weaken. Similarly, a well laid out institutional arrangement will not yield the desired result if it is not backed up by good quality staff and a good feedback arrangement. The study concludes that for M&E to be successful,all the factors that have been identified as key to the success of the system must be equally emphasized and provided with. A lapse of any one element, the report repeatedly argues, will lead to the collapse of the entire system.

at the planning stage where they also become the principal evaluators and through evaluative participation, empower themselves to drive the implementing process to targeted results.[7]

The past ECD works in Asia have also identified a variety of operational problems that affect administration of efficient evaluation. Principal among these is the lack of basic project data. In many countries,project offices are either dismantled or reduced in size after a project has been completed and most of the staff get shifted to another project. These changes affect undertaking of quality evaluation. Moving staff away not only means project knowledge is lost but in many cases, even the project documents get lost. Further, the lack of a Project Implementation or Completion Report (PCR) equally constrains the opportunity to capture updated data at the time of completion of implementation of a project. Efforts to build new evaluation capacities in developing countries must consider ways of collecting and collating project information in one place. Establishment of a documentation centre with briefs of completed projects will greatly assist in alleviating this problem. Further, active involvement of beneficiaries into the evaluation process,is also likely to help retention of data at the grass-root level.

In summary, it appears that as far as methodology is concerned, there appears to be a lot of room for improvement and problems with time and money indicate that more emphasis should be given to rapid methods and innovative research designs. Involvement of Universities/research institutes in evaluation research may also be seen as an important part of future ECD activities.

FEEDBACK: DEMOCRATISING THE PROCESS

Feedback relates to methods of linking evaluation findings to the decision-making process of the government, particularly its planning process. In Asia different types of methods and channels are currently being used to disseminate evaluation information. The most commonly used methods are: (i) reports/returns; (ii) review meetings; (iii) Workshops/ seminars; and (iv) newsletters and computer networking. As a result of ECD interventions,some countries have tried to institutionalise feedback arrangements, without much success. Further, many of the evaluation organisations seem to continue to function in isolation from the planning institution.[8] This institutional gap seems to be defeating the very purpose of evaluation. It was stated earlier how India had a system of sharing evaluation information with a wide audience including the Parliamentarians and the general public. However, falling governance standards, weak government commitment and politically vulnerable institutional arrangement seemed to have weakened that resolve. Increased democratisation

[7] *In Bangladesh,BRAC's (an NGO) Rural Development Programme seemed to have incorporated this methodology quite successfully.*

[8] *For example, in Sri Lanka the Performance Evaluation Unit of the Ministry of Plan Implementation is not institutionally linked to the planning process. New policies and programmes continue to be developed without much consultation with the Performance Evaluation Unit.*

of societies, legislative enshrinment of evaluation institutions,involvement of civil societies and most importantly, development of democratic culture in a country seem to represent the important elements, if not the key challenges of evaluation capacity development and feedback in the new millennium. Improvement to evaluation feedback must go hand-in-hand with other democratic reforms in a country.

All in all, while recent ECD works made important contributions in introducing and initiating evaluation practices in many Asian countries,further works are necessary to make these interventions institutionally better structured and operationally more efficient. There is indeed a need for holistic thinking on ECD and further work on this aspect of development management should weigh carefully the reform process within which each society is evolving. Work on ECD should contextualise the work on democratic reforms that are currently being undertaken in many countries and support institutional development in evaluation in such a manner that the activity itself becomes both an agent of change as well as the beneficiary of such a charge.

III. CHALLENGES OF ECD IN THE NEW MILLENNIUM

Overall, it is reasonable to state that the most impressive aspect of ECD interventions has been the recognition to and establishment of separate evaluation units in most developing countries, especially in Asia. There are also evidences that due to improved evaluation,project performance in some of these countries has also improved.However, despite these gains, practice of evaluation as an obvious and untempered element of good governance is yet to be fully realized. The new challenges of evaluation stem from a variety of factors. Broadly, these are institutional related, but there are several organizational matters which also need to be looked at.

THE CHALLENGES

As has been observed, a variety of challenges still remain:

- tentativeness with evaluation structures make these units vulnerable to changing government moods.
- confining institutional development of evaluation purely within government structures,leaves out a gap and tend to exclude the most important actor, the civil society - from an activity that is so vital to good governance.
- institutional development in evaluation capacity has been viewed by many as purely an aid management ritual and not as a tool for good development management. To ensure sustainability of evaluation, its institutional development (both as an organization as well as a function) may require legislative backing. Prevalence of a democratic environment in a society may also be seen as another important element of evaluation capacity development.
- limiting evaluation training purely within the government training institutes encountered usual difficulties associated with many government institutions of this nature - limited research, limited exposure to outside world, low incentives for high standards etc. risked quality and sustainability.
- poor staff quality often contributed to poor products coming out of evaluation, affecting the credibility of the organisation and even the merit of the practice.
- inappropriate and costly methodologies contribute to low reliability and poor acceptance of evaluation.
- inadequate or poor feedback remains a nagging problem.

THE OPTIONS

The new ECD measures must consider these difficulties seriously and come up with solutions, which are consistent with the changing political and economic environment of the world. It is in this context following suggestions are put forward:

- More emphasis should be given to the development of an evaluation culture rather than just building systems within government institutions.The former would require linking the civil societies - the Universities, research institutions, evaluation societies etc. - to the evaluation process of the government.

  In this context it is strongly suggested that future ECD measures should include in its various elements the aspect of providing support to the growth of evaluation societies in each and every country and through this foster a collegial link between the civil society, and the government.

- Not just for making evaluation effective, but to ensure growth of good governance, continuous efforts should also be made to systematically democratise all societies, create an environments

of free and fair exchange of information and give the ordinary people the opportunity to exercise their "right to know and right to tell'. It is only in an environment of openness that evaluation and for that matter, all other institutions of checks and balance and lessons learning can grow.

- In community development or in poverty alleviation activities, Patton's "Development Evaluation" (where evaluation is used to continuously monitor change and with lessons learnt make adjustments to achieve goals) and Chamber's "Empowerment Evaluation" (where beneficiaries are included in the evaluation process to drive project process to intended results) methodologies should be given high priority in ECD agenda.[9]

- Future ECD inputs in training should seriously consider (in addition to capacity building of government training institutions) providing support to the Universities/research institutions in building capacities in evaluation training and research.

Evaluation can be a highly political exercise. It is a game of truth. It can be both pleasant as well as unpleasant.However, truth must be told truthfully, but constructively. Evaluators need to be sensitive to the difficulties of implementers and be conscious of the environment within which they operate. It must be remembered that role of evaluator is not that of a development police but that of a development facilitator. More and more, evaluation should be seen as a resource, a reform agent and as an arbitrator of development justice. Constructively used, evaluation can be very effective in delivering both equity and justice as well as efficiency. Governments should be made aware of this.

The governments need to remove from themselves the fear of evaluation. They need to realise that evaluation is one tool that will help them doing the right thing at the right time. They need to be shown that it is an inseparable element in the journey to good governance and sustainable human development.

REFERENCES

Alauddin, M. and Tisdell,C. (1989), "Poverty Resource Distribution and Security. The Impact of New Agricultural Technology in Rural Bangladesh", *Journal of Development Studies,* Volume 24 Number 4, Pages 510 -570.

The Asian Development Bank (1993) *Country Synthesis Report on Post Evaluation findings.* The Asian Development Bank. Manila, 1993.

Chambers,R. (1997) *Whose Reality Counts? Putting the Last First.* London Intermediate Technology Publications,1997.

Khan, M. Adil (1989) *A South Asian Region Study on Current Thoughts and Practices in Monitoring and Evaluation.* The Economic Development Institute. The World Bank Washington D.C. 1989.

Khan, M. Adil (1991) *Initiatives in Efficiency: Experience of Malaysia in Monitoring and Evaluation.* Central Evaluation Office. Unites Nations Development Programme. New York. 1991.

Khan,M. Adil (1993) *Generic Issues in Monitoring and Evaluation: What Works and What Does Not.* Central Evaluation Office. United Nations Development Programme,New York,1993.

Patton,M. S. (1999) "Development Evaluation" *Evaluation Practice,* 1994,15 (3),311-319.

Picciotto, R. (1999)" The New Challenges of Evaluation Capability Development Programme" *Evaluation Insights. Evaluation Co-operation Group Newsletter.* June 1999 Volume 1,Number 1.

United Nations Development Programme. *Human development report,1999.* UNDP.

The World Bank (1993). *Evaluation results for 1991.* Operations Evaluation Department. The World Bank. Washington D.C. 1993 .

Yee, Y.L.(1999). "The ADB's first Country Assistance Programme Evaluation". *Evaluation Insights. Evaluation Co-operation Group Newsletter.* June 1999 Volume 1,Number 1.

[7] *These approaches also provide the general methodological framework for results-based M&E – which basically imply: (i) that the process must be seen as dynamic and flexible and not rigid;and (ii) that evaluation should be treated as a continuous planning tool and not merely a post-facto activity and institutionally linked as such.*

4 AID EVALUATION: A Donor Perspective

Niels Dabelstein, Head of Evaluation Secretariat DANIDA,OECD/DA CWP

In this intervention I would like to address four issues:

The DAC Evaluation Group; Aid evaluation from a donor's perspective; the move towards partner country responsibility and, what donors can do.

Bilateral and multilateral aid agencies have a long tradition of evaluation beginning in the fifties and becoming institutionalised and systematic in the mid-sixties.In most donor countries, development aid agencies/ministries established formalised evaluation procedures long before domestically oriented public institutions.

In order to enhance the quality and utilisation of evaluations the Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD), established a Working Party on Aid Evaluation already in 1982.

The purpose of the Working Party is to:

i. Strengthening exchange of information and experience in order to: improve evaluation activities; encourage standardisation of methodological and conceptual frameworks; and improve co-ordination in planning major evaluations;
ii. Contributing to improved aid effectiveness through the synthesis of evaluation lessons;
iii. Examining the possibility of launching joint studies of aid effectiveness;
iv. Seeking ways to promote developing countries' own evaluation capabilities.

Today, it has developed into the leading forum for exchange of information and experience between donor agencies; for development of methodologies; standardising of terminology; and co-ordination of joint evaluation efforts.The WP-EV today consists of representatives of evaluation units in bilateral donor agencies of OECD member countries and the European Commission. In addition, UNDP, the World Bank, IMF, and the regional development banks (AsDB, AfDB, IADB) hold status as observers and participate as such in WP-EV meetings twice a year.

The Working Party has established an evaluation definition widely recognised and adopted by its members. Evaluation is:

"An examination as systematic and objective as possible of an on-going or completed project or programme, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, developmental efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors."

Generally, aid evaluation performs several functions: it provides lessons for managers and staff of ongoing aid programmes thereby leading to improvements in implementation; it establishes information and documents experience that can be used in planning and designing future aid programmes; it is used by managers, strategy- and policy makers for revising existing — and for devising new — aid strategies and policies; and finally aid evaluation delivers information on efficiency and

effectiveness of aid programmes, and thus ensures accountability towards politicians and the public.

Government financed development aid is expenditure of taxpayers money outside national borders. In addition, aid is often spent through partners or co-operating governments over which the donor government has no formal control. This is a feature distinguishing aid evaluation from evaluation of most other publicly funded programmes. For that reason alone, the accountability function may very well be the most important for aid evaluation. Not the least in periods of aid fatigue and pressure on government budgets, continued public support for aid expenditure is contingent upon aid agencies' accountability and their ability to document aid results.

Development aid evaluation has continuously developed from an early focus on individual activities and projects to broader evaluations of sectors, programmes, and cross cutting issues or themes such as gender, environment, institutional development and sustainability. Aid evaluations are no longer simply about evaluating efficiency, effectiveness and impact of aid in terms of technical, economic, socio-cultural, institutional and environmental factors, but about evaluating the combined effects on developing societies of the complex of political, economic and technical interventions by donors.

The emergence of new types of aid interventions and the inclusion of novel themes in development aid pose new challenges to aid evaluators. Some of those challenges are methodological:how to evaluate support to good governance, human rights, civil service reform, and privatisation. Others are wider ranging: the current trend to move away from project aid and to more varied and flexible modes of assistance within a sector framework frequently addressing key policy and institutional problems.

These new trends and challenges need to be addressed. Donors alone should not bear the responsibility for evaluation alone.More emphasis on joint evaluations would be one way. Another is increased emphasis on evaluation capacity building. Through such emphases,many of the future evaluation challenges can be addressed.

The developing countries, in their own interest, should take on these challenges and further develop evaluation capacities to assess the performance of the public sector to which development aid is, after all, a small input.

Evaluation institutions exist in many developing countries, but most have little impact on policy making and management decisions,partly because there is little demand for independent and transparent evaluation. Governments are not yet fully results oriented. The need for accountability, participation, transparency, rule of law has only recently been considered necessary for improving governance practices. Moreover, the demand for evaluation comes more often from the donors community (which has limited influence), than from the partner countries' Parliament or taxpayers.

It appears that credible evaluation is a function of good governance i.e. demand for accountability more than of evaluation institution and professional capacity. Recent years' focus on public sector reform, on public expenditures management, on democratic reforms, and on development effectiveness has, I hope, increased the demand for evaluation as a necessary and useful element of good governance. There are many ways in which donors can support evaluation capacity:

- Promote an agency ECB support policy or strategy, particularly in view of new aid forms being introduced,including programme assistance for institution and capacity building as part of good governance initiatives at national and sector levels.
- Advocate and stimulate the evaluation issue in country dialogues and sector programme assistance.
- Provide technical advice and advise on training facilities and materials on evaluation issues.
- Support the establishment of twinning arrangements between other domestic evaluation institutions and host country institutions.
- Develop consistent evaluation methodologies and terminology.
- Co-ordinate their evaluation programmes with host countries and other donors in order to optimise the use of resources and the constrained capacity of recipient countries' evaluation systems.
- Arrange joint-evaluations with a genuine participatory approach, where the needs of both parties are incorporated from the start and where the capacity building element is taken into account specifically.

In my opinion joint evaluation is one of the most concrete tools to enhance evaluation capacity. To ensure that evaluations become efficient learning tools,promote good governance, enable the partners to be fully accountable, and are cost effective they must be planned and executed jointly with the recipients. Ultimately, both the donors and the recipient authorities are jointly accountable both to the taxpayers in developing countries and to the people of the developing countries.

PART II
Evaluation of Performance Management in the Public Sector

5

RESULTS BASED MANAGEMENT IN THE DEVELOPMENT COOPERATION AGENCIES: A Review of Experience

Annette Binnendijk,Consultant,UNDP

OVERVIEW

This paper provides a broad overview of some key concepts and elements of results based management (also called performance management), which has been a central feature of recent public sector reforms in the OECD countries. While most of the remarks made and issues raised apply generally to government agencies' efforts to implement results based management, the particular focus taken here is from the perspective of donor agencies. Experience and examples are drawn primarily from USAID, although other donors' experiences shared at a Workshop on Performance Management and Evaluation held in New York in October 1998 are also taken into account.

PUBLIC SECTOR REFORMS

During the 1990s, many of the OECD countries have undertaken extensive public sector reforms in response to economic, social and political pressures. Budget deficits, structural problems, growing competitiveness and globalization, lack of public confidence in government, and growing demands for better and more responsive services and for more accountability have all been contributing factors. Popular catch phrases such as "Reinventing government", "Doing more with less", "Demonstrating value for money",etc.describe the movement towards public sector reforms that have become prevalent in the OECD countries. Often, government-wide legislation or executive orders have driven and guided the public sector reforms.

While there have been variations in the reform packages implemented in the OECD countries, there have been many common aspects. For example:

- Focus on performance issues and on achieving results
- Devolution of management authority and responsibility
- Orientation to customer needs and preferences
- Participation by stakeholders
- Reform of budget processes and financial management systems
- Application of modern management practices

RESULTS BASED MANAGEMENT

Perhaps the most central feature of the reforms has been the emphasis on improving performance – that is, on ensuring that government activities achieve desired results. A recent study of the experiences of ten OECD Member countries

with introducing performance management showed that it was a key feature in the reform efforts of all ten. (See *In Search of Results: Public Management Practices,* OECD, 1997).

Performance management, also referred to as results based management, can be defined as a broad management strategy aimed at achieving important changes in the way government agencies operate, with improving performance (achieving better results) as the central orientation.

A key component of results based management is performance measurement, which is the process of objectively measuring how well an agency is meeting its stated goals or objectives. It typically involves several phases: e.g., articulating and agreeing on objectives, selecting indicators and setting targets, monitoring performance (collecting data on results), and analyzing those results vis-a-vis the targets. While performance measurement is concerned more narrowly with the production or supply of performance information, performance management is broader. It is equally concerned with generating management demand for performance information — that is, with its uses in policy, program, and budget decision-making processes and with establishing organizational procedures, mechanisms and incentives that actively encourage its use. In an effective performance management system,achieving results and continuous improvement based on performance information is central to the management process.

THE ROLE OF EVALUATION IN PERFORMANCE MANAGEMENT

The role of evaluation vis-a-vis performance management systems in many cases was not initially clarified, and still remains somewhat vague. In part, this is because evaluation was well established in many OECD governments before the introduction of performance management and the new approaches did not necessarily incorporate evaluation. New performance management techniques were developed partly in response to perceived failures of evaluation; for example, that the uses of evaluation findings were limited relative to their costs. Moreover, evaluation was often viewed as a specialized function carried out by external experts or independent units, whereas performance management, which involves reforming core management processes,was essentially the responsibility of managers within the organization.

Most OECD governments now see evaluation as part of the overall performance management framework, although perspectives concerning the degree of integration and independence vary. The view of evaluation as a separate or specialized function, but integrated into performance management has been gaining momentum. This is reflected in PUMA's *Best Practice Guidelines for Evaluation* (OECD, 1998) which was endorsed by the Public Management Committee.The Guidelines state that "evaluations must be part of a wider performance management framework". Less emphasis is placed on independence, and evaluation is seen as one of many instruments used in the overall performance management framework. However, some degree of independent evaluation capacity is being preserved; such as the evaluations conducted by central evaluation offices or performance audits carried out by audit offices.

Evaluation is viewed as complementary to — and in some respects superior to — routine performance measurement techniques. For example, evaluation allows for more in-depth study of performance, can cover a longer time period, may incorporate factors such as outcome or impact too difficult or expensive to assess through on-going monitoring, and can analyze causes and effects in detail.

Nevertheless, some concerns over the future of the evaluation function remains. Establishing performance measurement systems are taking considerable effort and resources, often without additional funds being allocated for this purpose. In situations of diminishing overall budgets, this may well lead to a competition for resources between performance measurement and evaluation activities within OECD government agencies.

USES OF PERFORMANCE INFORMATION

Performance should not be measured for measurement's sake. Performance information, both from performance monitoring and evaluation sources, should be used. One key use is for transparent reporting on performance and results achieved to external stakeholder audiences. In many cases, government-wide legislation or executive orders have recently mandated such reporting. Moreover, such reporting can be useful in competition for funds by convincing a skeptical public or legislature that the agency programs produce significant results and provide "value for money". Annual performance reports are often directed to ministers, parliament, stakeholders, customers, and the general public.

Performance information should also be used for internal purposes,such as for management decision-making and identifying areas for improvement. This requires that performance information be integrated into key management systems and processes of the organization; such as in policy formulation, in project/program planning and management, and in budget allocation processes.

Performance budgeting is the term generally used to refer to the infusion of performance information into resource allocation processes. The concept of performance budgeting is essentially the process of linking budget levels to expected (or actual) results, rather than to inputs and activities. Many OECD countries are experimenting with a variety of approaches to more closely associate expected performance with requested funding levels, as part of their broader reforms efforts to become more results-oriented.(See OECD, *Budgeting for Results: Perspectives on Public Expenditure Management,* 1995). A recent report by the U.S. General Accounting Office reviews some of the many challenges facing U.S. government agencies attempting to implement performance budgeting, including a variety of performance measurement issues, cost accounting shortcomings, and the essentially political nature of budget allocation processes.(See GAO, *Performance Budgeting: Initial Agency Experiences Provide a Foundation to Assess Future Directions,* July 1999)

When performance information is used in internal management processes with the aim of improving performance and achieving better results, this is often referred to as managing-for-results. Such actual use of performance information has often been a weakness of performance management in the OECD countries. Too often, government agencies have emphasized performance measurement for external reporting only, with little attention given to putting the performance information to use in internal management decision-making processes.

These two major uses of performance information may not be completely compatible with one another, and may require different types of results data and collection methods. Satisfying both needs or uses without over-burdening the performance management system remains a challenge.

KEY PHASES OF RESULTS BASED MANAGEMENT

Some key elements or phases of results based management include:

1. Identifying in clear, measurable terms the objectives or results being sought and developing a conceptual framework for how the results will be achieved.

2. Selecting indicators that will be used to measure progress towards each objective.

3. Setting explicit targets (i.e., planned results to be achieved by specific dates) for each indicator that will be used to judge performance.

4. Developing performance monitoring systems to regularly collect data on actual results achieved.

5. Analyzing and reporting actual results vis-a-vis the targets (or other criteria for making judgements about performance).

6. Integrating evaluations to provide complementary information on performance not readily available from performance monitoring systems.

7. Using performance information (from both performance monitoring and evaluation sources) for internal management learning and decision-making and for external reporting to stakeholders on results achieved.

The first three phases or steps generally relate to a results-oriented planning approach, sometimes referred to as strategic planning. The first five steps, together, are usually included in the concept of performance measurement. All seven phases combined are essential to an effective results based management system. That is, integrating complementary information from both evaluation and performance monitoring systems and ensuring management's use of this information are viewed in this paper as critical aspects of results based management.

OTHER COMPONENTS OF RESULTS BASED MANAGEMENT

Other reforms are often associated with results based management systems. Often, these other components act to stimulate or facilitate the use of performance information. Some of these organizational changes include:

- **Accountability** — instituting new mechanisms for holding agency managers and staff accountable for achieving results within their sphere of control or influence (e.g., results-oriented management contracts and personnel appraisals).

- **Decentralization** — delegation of authority outs to the field and down to the management level that's being held accountable for results. In other words, empowering managers at appropriate levels with flexibility to shift resources from poorer to better performing activities.

- **Client focus** — consulting with beneficiary groups concerning their preferences and satisfaction with goods and services provided, and being responsive to their needs and desires.

- **Participation** — including partners and stakeholders in all aspects of performance measurement and management processes (e.g., jointly setting objectives, defining indicators, collecting, analyzing and reviewing data, conducting evaluations, and using the information for learning and decision-making).

- **Reformed policies and procedures** — new policy and procedure directives for changing the way the agency conducts its business (e.g., new requirements, roles and responsibilities for strategic planning, for performance measurement and evaluation, and for use of performance information in external reporting and internal decision-making processes).

- **Supportive mechanisms** — various ways of assisting managers to effectively implement performance management, such as providing reengineering training, technical assistance services, new performance information databases, guidebooks, tips and best practices series, and other management tools.

- **Cultural change** — equally important for successful implementation of results based management is transforming the organizational culture and attitudes (e.g., reorientation towards achieving results rather than implementing inputs and processes, an openness to learning from failures as well as successes, and a commitment to objective and transparent performance reporting).

SPECIAL CHALLENGES FACING THE DONOR AGENCIES

As has been the case more broadly for the public sector of the OECD countries, the donor agencies have faced considerable external pressures to reform their management systems to become more effective and results-oriented. "Aid fatigue", the public's perception that aid programs are failing to produce significant development results, declining aid budgets, and the government-wide reforms have all contributed to the donor agencies' recent efforts to establish results-oriented management systems.

Thus far, the donor agencies have gained most experience with establishing performance measurement systems — that is, with the provision of performance information — and some experience with external reporting on results. There is less documented experience with the actual use of performance information for internal management decision-making.

Results based management and measurement processes may take place at various organizational levels within the donor agencies. The first level, which has been established the longest and for which there is most experience, is at the project level. However, some agencies such as USAID have also recently established more strategic country level systems within their country operating units. Moreover, establishing performance measurement and management systems at the third level — the corporate or agency-wide level — is now taking on urgent importance due to increasing public pressures and government-wide mandates requiring annual reporting on agency-wide performance and results. Effectively linking and aggregating performance measures across these various levels remains an issue.

The donor agencies face special challenges and issues in establishing their performance management and measurement systems, that are either unique or more pronounced than those confronting the domestic government agencies. This can make establishing performance measurement systems more complex and costly than normal. For example, these agencies:

- Work in many different countries and contexts.
- Have a wide diversity of projects in multiple sectors.
- Often focus on capacity building activities, which are harder to measure than direct service delivery activities.
- Are moving into new areas such as the environment, where there's little performance measurement experience.
- Often lack standard indicators on results that can be easily aggregated across projects.
- Are usually only minor actors affecting impacts, with consequent problems in attributing them to their agency's activities.
- Typically rely on outcome and impact data collected by partner countries, who have limited

technical capacity and resources, with consequent quality, coverage and timeliness problems.

In particular, these factors can complicate donor agencies'efforts to aggregate results across projects and programs to higher organizational and agency-wide levels.

PERFORMANCE MEASUREMENT AT THE PROJECT LEVEL

Performance measurement at the project level is concerned with measuring both a project's implementation progress and results achieved. Two broad types of project performance measurement might be distinguished. (1) *Implementation measurement* is concerned with whether or not project inputs and activities are in compliance with design budgets, workplans, and schedules, and (2) *results measurement* is concerned with whether or not actual results are achieved as planned. Results are usually measured at three levels — immediate outputs, intermediate outcomes and long-term impacts.

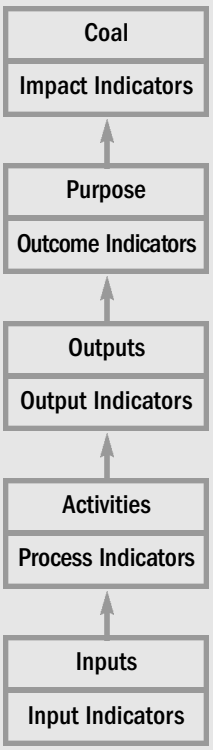
Whereas traditionally the emphasis has been mostly on implementation concerns, with the rise of results based management the focus is increasingly on measurement of results. Moreover, emphasis is shifting from more immediate outputs to longer-term outcomes and impacts.

Measuring performance at the project level can be divided into five elements or phases, as briefly outlined below. The importance of participatory or collaborative approaches in all phases of performance measurement is stressed — that is, involving not only donor agency project managers but also representatives from the implementing agency, the partner government, the intended beneficiary groups, and other stakeholders. This helps build agreement around the project's objectives and commitment to the performance measurement process.

1. FORMULATING OBJECTIVES

As part of project planning, the project's objectives should be clarified by defining precise and measurable statements concerning the results to be achieved (outputs. purpose, and goal) and then identifying the strategies or means (inputs and activities) for meeting those objectives.The project logframe is a favorite tool used for conceptualizing or modeling a project's objectives and the strategies that will be used to attain them. (See Figure 1). The logframe is based on a five-level hierarchy model with assumed cause-effect relationships among them, with those at the lower level of the hierarchy contributing to the attainment of those above. Thus, *inputs* are used to undertake project activities that lead to the delivery of *outputs (goods/services),* that lead to the attainment of the project purpose that contributes to project goal.

FIGURE 1: PROJECT LOGFRAME HIERARCHY LEVELS AND TYPES OF INDICATORS



2. SELECTING INDICATORS

Next, performance indicators are selected for measuring progress in implementing activities and in achieving results. The logframe provides a hierarchical structure around which the indicators are typically constructed.Indicators specify what to measure along a scale or dimension in order to gauge progress (e.g.,number of workshops held, percentage of farmers attending demonstration sessions, changes in crop yields, etc.). The relative importance of indicator types is likely to change over the project's life cycle, with more emphasis on input and process indicators at first, then shifting to output, outcome (purpose-level), and impact (goal-level) indicators later on as the project matures.Also, different management levels tend to place emphasis on different indicator types. For example, project field staff will find input and process indicators of most use, whereas project managers will be more interested in achievement of project outputs and purpose/outcomes. Senior agency officials will be interested in the longer-term and broader social and economic impacts of the project, which may not be evident until after the project is completed.These different intended uses and users need to be kept in mind when selecting indicators.

Often checklists of criteria are devised, against which proposed indicators can be judged and selected. For example, some commonly used indicator selection criteria include:

FIGURE 2: PROJECT PERFORMANCE MONITORING PLAN

(Format for Recording Key Aspects of Data Collection)

Type of Indicator	Indicators and Definitions	Data Sources	Data Collection Methods	Frequency and Schedule of Collection	Responsibility for Data Acquisition
Impact Indicators					
Outcome Indicators					
Output Indicators					
Process Indicators					
Input Indicators					
Risk Indicators					

- **Valid** — Does the indicator directly represent the result it is intended to measure?
- **Objective** — Is the definition precise and unambiguous about what is to be measured?
- **Reliable** — Is the data consistent or comparable over time?
- **Practical** — Can data be collected easily, on a timely basis and at reasonable cost?
- **Useful** — Will the data have utility for decision-making and learning?
- **Owned** — Do stakeholders agree that this indicator makes sense to use?

Tradeoffs among these indicator selection criteria may exist. Probably the most important, overarching consideration is that the indicators provide managers with the information they need to do their job. While on the one hand, indicator data should be of sufficient quality to be credible and ensure the right decisions are made, on the other hand it should be practical — timely and affordable. Care should be taken to limit the number of indicators selected to the minimum needed to adequately capture the key dimensions of a result. Keep the performance measurement system simple to avoid overburdening managers and staff with unnecessary data collection responsibilities.

3. SETTING TARGETS

Once indicators have been identified, actual baseline values should be collected for each, ideally just before the project gets underway. This will be important for gauging whether progress is being made later. Often agencies also set explicit targets for their indicators. A target specifies a particular value for an indicator to be accomplished within a given timeframe. (For example, 200 workshops to be held before September 2001, contraceptive prevalence rate increased to 65% by 2003.) Targets help clarify exactly what needs to be accomplished by when.It represents a commitment and can help orient and motivate project staff and mangers to the tasks at hand.

A natural tension exists between setting targets that are high enough to make project managers and staff stretch to achieve them, and yet low enough to be realistic and achievable.

If they are set unrealistically high and unattainable, confidence and credibility will suffer and may even set in motion perverse incentives to hide or distort the figures. Any information that helps to ground a target setting exercise and ensure its realism is useful. For example, it is useful to establish a baseline, identify historical trends, seek implementing agency views, survey expert opinion about what is possible, review research findings, or identify

benchmarks (i.e., compare what results have been achieved by similar projects with a reputation for high performance).

4. MONITORING PERFORMANCE (COLLECTING DATA)

Once indicators are selected and targets are set, actual data for each indicator is collected at regular intervals. *Implementation monitoring* involves the frequent, on-going recording of data on project operations — e.g.,tracking funds and other inputs, and processes. It involves keeping good financial accounts and field activity records, and frequent checks to assess compliance with workplans and budget. *Results monitoring* involves the periodic collection of data on the project's actual achievement of results — e.g., its short-term outputs, medium-term outcomes, and long-term impacts.This type of monitoring demonstrates whether a project is moving towards its objectives.

Project managers have found it useful to prepare performance monitoring plans to record key aspects of data collection,such as providing definitions for each indicator, source and methods of data collection, frequency/schedule for collection, and assignment of responsibility for collection. Figure 2 illustrates a matrix format that USAID has found useful for recording summary information about data collection plans.

Data collection approaches vary according to levels of the project logframe hierarchy. These common patterns (e.g., typical variations in data collection sources/methods, frequency of collection, and assignment of responsibility) are summarized in Figure 3. As one moves to higher and higher levels of the logframe hierarchy, there is the tendency for data collection efforts to become more expensive, time-consuming, and technically complex.Also, there is a tendency for data collection efforts to be conducted less frequently. The placement of responsibility for data collection also tends to shift from the implementing agency at the lower levels to the donor agency and/or to the partner government at the higher levels.

Data on project inputs, processes, and outputs are generated mostly by project staff and are based on simple reporting systems updated frequently. Data on outcomes are generally collected periodically (e.g.,annually) from low-cost rapid appraisal methods, mini-surveys or consultations with project clients. Measuring impacts usually require conducting expensive sample surveys or relying on already existing data sources such as national surveys, censuses, registration systems, etc. Impact data are usually only collected every few years or at the projects beginning and end (or ex post).

Data collection at the higher levels — especially at the impact level — is often considered beyond the scope of the implementing agency's normal responsibility. Donor agencies will need to make special arrangements with partner country organizations with statistical/data collection expertise for conducting or adding-on to planned surveys. Since

FIGURE 3: CHARACTERISTICS OF PROJECT DATA COLLECTION EFFORTS

(By Logframe Hierarchy Levels)

Type of Indicator	Data Collection Method	Frequency of Data Collection	Organizational Responsibility
Impact Indicators	censuses and surveys, national statistics	multi-year	partner governments, donor agencies
Outcome Indicators	customer surveys, rapid appraisals, consultations	annually	donor agency and implementing agency
Output Indicators	project records	quarterly, biannually	implementing agency
Process Indicators	project records	weekly, monthly	implementing agency
Input Indicators	project records, financial accounts	weekly, monthly	implementing agency

several donor agencies working in the same sector may share needs for similar impact-level data, it would be useful to consider coordinating or jointly supporting these data collection efforts, to avoid duplication of effort and to share costs. Moreover, to ensure valid and reliable data, supporting capacity-building efforts may be called for as well.

At what level should the focus of performance monitoring be placed? Concentrating on just one level of the logframe hierarchy may have unintended, even dysfunctional, consequences. For example, concentrating only on the output level may result in "doing the wrong things well". Concentrating only on higher outcome and impact levels may lead to lack of basic monitoring information about project activities and services, and result in poor implementation. The answer appears to lie in taking as comprehensive and balanced an approach as is possible, within reason/practicality.

Developing a more comprehensive performance monitoring system that recognizes the need for performance information at various levels is least likely to lead to distortions. Moreover, as already discussed, different stakeholder groups and management levels will have varying interests in these levels of results, so satisfying everyone means having a comprehensive system.

5.ANALYZING AND REPORTING PERFORMANCE DATA

Periodic management reviews,analysis and reporting of project performance monitoring data most typically emphasizes effectiveness in achieving targets, by comparing actual results with planned results. However, analysis of performance monitoring data may address a broad variety of issues. For example:

- **Economy** — the relationship between costs and physical inputs (i.e., an organization is economical if it is purchasing inputs as cheaply as possible).
- **Efficiency** — the relationship between costs and outputs (example: cost per kilometer of road built).
- **Productivity** — relationships between inputs and outputs (example: number of demonstrations handled per extension worker).
- **Excellence/quality** — producing high quality outputs (example: percent of units produced that meet technical standards).
- **Equity** — the extent to which needy/disadvantaged sub-populations have equitable access to results (example: percentage of students attending project schools who are female).
- **Customer satisfaction** — how well project outputs correspond to client preferences (example: percent of clients satisfied with health services delivered).
- **Effectiveness** — the extent to which results — outputs, outcomes, or impacts — are being achieved as planned (targeted).
- **Attribution** — the extent to which outcomes and impacts can be attributed to outputs from a particular project.
- **Cost-effectiveness** — the relationship between project costs and results attributable to the project.
- **Sustainability** — the capacity for results to extend beyond the formal life of the project.
- **Relevance** — the continued appropriateness of a project's results to the needs of the target population, the partner country's national development priorities and to the development agency's corporate-level goals.

Periodic reviews of performance data by project management will help alert them to problems or shortcomings vis-a-vis plans and targets, which may lead directly to taking actions or signal the need for more in-depth studies focused on specific performance issues. Routine performance monitoring alone may not be adequate for addressing some of the performance issues listed above(e.g., cost-effectiveness, attribution), which because of their long-term nature and/or complexity may require special in-depth assessments or evaluation studies.

A number of donor agencies have established performance rating systems whereby managers, drawing on data from performance monitoring systems, judge their project's performance by assigning a rating along a scale (e.g., highly satisfactory, satisfactory, unsatisfactory, or highly unsatisfactory), against a number of criteria (e.g., effectiveness, efficiency, relevance, sustainability, etc.). These performance ratings or self-assessments are typically reported to agency headquarters in standard reporting formats at specific times, such as at project completion or in annual reports. A particularly useful characteristic of project performance rating systems is that they enable consistent comparisons and aggregation across the project portfolio.

PERFORMANCE MEASUREMENT AT THE COUNTRY LEVEL

A few donor agencies have developed performance measurement systems for broader country programs — defined as sets of related projects or activities sharing the same development objective within a partner country, usually at a national sector level. USAID pioneered this approach during the mid-1990s,abandoning its previous focus on projects and moving towards more strategic and results-oriented country programming approaches as part of its broader reengineering reforms.

The country program approach is a much more comprehensive and strategic approach to performance management and measurement than the project approach. It focuses on a significant development objective within a country, usually a sector, sub-sector, or a crosscutting objective.Thus, the unit of analysis is not a single project but a country program that typically includes many projects or activities implemented by different donor agencies and partner organizations over a relatively long time period.

Performance measurement frameworks and systems developed at the country program level are thus comprehensive, long-term, multi-activity and multi-site endeavors that usually include many projects and different organizational actors/contributors within a given country sector or sub-sector.

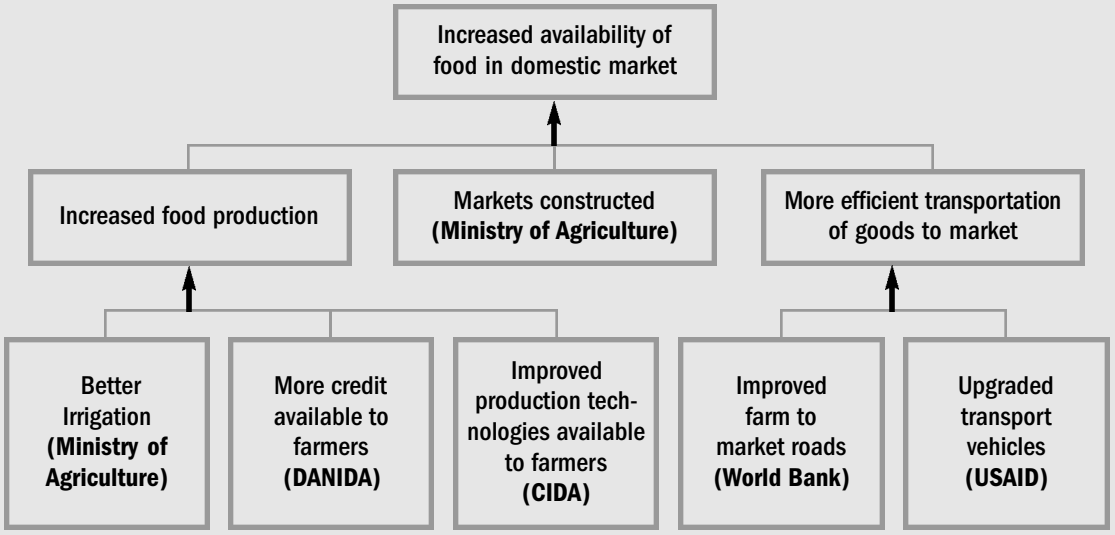
A conceptual tool that is being used by USAID for strategic planning and performance measurement at the country program level is the *results framework.* A results framework is a graphic display of the strategies necessary and sufficient for achieving a significant or strategic development objective in a developing country. The results framework relies on objective tree concepts, and diagrams the logical cause-effect relationships between activity outputs at the bottom, intermediate results or outcomes in the middle, and the strategic development objective at the top. Thus, it embodies the development hypotheses underlying multiple partners harmonized strategies for achieving a shared development objective.

See Figure 4 for a hypothetical example of a results framework. Results frameworks should be developed via collaborative processes involving all donor agencies and other development partners working towards a shared development objective, ideally under the leadership of the partner country government.

Results frameworks are useful as strategic planning and management tools as well as structures for performance measurement. They help identify what program strategies are necessary and sufficient to achieve a significant development objective, and then enable collaborating partners,working in harmony, to sort out their individual responsibilities or contributions to the overall strategy. This can help donor agency operating units to better align (focus and concentrate) their assistance activities into

FIGURE 4: HYPOTHETICAL ILLUSTRATION OF A COUNTRY PROGRAM RESULTS FRAMEWORK

(Development Objective and IntermediateOutcomes)



those program strategies for which they have taken responsibility, rather than just have a diverse portfolio of seemingly unrelated projects. The country development objectives and intervention strategies selected by a unit usually have to be in line with the donor agency's overall corporate goals and areas of comparative advantage.

The framework is also a performance measurement tool — providing a structure for measuring and monitoring progress towards the achievement of those results for which the unit is responsible. Performance data from the monitoring system is used to alert managers when actual results are not meeting targets as planned,indicating the need for adjustments to be made in relevant projects and activities. It may be useful to occasionally supplement more routine reviews of the performance monitoring data with complementary strategic, program-wide evaluations that assess the relative effectiveness and cost-effectiveness of alternative strategies and activities for achieving the development objective.

Whereas the project approach puts equal weight on monitoring all elements of the logframe hierarchy, and may even traditionally have favored implementation monitoring, the country program results framework puts the higher-order development objective and intermediate outcomes at center-stage. It is less concerned with defining the individual project means (inputs/processes) and outputs, and much more concerned with measuring and achieving the higher-level results. The shift from individual projects to programs also implies a different time-frame dimension,freed from the confines of a single project's life cycle. By focusing on country level development objectives and intermediate outcomes, the timeframe now becomes longer-term,outliving the comings and goings of individual project activities.

Individual project activities tend to be less well defined in this approach,allowing for more flexible designs and implementation, rather than rigid "blueprint" approaches. Moreover, in USAID, headquarters no longer approves projects. Instead, authority is delegated to operating units in the field so they can shift course mid-stream if results monitoring information indicates certain activities are not working well. Nevertheless, it is important to be able to link individual project activities, outputs and their associated costs within the broader results frameworks.

This country program level approach puts a premium on partnerships and more collaborative approaches, since achieving a strategic,long-term development objective is clearly dependent on the activities of many development partner actors — e.g., various donor agencies, the NGO community, and of course the partner country government. Some of the tools developed for country program level strategic planning and performance measurement should be particularly well suited to new modes of development assistance based on joint multi-donor/partner sector programs in which investments and activities are harmonized to achieve shared country development objectives.

While this approach holds considerable promise, its actual use thus far may be falling short of its potential. For example in USAID, country operating units have too often tended to develop results frameworks in relative isolation and from their own agency perspectives.

While there is typically some limited participation by their implementing agency partners and immediate stakeholders,the focus is usually on USAID's own programs and strategies, rather than placing equal focus on all relevant partners' programs.

PERFORMANCE MEASUREMENT AT THE AGENCY LEVEL

Largely driven by domestic public pressures and government-wide legislation for annual reporting on agency performance, the donor agencies — like government agencies more generally — are clarifying their overall goals and seeking ways to summarize their achievements vis-a-vis those goals.Measuring and reporting on results at the agency-wide level poses a significant challenge for the development agencies.They face a number of obstacles in attempting to aggregate results, some of which are either unique to or complicated by the nature of development cooperation work (See previous discussion).

Agencies such as USAID have recently developed and issued policy papers or strategic plans that clearly articulate the agency's overall mission and the key development goals or priority areas on which they will concentrate. Usually the agency goals are sector-oriented (e.g., better education, improved health, etc.) although some may be crosscutting special concerns (e.g.,gender equality, partnerships) or internal management efficiency goals. These statements about agency goals serve to articulate to external audiences what the overall aims of the development assistance program are, and provide a framework or structure for gathering and reporting data on overall agency results achieved.

This is viewed as important in an era of declining aid budgets, increasing competition for funding, and growing public skepticism about the effectiveness of development aid. Clarifying agency-level goals has also been useful as an internal management tool for strategic planning — that is, for focusing and concentrating the agency's assistance portfolio and resources within priority goal areas.In the case of USAID, country operating units have been asked to align their country strategic objectives and program activities within the new USAID goal structure.

USAID has further elaborated its agency level goals into several sub-categories, forming a multi-level framework or hierarchy of objectives. Such a multi-level strategic framework can serve to clarify even further what the agency aims to contribute towards achieving and how it intends to contribute. The hierarchies serve as detailed structures for reporting on agency results at several levels. For example, a typical three-level hierarchy structure might include agency goals, sub-goals, and supporting program approaches. USAID has also found it useful to present their strategic framework as graphic diagrams or visual displays, using objective tree concepts. Figure 5 provides an example for the environmental goal from USAID's strategic framework.

Collecting data for agency-wide performance assessments and reporting may take place from several basic sources;(1) from existing international sources/databases that maintain country level statistics on sector development trends, (2) from the project/program performance measurement systems

FIGURE 6: HYPOTHETICAL ILLUSTRATION OF A COUNTRY PROGRAM RESULTS FRAMEWORK

(Development Objective and Intermediate Outcomes)

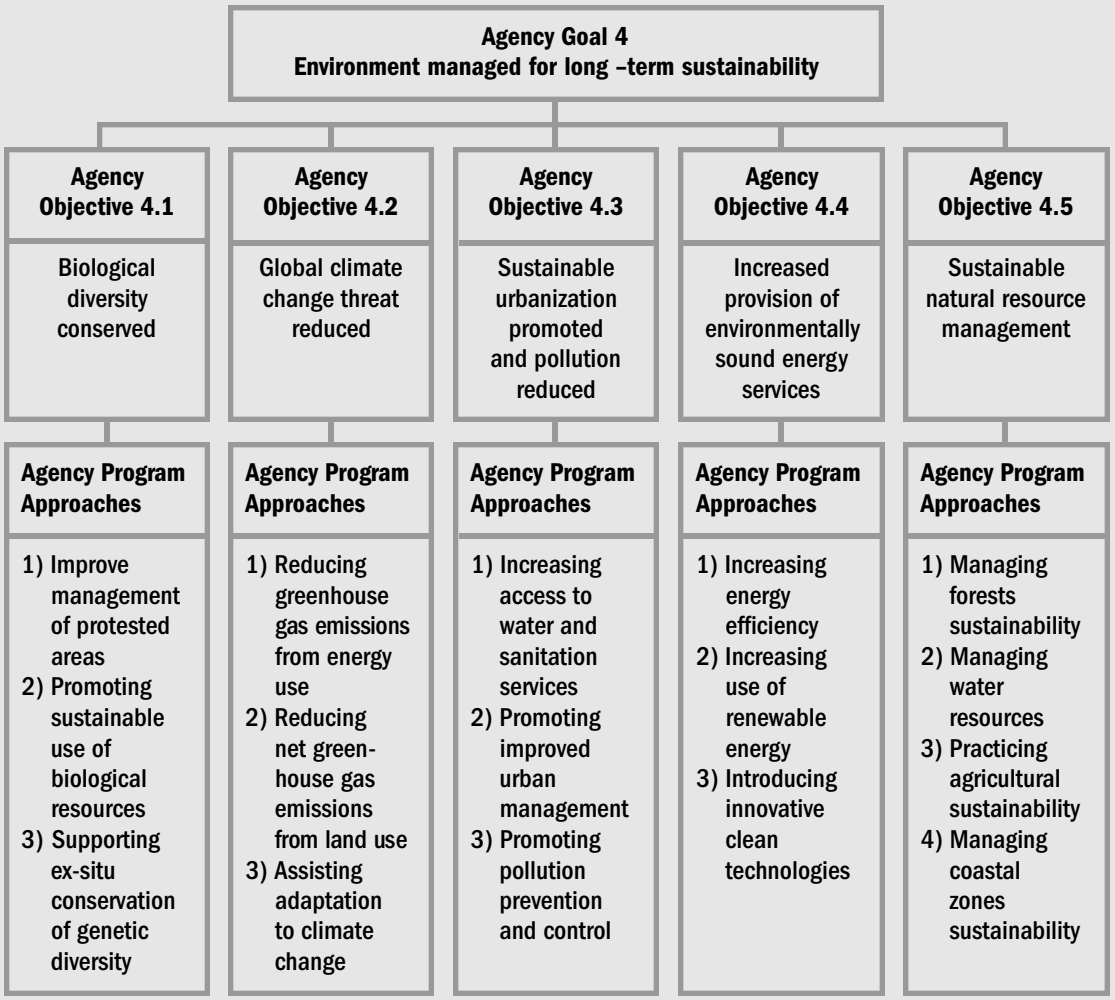


FIGURE 5: USAID'S STRATEGIC FRAMEWORK FOR THE ENVIRONMENTAL GOAL

Core Set of Indicators	Core Set of Indicators
Economic well-being Reducing extreme poverty The proportion of people living in extreme poverty in developing countries should be reduced by at least one-half by 2015	1. Incidence of extreme poverty 2. Poverty gap ratio 3. Poorest fifth's share of national consumption 4. Child malnutrition
Social development Universal primary education There should be universal primary education in all countries by 2015	5. Enrolment in primary education 6. Completion of primary education 7. Adult literacy rate
Gender equality Progress towards gender equality and the empowerment of women should be demonstrated by eliminating gender disparity in primary and secondary education by 2005	8. Gender equality in education 9. Gender equality in adult literacy
Infant and child mortality The death rates for infants and children under the age of five years should be reduced in each developing country by two-thirds the 1990 level by 2015	10. Infant mortality rate 11. Child mortality rate
Maternal mortality The rate of maternal mortality should be reduced by three-fourths between 1990 and 2015	12.Maternal mortality ratio 13.Births attended by skilled health personnel
Reproductive health and population Access should be available through the primary health care system to reproductive health services for all individuals of appropriate ages, no later than the year 2015	14.Contraceptive prevalence rate 15.HIV prevalence rate
Environmental sustainability and regeneration Environment There should be a current national strategy for sustainable development,in the process of implementation, in every country by 2005, so as to ensure that current trends in the loss of environmental resources effectively reversed at both global and national levels by 2015	16.Countries with national environmental plans 17.Access to safe water 18.Intensity of fresh water use 19.Biodiversity: Land area protected 20.Energy efficiency 21.Carbon dioxide emissions

maintained by the agency's country operating units, or (3) from the agency's evaluation reports. In most cases, these data are entered and stored in automated, central agency databases to facilitate agency-wide analysis and reporting.

Computer databases and software programs facilitate data sorting, aggregation, statistical analysis and graphic presentation of results.They can greatly aid the work of analyzing large amounts of performance/results data across project/program portfolios. Results of these agency-wide analyses of aggregate project/program performance and results are usually reported in annual performance reports.

Donor agencies have a number of basic options to consider for aggregating or summing up performance and results achieved at the agency-wide or corporate level. At the two extremes — project outputs and country-level sector statistics — aggregation of

indicator data may be relatively easy. But in the case of outputs, the question "so what?" may be raised. With country level statistics, it is rarely possible to link changes credibly to a single agency's interventions, especially on a year-to-year basis. In the middle are project outcomes, which should be both significant yet have clearer linkages to agency activities than national statistical trends. The problem here is that often there is great diversity in projects' objectives and in their performance measures, so aggregating across standard indicators is often not possible. Some agencies have overcome this by developing rating systems that score a project's success in meeting its objectives and then summing across projects the numbers and percentages that were successful or unsuccessful in achieving outcomes.

These three basic options for aggregating results are discussed below.

1. SELECTING THE PROJECT OUTPUT LEVEL FOR AGENCY-WIDE REPORTING ON RESULTS

Outputs of projects (such as number of units of goods and services delivered or numbers of beneficiaries/clients reached) are relatively easily summed up across similar types of projects,assuming they are comparable. For development agencies with fairly centralized structures and a standard set of project approaches with comparable outputs (goods/services) across country settings, this approach may be feasible. However, reporting at the output level will only be valuable to the extent that the intended external audiences/stakeholders will be impressed with this level of results. If the response is "Is that all we're achieving?", summing and reporting on outputs may be counterproductive in terms of defending the aid program before parliament or the taxpaying public.

2. SELECTING LONG-TERM SECTOR DEVELOPMENT TRENDS FOR AGENCY-WIDE REPORTING OF RESULTS

Another option for reporting on results achieved at the corporate level is to report on long-term social and economic improvements at the country sector and global levels using international statistical datasets,which have some measure of comparability across countries. Advantages of this approach include its appeal in terms of reporting on significant impacts that matter to stakeholders, (e.g., alleviate poverty, reduce infant mortality, achieve universal primary education) and the ready availability of international indicator datasets covering many (not all) of the sector concerns of the development agencies. On the other hand, there are some serious issues with using this approach, especially in the context of reporting on agency performance. Attempting to link and attribute these country-level and global-level socio-economic improvements to the activities of a single donor agency is a wide stretch of the imagination that many will question. Donor agencies using this approach would be advised to adopt the goals and indicators from among those agreed to by the international community (as articulated in the DAC report, *Shaping the 21st Century: The Contribution of Development Co-operation,* 1996). Moreover, their reports might clarify that these results are the consequence of many partners' contributions, and cannot be attributed to individual agencies. (See Figure 6). Another complication is that agency performance reporting is usually required annually, whereas data on country development trends is often only available at intervals of several years apart. Moreover, even if it was available annually, the long-term nature of development improvements at this level means year-to-year changes may not be significant.

3. SELECTING THE PROJECT/PROGRAM OUTCOME LEVEL FOR AGENCY-WIDE REPORTING OF RESULTS.

In between project outputs and macro-statistics, there's the level of intermediate outcomes.

A major advantage is that performance monitoring systems at the project or program level are generally already established and thus data on project/program outcome achievement should be available. A key problem with aggregating project outcomes for agency-wide reporting is the typically great diversity of outcomes and their indicators,especially in decentralized agencies such as USAID. Without standard indicators of project/program outcomes, direct aggregation is not possible. Agencies may deal with this problem in different ways. For example, by developing standard outcome indicators for common "program approaches" (groupings of like projects). However, this approach is only advisable for more centralized agencies with fairly structured program approaches. Another approach for getting around this incomparability problem is to devise rating systems that score a project's success in meeting its objectives. The agency can then aggregate across projects within an objective or program area with statements like "85% of projects or programs aimed at improving child survival have successfully met or exceeded their outcome targets". Issues with this approach may include the extent to which standard criteria for making judgements about scores are applied across projects, and the

reliability of "self-assessment" ratings especially when managers may fear the consequences of poor scores.

None of these options for aggregating results to the agency-wide level appears to be ideal, and the donor agencies face considerable challenges in their current efforts to summarize and report performance at the corporate level. Given this state-of-the-art, agencies may benefit from mixing or balancing the various approaches available. How to best co-ordinate, synthesize and integrate findings from evaluation reports into annual agency performance reports is another issue needing attention.

THE ROLE OF EVALUATION VIS-A-VIS PERFORMANCE MEASUREMENT

Performance measurement and evaluation are generally viewed as two distinct but complementary sources of performance information, both of which are necessary for effective results based management.

In USAID, performance monitoring is defined as a process of collecting and analyzing data to measure the performance of a program, process or activity against expected results (targets), whereas evaluation is defined as a relatively structured analytical effort undertaken selectively to answer specific management questions regarding programs or activities. Some further distinctions often made by agencies follow:

- Performance monitoring reports are self-assessments by project or program managers, whereas evaluations are typically conducted by larger evaluations teams, often comprised of external evaluators that can provide an independent judgement about project/program performance. However, trends towards more participatory forms of evaluation in some agencies may make this less of a distinction.

- Performance monitoring reports are typically *mandatory* for larger projects or programs and thus provide a reasonably complete coverage of the overall portfolio, whereas evaluations are often conducted on a much more *selective* (i.e., occasional, optional) basis for projects/programs of particular interest or concern.

- Performance reports involve relatively straightforward presentations of performance data (e.g.,actual results achieved vis-a-vis expected results, actual expenditures data vis-a-vis budgets, managers' project/program performance self-ratings, etc.) Typically performance data are presented in standard, comparable formats that can be easily entered into databases and analyzed across the portfolio. They are meant to provide consistent types of information covering a broad range of performance issues and results, but without great depth of analysis. Evaluations, on the other hand, usually are less standardized and follow individual scopes of work.Moreover they tend to focus on fewer performance issues but analyze them in greater *depth.*

- Performance monitoring reports focus mostly on *whether* or not results were achieved as planned, whereas evaluations can better explain *why and how* they were achieved or not. In other words, evaluations seek to analyze and understand the projects or program's context and factors influencing performance, both internal (within managers' control) and external (beyond managers' control).

- Routine performance monitoring can serve as an early warning system to alert managers when there are performance shortfalls. However, they do not assess the causes of the shortfalls nor make recommendations for appropriate management actions, as do evaluations.

- Because of timing as well as the need to use more rigorous methods and in-depth analysis, some performance issues, such as long-term impact, attribution, cost-effectiveness, and sustainability, can probably be better addressed by evaluation than by routine performance monitoring reports.

Thus, evaluations and performance measurement/monitoring can be viewed as distinct but complementary functions. Both are management tools. Both are important sources of performance information that together can contribute to management learning and decision-making processes and to external performance reporting requirements.

However, there is some concern that performance measurement and reporting tasks, often required by government-wide law or executive orders, may be "crowding out" evaluations. That is, they may be competing for the same, increasingly scare staff and other resources. For example, a survey of evaluation offices within the U.S. federal government agencies found this to be a concern (See U.S. General Accounting Office, *Program Evaluation: Agencies Challenged by New Demand for Information on*

Program Results, April 1998). Typically, agencies have not been given additional funding to establish their performance management and measurement systems, while overall agency budgets have often been on the decline.

These overall trends and concerns are evident in USAID. For example, since the mid-1990s when reengineering reforms mandated the establishment of performance management and measurement systems, the size of USAID's central evaluation office staff and resources has declined rapidly. Moreover, the number of evaluations conducted by USAID's country operating units also declined sharply, from 489 reports in FY1994 to 70 reports in FY1998.

KEY USES OF PERFORMANCE INFORMATION

In results based management systems,performance information (drawn from both performance measurement and evaluation sources) serve two primary uses or purposes. One use is as an internal management tool for making program improvements; the second is for external accountability reporting.

INTERNAL MANAGEMENT IMPROVEMENT (MANAGING-FOR-RESULTS)

This first intended use of performance information is for continuous feedback to managers about the results they are achieving, so they can then use the information to improve their performance even more. Sometimes discussions of this internal management use are further sub-divided into two related aspects or processes — promoting learning and facilitating decision-making.

(a) Promote Learning. Performance information promotes continuous management learning about what results are being achieved by their projects/programs and why — i.e., what factors are influencing good or poor performance. Improved knowledge is a prerequisite for better decisions.

(b) Facilitate Decision-making. Management's learning in turn facilitates their making appropriate decisions.Continuous performance information about progress towards results and about the factors influencing performance will facilitate good decision-making and timely action. Lessons from experience can help agency managers to continually improve their development assistance projects and programs.

A special type of decision-making that performance information is increasingly being called upon to influence is resource allocations. In USAID, for example, country operating units must submit annual reports (called *Results Reports and Resource Requests*) that ties their request for funds to the results they expect or plan to achieve with those resources. The reports also contain self-assessments by operating units of their actual program performance (extent to which results targets were achieved) over the previous year. Their performance ratings are then ranked across countries and programs, and this information further influences the budget allocation process across countries and programs. On the margin, better performing programs tend to get more funds, and poorer performing programs get less. However, performance remains a comparatively minor factor in these allocation decisions, outweighed by predominant foreign policy criteria and constrained by Congressional earmarks.

EXTERNAL REPORTING (ACCOUNTABILITY-FOR-RESULTS)

The second key use of performance information is to report agency performance to various stakeholder audiences. Donor agencies, like other domestic government agencies, are accountable for achieving and reporting results to the taxpaying public and their elected representatives, and to designated central oversight agencies. Often there are now government-wide legal requirements for reporting results, at certain times and in specific formats, which are being audited by oversight agencies. Moreover, overall agency accountability for results is increasingly being devolved and translated into accountability at lower organizational levels (e.g., operating units, teams, or even individual managers). USAID for example is experimenting with management contracts and personnel appraisal systems that specify what results are to be achieved, when, and by whom. In addition to being responsible to domestic stakeholders, the donor agencies are also accountable to their partner country governments and ultimately to their intended beneficiary groups.

POTENTIAL CONFLICTS AMONG USES

As experience with using performance information grows, the potential for conflict between its two key intended uses is emerging. Managing-for-results implies a shift in focus in from inputs and processes to outputs, and from outputs to even higher outcomes and impacts. Not only is it important to know what results are being achieved at these different levels,but also to understand the

cause-effect linkages between them, e.g., why an activity is successful or not, which approaches work better, and under what conditions or contexts. Emphasis on accountability-for-results, however, may ironically shift focus back down to outputs, which can be more easily attained and attributed to agency activities, and for which data can be easily collected on an annual basis. Managers have relatively greater control over outputs and thus are understandably more comfortable with being held accountable for this lower level of results than for outcomes or impacts over which they have less influence. Moreover, outcomes and especially impacts are longer-term changes that may not show improvements quickly or annually. Since performance reporting is generally conducted annually, this further encourages managers to focus and report on lower-level results that will show changes faster. Furthermore, there is a growing concern among auditors and oversight agencies with attributing results to agency interventions. Since demonstrating attribution becomes increasingly difficult for higher-order outcomes and impacts, this also acts to encourage managers to focus and report at lower results levels. Furthermore, accountability reporting tends to emphasize measuring what is being achieved (and comparing it to pre-set targets), rather than analyzing why or how it is being achieved.In contrast,a management improvement approach is equally concerned with analyzing the context and factors influencing performance, and with drawing lessons for improving performance.

Accountability reporting versus management improvement uses also implies different data collection and analysis approaches. For example, attributing outcomes and impacts to specific agency interventions requires rigorous designs and data collection methods. It also implies extensive attention to data quality, validity and reliability, and to independent verification. On the other hand, a management improvement approach would tend to emphasize more rapid and low-cost data collection/appraisal techniques, with data of sufficient quality for decision-making needs but not necessarily up to standards required for social science research. Moreover, it would favor a self-assessment approach to ensure management's ownership and a first-hand learning experience, and also would encourage more participatory methods and stakeholder involvement. In contrast, an audit/accountability approach might either call for more independent assessments or for a system of spot-checks, reviews and verification of management self-assessments.

These conflicting aims present a dilemma for donor agencies, as it does for other government agencies implementing results based management. Both uses should be kept in mind when establishing performance measurement and evaluation systems. To the extent possible, the systems will need to address both uses and mix or balance data collection and analysis approaches to satisfy both interests. For example, an independent series of central impact evaluations might be undertaken to address auditor's concerns about attribution, while managers in the field might be encouraged to conduct self-assessments employing more rapid appraisal and participatory techniques.

Another potential conflict among performance information uses is more unique to the donor agencies. Donor agencies are accountable not only to domestic stakeholder audiences but also to the partner country stakeholders.

To the extent that donor agencies' performance measurement and reporting systems may vary considerably for one to the next, partner country governments will have to deal with trying to coordinate, compare and make sense of widely different donor agency approaches, frameworks, indicators, data, etc. Harmonization among donor agencies of their performance measurement and reporting systems,particularly at the country level, would lessen the burden on partner country organizations. However, the extent to which this can be accomplished may be limited given the variation in government-wide performance reporting systems that have evolved in different OECD countries that may dictate the donor agencies' approaches.

6 EVALUATION CAPACITY DEVELOPMENT IN THE PEOPLE'S REPUBLIC OF CHINA: Trends and Prospects

Ray Rist, Evaluation Advisor, The World Bank Institute

INTRODUCTION

Evaluation is relatively new in the People's Republic of China (PRC). Indeed, before the early 1980s, it was unknown there. This unfamiliarity with evaluation reflected the orientation of the social science at that time, the virtual absence of any evaluation literature published in Chinese, and the lack of systematic contacts by chinese with those practicing evaluation in other parts of the world. Some activities under way, however, whithin the PRC did come to resemble evaluation, including some policy analysis,economic and management studies, survey research, project completion reviews, and what was broadly termed "experience summarization". But these efforts were not called evaluation, nor were they systematic or focused on the issues now encompassed inour understanding of evaluation.

In the 1950s, the PRC established policy and economic research institutes in national and ministerial, provincial and even some county governments. Over the past nearlyt 50 years, these institutes have undertaken a wide variety of studies using an array of analystic methods. Early work was largely economic, but the later work has branched out into studies of financial systems, social affairs, environmental protection, and sustainable development, to name four.

Although there are few formal studies of the use of material from these institutes in national planning, a general consensus is that the institutes have been helpful. The policy community is aware of their work and sees them increasingly as sources of pertinent data of analysis. The material is largely anecdotal, but some important findings from the institutes have directly affected government decision making, especially after the economic reforms and openings to the world in 1978. As the reform movement in the PRC has grown over the past 20 years, a number of studies in agriculture, fiscal and tax policy, financial policy, foreign trade, and enterprise management have contributed signicantly to the formulation of reform policies.

As the impetus for development grew from the 1950s to now, the PRC built methodological experience in how to study development projects and programs. It focused on the technology, engineering, and cost-effectiveness of development initiatives such as dams, highways, bridges, power plants, construction projects, irrigation, heavy industry, and railroads. Studies of these projects examined cost and quality control, financial benefit, and compliance with development objectives.

As the 1978 reforms took hold, more capital and development assistance came into the PRC. As researchers built contacts elsewhere, comprehensive and regular evaluations were more often undertaken. The past 18 years have seen a growing capability in and understanding of technological and engineering analysis,

environmental impact analysis and modeling, social impact analysis, environmental analysis, sustainability analysis, and implementation studies, though in paractice not all of these are equally conducted in a satisfactory manner. Building skill and experience in these areas for the PRC has been undertaken via a two-way flow of information and expertise — many persons from Cchina have gone elsewhere to study these techniques, and many consultants and experienced researchers in these areas have traveled to China.

THE PRESENT SITUATION: THE MAJOR ACTORS

The driving force for evaluation in China is the massive and sustained surge in national development and economic growth (the annual GDP has increased by 8% per year for the last 8 years). The attention and capability of the country to address evaluation questions comes form this concer with development. Although most evaluation questions comes from this concern with development. Although most evaluation is ex-post project assessment, there is increasingly recognition that evaluation issues are also embedded in all stages of the development project cycle. For this reason, there is a growing awareness within China that the evaluation function is applicable at all stages of the project cycle. There is now interest in linking evaluation to project and program fomulation and implementation, and some ongoing evaluation has already been undertaken, though comprehensively doing so still infrequent. (cf. Valadez and Bamberger, 1994).

Before the reorganization of the government that is now under way, the major national agency in China for project and program evaluation was the State Planning Commission (SPC), whose major functions also include formulating macro regulation policies,preparing what large and medium projects need to be constructed each year all over the country. The SPC organized evaluation studies through its Key Construction Department. The SPC organized evaluation studies directives, established policies and guilines for evaluation, and planed what projects or programs need to be evaluated and when. The major executing agency for the SPC in project appraisal and evaluation was the China International Engineering Consutling Company (CIECC). The CIECC, located in Beijing, has undertaken evaluations, of many large nationally entities and line ministries. The CIECC is systematically studying the policies and methods of evaluation, training staff from mulitple and handbooks on evaluation practices.

By June 1998, the CIECC had completed evaluations of 58 national government projects and also a large-scale evaluation study of the projects constructed in accordance with the eighth state five-year development plan (spanning from 1990 to 1995). It is now assisting the Asian Development Bank (ADB) to evaluate the performance of ADB's whole portfolio in China. These studies can be characterized as generally a combination of performance measurement,financial analysis, cost-benefit assessments, implementatin monitoring, and technical and engineering analysis. The CIECC to date has done few systematic ex-post or impact studies of large-scale national projects.

A recent significant development is thatt, during the current government restructuring, the SPC has been restructured as State Development Planning Commission (SDPC) and the CIECC is transferred from being the executing arm of the SPC into an advisory group of the State Council. Compared to the SPC, the SDPC focusses much more on macro economic regulation and long-term state development programming, and less on annual investment planning. As a matter of fact, the annual investment plan of the SDPC covers only state financed large projects.

In evaluation,however, the SDPC has an important new duty of monitoring and supervising the implementation process of state key projects. The Project Supervision Office of the SDPC, comprising mainly senior officials,is set up exclusively for this obligation. It has not yet been determined whether the SDPC will conduct evaluation as the SPC did before, but it is clear that if it will, only state financed projects will be focused, and those projects involving financing from bilateral and multilateral sources. It will not evaluate any projects wholly financed by local governments with no funding from foreign sources. This means that from now on,local projects will need to be evaluated by local governments themselves.

The new functions of CIECC have not yet been defined. But it can be presumed that its role in evaluation will by no means be weakened,but only strenghtened. In fact, the CIECC has now a better institutional location than before in terms of carrying out evaluation. Operating under the direct supervision of the State Council, it now

enjoys much more political clout and is almost wholly independent of administrative agencies. As a result, it can now access more easily information and data relevant to evaluation and can report directly to the State Council any evaluation findings and recommendations. It is essentially now in the position to monitor the performance of the public sector (cf. Mayne and Goni, 1997). Further, its evaluation conclusions will hopefully be taken more seriously by administrative agencies and project entities. However, there will be conflict of interest which will harm the objectivity of its evaluation work if the CIECC continues to do project appraisal. (Dr. HONG: WHY IS THIS SO? CAN YOU EXPLAIN??)

Another important agency, the State Audit Administration (SAA), which in recent years has made significant progress in the evaluation field, also merits increased attention. Reporting directly to the premier, the SAA conducts audit of financial compliance and funds utilization as required by the State Audit Law. The top management of the SAA attaches great importance to evaluation work. They have had extensive exposure to international practice and have a good understanding of the linkage between evaluation and auditing. The SAA has to date conducted considerable training for their staff on evaluation methodology. They have also finished audits of financial and economic benefits for 21 national key projects that were partially financed by lending from international organizations. The SSA is now preparing guidelines,establishing data banks, and designing indicators for preparing guidelines, and designing indicators for evaluation operations. According to its plan, the SSA will have conducted a formal evaluation in 1998 and 1999. (DR. HONG. DO YOU KNOW HOW MANY EVALUATION THEY PLAN TO DO IN 1998?). The significant advantages of doing evaluation in the SSA is that it promises almost 100 percent independency from administrative agencies, holds the highest political clout, and is close to the top state leader. In its audit reports, the SSA often makes critical but objective recommendations regarding what needs to be improved in investment management. China,will now, with the evaluation capacity of the SSA,be able for the first time to link up the budgeting proces with the evaluation process (cf. Gray, Jenkins, and Segsworth, 1993). The SSA can collect information and data quickly and effectively, and provide good feedback on its findings to the State Council and relevant administrative agencies.

There are evaluation activities in the line ministries as well. Many line ministries have conducted evaluations for a number of projects in their own sectors,as required by either the SPO or their own plans, with the assistance of their affiliated research or designing institutes.

The ministries and their institutes actively participate in (and even sponsor) training seminars and workshop of study tours to foreign countries on evaluation procedures and methodologies. Some of the ministries have developed their own evaluation guidelines and directives, such as the Ministry of communications, Ministry of Railway, Ministry of Agriculture, Ministry of Health, and Ministry of Water Resources.

Admittedly, the evaluation capacity in the provinces is still relatively weak, but a number of provinces have had in recent years good exposure to evaluation. They have conducted self-evaluations as required by SPC and some international organizations, participated in training courses on evaluation methodology organized by central agencies, and some of them even have prepared their own evaluation guidelines. It will be a significant challenge for China in the coming years to effectively design and implement an intergovernmental evaluation system that will allow evaluation work to be linked and coordinated at the various administrative levels (cf. Rieper and Toulemonde, 1997).

One sector significantly influencing the development of evaluation in China is banking. Many banks in China also evaluate projects they themselves have financed. They have explicit requirements regarding what percentage of their projects should be evaluated, when to conduct an evaluation, what indicators to use, and what kinds of data are to be collected. Since many banks have been extensively commercialized during the past few years, their demand for evaluating their own lending operations is tremendous. The Construction Bank of China (CBC),formerly named the People's Construction Bank of China, for example, has developed systematic guidelines and evaluated about 240 projects within its portfolio. The State Development Bank (SDB), as a non-commercial bank,has also issued an array of evaluation policies and guidelines. It also conducts evaluation regularly. In the banks, however, evaluations largely assess financial benefits, management performance, compliance with terms of loans, and risks on loan recall. There is little in the way of evaluation outcomes or impacts,studies of stakeholder satisfaction, and sustainability analysis. The banks evaluate against criteria they have as lending institutions, evaluation policies largely independently of the SPC.

The last key set of actor in evaluation within the PRC is made up of the bilateral and multilateral development organizations, most notably the World Bank, the Asia Development Bank, and the United Kingdom's Overseas Development Administration (ODA). These three organizations, in particular, have demonstrated sustained support for the development of an evaluation capability within China with their support, China has undertaken the following activities to strenghten its evaluation infrastructure:

- It has conducted studies of how other countries in both the developed and developing world have established their national evaluation system and what are the possible options China to establish its own evaluation system;
- It has drafted evaluation guidelines, manuals, and handbooks;
- It has financed senior Chinese officials to travel to both developing and developed Countries to study different evaluation systems;
- It has provided training in evaluation to about one thousand officials, professionals, researchers; and
- It has organized a ministry-level seminar on evaluation.

Many of these activities have been carried out within the past four to six years, and there have already been notable results. First, the CIECC and the SDB have both set internal units to focus exclusively on evaluation. Second, the CBC, similar banks,and a few sector ministries have established evaluation divisions. Third, the SDPC has set up a unit responsible for supervision of project implementation. Fouth, the SAA has begun to incorporate evaluation into its benefit audit function, and Fifth, the Ministry of Finance has established a function of cost-benefit analysis for state key projects. Finally, in the area of knowledge dissemination, some of those trained have now begun training others on their own; and other central agencies, major universities have incoporated on evaluation into their relevant modules, and some graduate students have chosen evaluation as the main subject of their degree dissertations.

WHERE TO GO FROM HERE

China has begun to (a) establish an evaluation culture that legitimates serious inquiry into public sector performance, (b) build the necessary expertise to undertake evaluation studies, (c) cultivate understanding and recognition in the program and policy communities of the utility of evaluation information, (d) find appropriate means to convey evaluation findings to the policy community, and (e) create the development of evaluation as a sustainable effort. China is building the foundation. However, there is no grand edifice in place. In the Chinese governmental structure and administrative hierarchy, several key tasks appear necessary at this time if evaluation is to continue to develop.

First, it is vital for the PRC to have a strong central organization for overall evaluation management and coordination. Such a central organization would carry out independent evaluations for state financed key projects and state development programs and portfolio, report relevant evaluation findings to the State Council, and disseminate evaluation results to relevant stakeholders and even the interested public. Further, given the magnitude and diversity of the national administration, the organization would also set comprehensive national evaluation policies and guidelines so that any national organization unit undertaking an evaluation will have a common understanding of policies and methods necessary. In addition, there is a strong need for alignment among the key organizations involved in evaluation so that they share a common understanding and build a common set of approaches to different evaluation tasks. This coordination has not existed to date, but many in these different sectors see the need to do so.

Second, it is important that the PRC establish formal evaluation units, policies, and guidelines in the ministries and banks. Such units would be independent of their operational departments and adequately staffed and they might be required to evaluate not only their own projects, but also expand the analysis to include programs portfolios. Such evaluations are an important instrument for the ministries and banks to maintain the quality of their portfolios as well as a basis for discussions and decisions at the central level as to whether further evaluations are required.

Third, it seems to be an appropriate time for the provincial-level and local governments to start building their own evaluation capability. The local governments have now almost 100 percent responsibility for their own investment activities, whithout interference from the central authorities. Their investement portfolio are huge. Therefore, the local governments need to establish an evaluation authority operating directly under the governor's supervision. Such an authority would be responsible for evaluating key projects and programs, drafting evaluation policies and guidelines, organizing

evaluation training, and disseminating evaluation information. It appears essential to build a formal evaluation function in the provincial operational departments as well, especially for those that have a large investment portfolio. Some departments have already conducted evaluations and have relevant expertise and experience, but they do not have a specialized evaluation unit and their evaluation capacity needs to be strenghtened considerably.

Fourth, the PRC needs to set up in the SAA an auditing process of the evaluation function so that there can be ongoing oversight and auditing of the evaluations undertaken within the line ministries and evaluation policies and guidelines issued by the central evaluation organizations, the relevant ministries, provinces, and banks. Evaluation units need to be accountable and must adhere to standards of high performance, and the SAA has the right institutional location and legitimate authority to supervise this.

The oversight responsibility would be much like that performed by the U.S. General Accounting Office, The Netherlands Court of Audit, the Swedish National Audit Office, and the Office of the Auditor General in Canada.

Fifth, the PRC needs to develop advanced evaluation methods across these units and organizational entities. Much evaluation work to date has not been performed by such methods. Among the aspects that mostly need to be strenghtened are environment and social impact analysis, sustainability analysis, institutional assessment, program and portofolio evaluation, macroeconomic analysis, and evaluation of social sector projects (cf. Valadez and Bamberger, 1994). Learning new approaches and bringing them into evaluation requires training, building curricula, translation of materials,and mentoring from those more experienced in evaluation.

Sixth, the PRC needs to introduce the monitoring and supervision function into its investment management agencies. Monitoring and supervision provides direct examination and control of the quality of project implementation. Further, it helps collect data that are needed for evaluation. Without good monitoring and supervision, it is difficult to conduct good evaluation. The PRC has begun to recognize the importance of monitoring and supervision. Apart from the SDPC,the State Council has set up a supervision office in the Ministry of Personnel specializing in monitoring and supervising the performance of the managers of large state owned enterprises. It is important to further disseminate such practices to the line ministries and local governments.

Finally, the PRC needs to develop a supply of well-trained evaluators for the many national ministries, provinces, and banks moving into the evaluation arena. The supply at present is far too small to meet demand. Indeed, demand will probably outstrip supply for some considerable time in the future. Building training centers, recruiting talented students and government employees, legitimating the professional status of evaluators, creating means of sharing experiences and developents through journals and conferences, and linking evaluation skills to subject matter expertise will take considerable time and effort. But if China is to build the necessary capacity and expertise, all these efforts need careful, sustained attention.

IN THE SHORT TERM:

One task viewed as an important initial step to the achievement of the evaluation goals set out above is the creation of a special central office, which would have oversight responsibilities in the following areas:

- Drafting of guidelines and procedures for the ministries in the area of evaluation.
- Coordination of training, development of materials, and use of consultants;
- Development and maintenance of data banks for its own work, the sector ministries,provinces, banks, and the SAA;
- Building of evaluator networks and establishment of a national evaluation association;
- Building the institutional capacity in universities and research centers in order to ensure for the future supply of persons trained in evaluation; and
- Dissemination of evaluation findings.

Given the movement in China toward a national evaluation system, the creation of such an office is an essential first step. At present, there is no national coordination or oversight to ensure the best use of resources, prevent duplication, and set high standards.

Since the CIECC is now the advisory organization of the State Council and has relatively good expertise in evaluation, such a central office could be well placed in the CIECC.

IN THE LONGER TERM:

The experiences of developed and developing countries suggest that establishing national evaluation system is a long and difficult process. It takes system coordination and alignment that is not easily achieved, even in small countries with relatively sophisticated public sectors. China is not small, and its public sector is still developing.

China can best approach this undertaking of evaluation capacity building through a number of pilot projects targeted at sector ministries, provinces, and banks with a strong emphasis on understanding what careful evaluation work requires. This select group could test activities related to capacity strenghtening, the building of methofological skills, development of reporting formats, and training. The emphasis would be on piloting different approaches to the institutionalization of evaluation in the public sector. In the beginning, the emphasis would be more on organizational learning than on organizational accountability, there may be a reluctance to test out new approaches-something essential in a pilot phase. As lessons emerge from these pilot efforts, the system can expand ministries and levels of government. Over time, the balance between learning and accountability can be worked through, and an evaluation culture can be developed.

The demand for evaluation in china is great. Demand is being driven by social, technological, and environmental development of the country, which is moving in a high and sustained level. Tracking the changes and assessing the impacts associated with the development will necessitate an evaluation infrastructure, considerable expertise in both the subject matter and evaluation methods, and development of routes for conveying evaluation findings into the policy communities. The second source of demand comes from development assistance given by the bilateral and multilateral lending organizations. The loan portofolio of the PRC now represents billions of dollars. China is the one of largest borrowers from the World Bank and the Asian Development Bank. The emphasis by both the lenders and the Chinese government upon portfolio performance, indicators, and evaluation of outcomes drives still more need for evaluation capability.

Finally, as China moves into the international arena, its policy communities will encounter new evaluation developments,approaches,and forms of utilization that will apply to China. There will be encouragement to bring these developments home and test them out in the Chinese context.

The major concern in the Chinese public policy community is whether the initiatives and efforts described in these pages can be sustained. Building a national evaluation system in China will take decades. The infrastructure has to be built piece by piece across ministries, provinces, and local jurisdictions. The first steps we have described above are being taken, but sustaining this marathon effort will take financial support, technical assistance, the perseverance of officials, and patience.

7 EVOLUTION OF EVALUATION IN THE PEOPLE'S REPUBLIC OF CHINA

Peter C.Darjes,Operations Manager ADB

The process of building performance monitoring and evaluation capacity has been evolving in the PRC since the early 1980s. It was,however, not until 1992, when the China National Audit Office (CNAO), the Ministry of Finance (MOF), the State Planning Commission, and other key central agencies, decided to develop specific proposals for recommendation to the State Council. The World Bank and the Bank have supported this process by providing a series of technical assistance for institutional strengthening and training. The World Bank supported a steering committee composed of representatives of the major public agencies under MOF leadership. This group drafted a proposal with various alternatives for an evaluation system. In 1994, the Bank provided a small-scale TA to develop project performance monitoring and evaluation capacity in the PRC.[1] Under the TA, evaluation methodologies were developed and training was provided to selected Government officials. Similar objectives were pursued through TA to the State Development Bank.[2] While the two TAs focused on selected agencies,[3] the third Bank TA looked at evaluation as a governance function and therefore more as a generic activity.[4] The TA provided a useful overview of the existing institutional arrangements,policies,systems,and practices for monitoring and evaluation in the PRC.The main objectives were to develop a plan of action to build long-term evaluation capacity, improve the institutional arrangements for better coordination among evaluation agencies, and harmonize the evaluation system, methodology, and practices to be used by all evaluation agencies.

I.INSTITUTIONAL SET-UP

1. The evaluation system of the PRC involves many institutions. While CNAO has an oversight function for instituting evaluation at all government levels, its evaluation practice is confined to financial audits. Evaluation units have been established in most key central agencies. For example, the State Development Bank, Ministry of Construction, the China Engineering Consulting Corporation (CIECC), a Government owned consulting firm with close links to SDPC, and others have set up evaluation units. In all these cases,the evaluation units are being located at the highest levels of each organization to ensure independence from operational line responsibilities.

[12] *TA No. 2133-PRC:Developing the Performance Evaluation Capability ofthe People's Bank ofChina,for $100,000, approved on 9 August 1994.*

[13] *TA No. 2664-PRC:Institutional Strengthening ofthe State Development Bank ofChina,for $500,000, approved on 16 October 1996.*

[14] *The agencies covered by the TAs included the China National Audit Office (CNAO); State Planning Commission (SPC); China International Engineering Consulting Corporation (CIECC); State Development Bank (SDB)*

[15] *TA No. 2821-PRC:Strengthening of Evaluation Capacity in the People's Republic of China,for$500,000, approved on 4 July 1997.*

2. SDPC is at the apex of the country's planning and performance evaluation activities. It is directed by the State Council to issue the regulations and to commission evaluation of public sector programs. In the process of a major reorganization in May 1998, SDPC has established an independent evaluation department, the Key Project Inspectorate Office (KPIO). Currently, 85 KPIO staff monitors about 110 projects countrywide. As the number of projects will grow to eventually 700, the staff strength is projected to increase to about 200 in the medium term. SDPC's evaluation work has been primarily focused on ex-post evaluation of projects. Thus far, very little work is being done in monitoring activities and evaluating operational performance during project implementation.

II. FUTURE CHALLENGES

3. While proposals have been made from time to time to devolve evaluation responsibilities to provincial public sector agencies, no formal links have yet been established. This is the challenge to be addressed by SDPC in the future. The significance of performance management and evaluation needs to be assessed in the context of effective governance. It emanates from the delegation of authority to subordinated line agencies that will be held accountable for the accomplishments of tasks assigned to them. Thus, evaluation is a necessary corollary to results-based resource allocation. In the PRC, this thinking is slowly evolving. Currently, the emphasis of performance management is on inspection and control by the central government.

4. At the provincial government level, project performance management and evaluation is all but unknown. Apart from the central agencies mentioned above, there is no authority at the provincial level that has a mandate or an appropriate organization to undertake project performance evaluation. In the process of decentralization, the Provincial Planning Commissions were given project approval authority. As a result, there are now a large number of projects under provincial management. However, even at this level it is being increasingly recognized that there should be a linkage between performance evaluation and project planning, design and approval.

5. During the past two years or so, project performance evaluation has assumed renewed urgency in the PRC. To stimulate domestic demand and to counter deflationary tendencies in the economy the Government has embarked on a massive program of infrastructure investments. At the same time, poor implementation results of large public-sector projects have heightened the awareness of the Government for more rigorous performance management systems. In February 1999, a national conference on project quality was prompted by a series of severe accidents and project failures relating to major investment projects.

6. In the PRC, despite the efforts exerted in the past, evaluation is still in its infancy. Therefore, the human resource requirements are significant and the need for both,qualified staff and training must be met as a matter of priority. Most KPIO staff has undergone some training at various PRC universities and in the context of external assistance. KPIO management has, however, expressed concern that this training is inadequate given the tasks ahead. Previous training activities are either considered to be too short or not commensurate with best practices. As KPIO is mandated with the training of future evaluators and project managers, it must first focus on building up a nucleus of qualified trainers within its folds. Currently, there is no clear notion of the competencies and skills that would be required for project performance management. The formulation of job descriptions will be a key task of the proposed TA.

III. PROPOSED ADB TECHNICAL ASSISTANCE

A. OBJECTIVES

7. The overall objective of the TA is to continue the process of building performance management capacity within the PRC administration. The focus will be on the central government level. This will mean to improve the ability of KPIO to develop its capacity and culture of evaluation. In this regard, the TA will act as a catalyst to enhance KPIO's abilities,skills,and knowledge, procedures and attitudes that will enable it to improve the effectiveness of public sector investments on a sustainable basis. Specifically, the TA will foster project performance management,including diagnosis, monitoring, and impact evaluation. A logical framework for the TA is given in Appendix 1.

B. SCOPE

8. The scope of the TA will follow the recommendations and the action plan prepared under the previous Bank TA. As such, the TA will focus on human resource development of KPIO. Towards this end, the scope will include:

   (1) Forming a partnership with one or more international partner agencies that have a comparable mandate. This would most likely be a government department. It is considered that the most effective way of transferring know-how would be through such arrangements;

   (2) Identifying skills and key competencies required for KPIO staff. The foreign partner will help introduce appropriate procedures for project performance management;

   (3) Based on the above activities, preparation of a training curriculum and a training program;

   (4) Implementation of training activities: Training will center on seminars and workshops to be conducted in the PRC. A limited amount of the TA funds will be provided for overseas training. All training activities will be programmed and carried out in close association with the foreign partner institution;

   (5) Preparation of project performance evaluation guidelines. The guidelines should include the use of teaching methods and tools;

   (6) Evaluation of selected Bank-financed projects as case studies;

   (7) Preparation in outline form of a program of activities for a follow-up phase of capacity building TA, which would focus on spreading project performance management practices to provincial governments, and;

   (8) Provision of office and training equipment.

IV. KEY ISSUES

APPROPRIATE ENTRY POINT

9. Institutionally, there is not a high-level central authority in the Government to provide overall coordination and guidance for all evaluation activities in the country. In the absence of these preconditions, it is difficult for evaluation methodologies to be harmonized and refined, an evaluation culture to be formed, and for the evaluation profession to be properly recognized. KPIO's mandate includes all facets of impact evaluation, including reviews of project design, monitoring of benefit development, bidding procedures, construction schedules, quantity and quality of physical progress as well as costs, prices and fund utilization. KPIO has shown a keen interest in adopting the Bank's Project Performance Management System (PPMS) and has translated the Bank guidelines into Mandarin. KPIO would appear as the most promising entry point for transmitting that system to the PRC evaluation administration.

GETTING THE PROVINCES ON BOARD

10. The world over, the significance of evaluation as an element of governance has evolved in the process of delegating responsibility and accountability from central government planning levels to line agencies and subordinated provincial administrative levels. For KPIO to become effective, the provinces must over time match KPIO's efforts in undertaking project performance management. This should, as a start, include all projects under central Government financing executed by the provinces, and should later cover projects financed by provincial governments alone. The longer-term need to appropriately including the provincial administrative level in performance management is addressed by the proposed scope of the TA.

LONG-TERM PERSPECTIVE

11. Building an effective evaluation capacity in a country is a long-term developmental process and the process may take decades. The process has been evolving in the PRC since the early 1980s. During this relatively brief time span, remarkable progress has been achieved in the PRC. However, given the size and complexity of the country, particularly with its ongoing economic transformation, a major effort would still be required to build an effective project and program performance evaluation system covering all concerned entities in the Government.

TRANSFERRING THE REQUIRED KNOW-HOW

12. Project management will have to consider the optimum source of technical assistance for each type of activity under the project. In this regard, a priority activity for the TA will be the selection of one or more organisations to enter into a partnership arrangement with KPIO. A partnership arrangement is more complex and demanding than a twinning arrangement. Twinning arrangements are often based on an expert-client model. A partnership approach may involve some twinning, but takes a more developmental approach based on an analysis of the current and future skills and needs of both organisations. While a partnership will have some necessary focus on technical assistance, it will largely be intended to assist staff to understand and master a process. Organisations interested in this partnership arrangement will have to show how they intend to develop a relationship based on mutual learning, equality and respect, with sustainability of process as a long-term goal.

The partner agency will have a number of functions, including the provision of technical assistance, professional development support and mentoring for KPIO curriculum development and in-service staff, as well as the provision of work attachments. The partner organization(s) would have to be selected through competitive bidding. This will not be easy given that the prospective organizations would likely be in the public sector.

8

LINKING EVALUATION TO POLICY FORMULATION AND BUDGET PROCESSES: Lessons Learned from Australia

Leonard J. Early, Dept. of Finance and administration, Australian Public Sector

INTRODUCTION

Australia has had a well-developed system of policy evaluation that has been firmly linked to the Budget process. The system developed over many years,but is now in a state of flux – its future is far from assured.The growth and possible future decline of the Australian model can provide some interesting lessons about development and maintenance of evaluation capacity in other countries.

THE HISTORY

There has been a long history of policy evaluation in Australia. A number of policy research bureaux (e.g. the Productivity Commission, the Australian Bureau of Agricultural and Resource Economics and the Bureau of Transport and Communications Economics) have had a major impact on policy formulation in Australia for many decades.[1] In addition, major public enquiries have frequently been commissioned to evaluate economic and other government policies. For example, the Vernon Committee[2] undertook a major review of Australian economic policy in the early 1960s, and the Crawford Study Group[3] evaluated industry policy in the late 1970s. Evaluation was given significant impetus with the election of the Hawke Labor Government in 1983.In the period 1947 to 1983, Labor had been in Government for only three years, when Gough Whitlam was Prime Minister, from 1972 to 1975. The Whitlam government never achieved legitimacy in the eyes of many Australians, and it was dismissed in controversial circumstances by the Governor General. The Hawke Government was led by young Ministers with tertiary qualifications in law, economics and commerce, who were determined not to repeat the mistakes of their predecessors. They intended to stay in office sufficiently long to stamp their philosophy irrevocably on Australian public policy.

The Hawke Government had come to office with a well-articulated public service policy, designed to ensure that Ministers were firmly in control of the public policy agenda. Upon its election, the Government published a White Paper on public sector reform[4] and embarked on a reform program, aspects of which included:

- The introduction of program budgeting with its focus on program objectives and program performance rather than on inputs; and

[1] *Participating in other agencies' evaluation activity became a major part of Department of Finance budget officers' work. By around 1993, the Department of Finance had to limit its involvement to major evaluations.*

[2] *The Hon David Kemp, Minister Assisting the Prime Minister for the Public Service, Ministerial Statement,Reforming the public service to meet the global challenge, Public Service and Merit Protection Commission,1998.*

[3] *This action was unprecedented in the Commonwealth Government in Australia.*

[4] *Within a set of fairly loose controls.*

- Devolution of authority and responsibility to program managers to enable them to achieve the desired objectives.

THE INTRODUCTION OF A FORMAL EVALUATION POLICY

By the middle of the 1980s, it was clear that further reform was required if the Government's objectives were to be met. Program objectives and performance were not well integrated into the Budget process. Program resourcing did not depend on demonstrated program performance and, as a result, program managers did not have the incentive to produce good performance information. Program performance information was poor, and public service managers were not performance-oriented. Program budgeting had not delivered the requisite cultural change in public sector management.

Further public sector reform in 1987 and 1988 included the introduction of a formal evaluation strategy. This strategy had a number of elements:

- Each policy program was evaluated every three to five years.
- Each year, each portfolio submitted an evaluation plan to the Department of Finance spelling out the portfolio's evaluation program over the next three years.
- Each new policy proposal submitted for Budget consideration included a proposal for future evaluation of the initiative.
- Evaluations were normally to be published.

This strategy recognised the full range of evaluation activity. It did not only focus on more formal evaluation processes. It encompassed different forms of evaluation, addressing program appropriateness, effectiveness and efficiency. For example, the strategy included coverage of:

- High level external evaluations (usually on politically more sensitive issues), requiring particular expertise, complex analysis and possibly leading to significant change.
- Evaluations involving both the Department of Finance and the line agency; usually taking less than one year and typically focused on potential program improvements – including of program objectives and performance information.
- Internal evaluations, undertaken entirely within agencies, usually short exercises aimed at improving program management and efficiency.

BEDDING DOWN THE NEW ARRANGEMENTS

In the early years,some agencies regarded the need to produce evaluation plans as an unnecessary bureaucratic requirement. The quality of plans was variable and, in many cases, evaluation planning was not well integrated into strategic planning within agencies. Evaluation capacity in agencies was mixed, as was the quality of evaluations.

Into the 1990s, a separate evaluation unit was established in the Department Of Finance to oversight the evaluation strategy, and to assist agencies with evaluation plans, with advice on evaluation methodology and (in some instances) with evaluations. That unit delivered evaluation training to hundreds of policy analysts and produced evaluation handbooks.

Evaluations were increasingly used and demanded by Ministers in taking Budget decisions. (Indeed, through the 1990s, it became increasingly common for the Expenditure Review Committee – the Cabinet sub-committee responsible for producing the annual Budget - to refuse to consider policy proposals if scheduled evaluations had not been completed.) Evaluation units (of varying sizes and capacities) were established in most of the bigger portfolios.

Most agencies quickly learnt that major Budget proposals were more likely to be accepted if they were supported by sound evaluation (particularly if that evaluation had also involved the Department Of Finance and if that department supported the evaluation results). As a consequence, the Department of Finance was routinely involved in steering committees of major evaluations[5] and evaluation became firmly embedded in policy development and in Budget processes.

[4] *For example:*

- *The Department of Finance and Administration was formed by amalgamation of the Department of Finance and the Department of Administrative Services. On its formation in October 1997,the Department had 2,300 employees. As a result ofa series of reviews (some of which had been commissioned before the Department was formed) activities have been abandoned, re-engineered or contracted out. As a result,employment in the Department has now fallen to around 700;*
- *The provision of IT services to all government agencies is being substantially outsourced to the private sector; and A new agency, Centrelink, has been created and required to compete with the private sector to deliver employment services to the unemployed.*

By 1997, the Australian National Audit Office was concluding that evaluation plans and the quality of evaluations was generally satisfactory (although more needed to be done to ensure evaluation results were used to improve program performance).

However, at the same time, portfolio evaluation plans had exploded in length and complexity, until reports of 120 pages or more were the norm. The Department Of Finance and some agencies were becoming concerned at the cost of this bureaucratic overhead, and beginning to question whether evaluation had become sufficiently embedded in public service culture that the reporting burden could be significantly rationalised.

THE NEW PHILOSOPHY

The conservative Howard coalition government was elected to power in 1996. The new government was highly sceptical of the efficiency and effectiveness of the public service. It has stated[6] that on coming to office it was confronted by:

- A public service with major inflexibilities in its administrative, industrial and financial dealings;
- Costs of many processes that were high because of a culture of compliance with detailed central rules and regulations;
- A Commonwealth bureaucracy that lagged behind other governments and the private sector in its efficiency and in the quality of its management; and
- A cash-based accounting system that made accurate assessments and comparisons of public sector activity impossible.

The Government rapidly set out to address these perceived deficiencies. It dismissed six of the 17 portfolio secretaries (chief executive officers);[7] introduced a Budget which cut $A8 billion (or around two per cent) from Commonwealth Government outlays and introduced a sweeping range of public service reforms.

The government substantially rationalised central controls on agency operations. Agencies became free to set their own wages and terms and conditions of employment; financial controls were substantially rationalised; and most central reporting requirements were rationalised or abandoned.

The winding back of central controls included abolishing the requirement to produce portfolio evaluation plans. Agencies were to report (very briefly) evaluation plans in Portfolio Budget Statements. In practice, the specification of evaluations in Budget documentation has been patchy, with some agencies (most notably the Department of Finance) paying this requirement scant attention.

The evaluation unit in the Department Of Finance has been disbanded and the department no longer sees its role as assisting agencies with evaluation methodology (although it does continue to sit on steering committees of significant evaluations – often at the request of Ministers or Cabinet Committees).

The Government has required managers to review all their activities to see if they should be discontinued, privatised, devolved to another level of government or otherwise improved, e.g. by market testing. This entails a fundamental evaluation of the appropriateness,effectiveness and efficiency of all government operations. There have been some significant changes, but most agencies are yet to commence the process.

In 1999, the Commonwealth government produced its first accrual Budget. The accrual Budget is produced in terms of outputs and outcomes. This provides a link from Budget resourcing through the outputs produced by agencies to the outcomes the government wants to achieve for the Australian people.

Budget documents detail the proposed price, quantity and quality of agency outputs, and link these to desired outcomes. The documents spell out detailed performance targets and performance against those targets is to be reported in annual reports.

[6] *Not surprisingly, the quality ofthis material is highly variable in this year's Budget,with much progress yet to be made.*

[7] *Government activities are classified either as Departmental outputs (the goods and services produced by Departments and agencies on behalf of Government for organisations or individuals) or administered items (controlled by the Government and managed or oversighted by Departments and agencies). Examples of Departmental outputs are policy advice, evaluation activity or the activity of payment of pensions or benefits. The benefits (e.g. unemployment benefits and age pensions) which are determined by Government and that are a statutory right of people who meet the eligibility criteria,are, administered items,as are most financial payments to the States and Territories.*

The concept of program essentially disappears. Thus the requirement to review programs every three to five years effectively lapses (although it has not formally been repealed).

There is a rolling program of price reviews that effectively evaluate performance in delivering Departmental and agency outputs. However, there is a gap, in that there is no longer any formal requirement or process to regularly review administered items.

IMPLICATIONS OF THE NEW PHILOSOPHY

Prior to the election of the Howard Government in 1996, the efficiency of government arrangements had not been subject to much scrutiny or evaluation. The major instrument for improving efficiency had been the efficiency dividend, which had been introduced in 1987. With the efficiency dividend, agency running costs were reduced by one per cent each year. While some agencies found this regime difficult, in fact it was highly ineffective in improving efficiency. Most agencies had little difficulty in obtaining new resources through the new policy process in the annual Budget, and managers had little incentive to re-engineer or restructure their operations to improve efficiency or effectiveness. Most government operations were effectively sheltered from any effective scrutiny or competitive pressure.

Evaluation activity was mainly focused on programs. Few evaluations shone the spotlight on the appropriateness,effectiveness or efficiency of those government operations that delivered the programs.[12]

This situation has changed dramatically. Attention is now firmly focused on the public service and its effectiveness and efficiency. The intention is to subject the vast bulk of government activity to competitive pressure – with agency resourcing based on competitive prices of outputs rather than on the cost of their production.

The additional evaluation of government activities is a welcome development. However, the risk is that this entails a shift in the focus of attention away from the effectiveness of programs. Given that departmental expenses comprise only around 10 to 20 per cent of the Commonwealth Budget, any such development would entail obvious risks.

LESSONS FROM AUSTRALIAN EXPERIENCE

In implementing an evaluation strategy, Australia started with some significant advantages, including:

- A history of evaluation at least of major policy initiatives;
- A strong central statistical unit producing quality data;
- A well educated bureaucracy, with a good leavening of policy and evaluation skills;
- A history of public provision of information on program performance; and
- A sound Budget system, including four year forward estimates of Budget outlays and revenues.

Even so, it took ten years for the evaluation strategy to become firmly embedded in the Australian public service culture.

Mackay has identified factors in the success of Australian evaluation policy as:

- The creation of an explicit, whole-of government evaluation strategy;
- The existence of a powerful central department that has been a committed champion of evaluation;
- Sustained commitment and support the evaluation strategy over a decade; and
- Implementation of related public sector management reforms that have given considerable autonomy to line managers and that emphasise bottom-line results – providing incentives to conduct and use evaluation findings.

I agree with this analysis as far as it goes. However, to the list of factors Mackay identifies, I would add the existence of a strong supportive group of key Ministers throughout the period. (I see the consistent demand of the Expenditure Review Committee (ERC) for evaluations as one of the most important reasons that evaluation flourished in Australia over this period).

I think the Australian experience also illustrates the potential fragility of the evaluation culture.

While current Ministers are, if anything, even more firmly focused on performance than their predecessors, the assumption is that greater management flexibility, in tandem with increased market discipline (including by privatisation) will underpin performance improvement.

Evaluation has a role in the process,but this role is greatly diminished, in that the regular evaluation of the appropriateness, effectiveness and efficiency of the bulk of Government outlays is no longer guaranteed.

A moot point is the extent to which evaluation will continue to flourish in this environment. Developments thus far suggest that the culture is unlikely to change quickly in the bigger Departments and agencies where evaluation is firmly embedded – e.g. in those Departments that have established their own evaluation units. In other organisations, the stimulus to evaluation is likely to come from the Department of Finance and Administration in the annual Budget process. However, given that only around two per cent of Government outlays typically come up for detailed scrutiny in the annual Budget,this implies a significant potential reduction in evaluation activity in some areas.

IMPLICATIONS FOR OTHER COUNTRIES

An interesting question is the extent to which other countries might be able to learn from and emulate the Australian experience, particularly if their starting environment appears less conducive to success.

In my opinion, the Australian experience does provide some useful guidance. It would suggest that introduction of a systematic process of evaluation is more likely to be successful if:

- There is an explicit whole-of-government policy championed by the Budget agency;
- Ministers strongly support the evaluation policy and demand and utilise evaluations in the Budget process; and
- Public sector reform has given line managers sufficient autonomy to manage their resources to achieve specified objectives.

However, there is no miracle cure. Building an evaluation culture requires sustained effort over an extended period.

In my view, there is no doubt that a requirement to report evaluation plans and performance is necessary to get the requisite response in Departments and agencies. However, this reporting requirement should not be allowed to burgeon into a bureaucratic monstrosity. Reporting regimes should be kept simple. Evaluation resources in Departments and agencies should be directed to evaluating, not to completing lengthy reports for central agencies. Is it necessary for countries to "walk before they can run"? In other words, is it necessary to first focus on provision of simpler performance information before moving to implement more sophisticated analytical techniques? I would suggest not – in my view, it would be desirable to move on both fronts simultaneously, albeit the adoption of more sophisticated analytical techniques would need to be at a pace that recognised the availability of skills to support the analysis.

Public sector reform typically comprises a focus on outputs rather than inputs, and greater devolution of authority and responsibility to line managers. Performance objectives and performance information clearly have a role to play in this process. However, the Australian experience suggests program managers are unlikely to produce good performance information unless it is an integral part of the Budget process.

However, the use of performance information in the Budget cannot be automatic. There will be a need of judgement in interpreting the performance information. Evaluation is necessary if performance information is to give up its meaning.

For example, a failure to perform satisfactorily could call for:

- A reduction in Budget allocation (e.g. if the program is proving too expensive and the outcome does not justify the cost);
- An increase in allocation (e.g. if the program is a high priority for the Government and cost is a lesser issue); or
- No change (e.g. if a road program has been delayed by inclement weather).

The important point is that the appropriate Budget response can only be determined in the light of evaluation and analysis of the performance information. Evaluation is necessary for a proper interpretation of the results. One needs to ask questions such as:

- Are results good because targets were set too low?
- Are results poor, despite good performance, e.g. because the agency is poor at producing accurate and timely performance information?

- Has there been goal displacement (e.g. lowering service quality in other areas) in order to achieve the target?

In summary, one needs to understand the underlying reasons for good or bad performance and evaluation is the most effective means of shedding light on such questions. Sound evaluation both requires and increases the value of good performance information. The two go together.

9 PUBLIC PERFORMANCE EVALUATION AND IMPROVEMENT

Marc Holzer, Executive Director,
National Center for Public Productivity, Rutgers University , NJ, USA

PRESSURES FOR PERFORMANCE

"Government must be accountable for results.""Government must do more with less." The 'bottom line' for government is that such statements, voiced repeatedly by policy makers, citizens, and the media, are deceptive in their simplicity.

They raise some very basic questions: What are the desired "results" of government? What does accountability really mean? Faced with competing demands and expectations, and social issues that seem to defy measurement, the "results" of public sector initiatives are difficult, at best, to define. Although we might achieve consensus on broad platitudes, the "devil is in the details." How can we develop an objective approach to accountability, to evaluation and improvement, which will be accepted as impartial, utilized as a means of adding value to public decisions, and applied to positively impact the quality of the lives of citizens? Our argument is that these multiple objectives can only be achieved if policy makers and citizens are actively involved in the process of performance measurement - the evaluative process of holding agencies accountable for their expenditures, actions and promises.

Until recently, government accountability was largely a matter of financial accounting, hence its name. When public funds were appropriated, the key accountability questions focused on how much money was spent and on what specific items: personnel, supplies, travel, communications, etc. Today, the concept of governmental accountability has taken on a much broader meaning. According to Romzek and Dubnick (30) "Accountability is a relationship in which an individual or agency is held to answer for performance that involves some delegation of authority to act. Accountability mechanisms are the means established for determining whether the delegated tasks have been performed in a satisfactory manner or not.

Thus governments,in an effort to demonstrate accountability, need to show their policy officials and service users:

(1) what they are getting for their tax dollars, in terms of services (and sometimes products): public hospitals, roads, airports, libraries, water supplies, etc...

(2) how efficiently and effectively their tax dollars are spent: cost per mile of roads built to certain standards, cost per pupil of students graduating from high school or university.

(3) how such expenditures benefit their lives or the lives of those they care about: eradication or suppression of diseases such as encephalitis or AIDS; high sense of security in a neighborhood; safe and reliable water supplies.

This type of accountability holds government responsible not only for its actions,but also for the results of its actions - the impact of each service on each citizen.

In order to emphasize to policy makers and service users what they are getting for their tax dollars, agencies need to be able to evaluate-measure and report- what they are accomplishing. Measures which are unreported,suppressed or poorly reported are virtually useless in improving performance. Well reported (i.e. well displayed and widely distributed) performance measures emphasize how well a government is doing at meeting its citizens' needs and at living up to its commitment to provide quality services for the taxes people pay. Performance measures can help officials, public managers, and citizens evaluate the impact that various expenditures have on the quality of their lives and the lives of those they care about. Performance measures also enable officials, public managers and citizens to see the outcomes, or results, of specific government services and to place a relative value on each service provided.

Evaluation and measurement of performance has always been implicit in questions of outputs and outcomes: Is crime increasing? Are the streets cleaner? Is the air quality better? How well are our children doing in school? In short, is the program producing as promised; as expected? What are the results? Performance measures provide an opportunity to answer such questions with specific data for evaluation, rather than the more common situation: subjective statements based on vague assessments of efficacy: the neighborhood seems safer. "Litter here is worse than ever." "The schools in this town are very good!"In the absence of data, these statements are merely impressions.

PART I: DATA FOR EVALUATING PERFORMANCE

It is helpful to think of performance measurement as a process or system of measures and procedures, whereby organizations assess how well they are doing compared to previous performance, to other organizations ("benchmarking"), and in terms of how well they are achieving their stated goals and objectives. A well-designed evaluation and performance measurement system should clearly articulate service goals and objectives, define service outputs and outcomes, and specify the expected quality levels for these outputs and outcomes.

Performance measures can be quantitative (average response time) and qualitative (how safe people feel in their parks during the day or at night). Managers can develop performance measurement systems incorporating a variety of performance indicators, such as those defined by the Government Accounting Standards Board(4) in the United States:

Input indicators. These measure the resources expended on a program, such as the amount of money spent or the total number of employee-hours needed to deliver a service.

Capacity indicator. These measure the ability of an organization to deliver services. Capacity measures helps managers evaluate the level of training (how recent, percentage of employees actually trained, level of training, etc.), the state of physical facilities (space, comfort,safety, etc.), the readiness of systems (Y2K compliant, computer networking, memory and speed, etc.).

Output indicators. These report the quantity of products or units of service provided to a service population.They also include "workload" measures that reflect the amount of effort expended to produce a product or provide a service. Examples of output indicators include: number of meals delivered;miles of road paved;number of students passing the high school proficiency test.

Outcome indicators. These measures report the results of programs and services. Outcome indicators possess quantitative and qualitative aspects. Examples of outcome indicators include: the number of individuals employed six months after participation in a job training program or percentage of residents who frequently enjoy using a park.

Efficiency and cost-effectiveness indicators. These measures focus on how a goal is achieved, rather than what was achieved. Specifically, efficiency indicators refer to the ratio of the level of service (tons of refuse collected, number of meals delivered) to the cost, in dollars and labor, of providing the services. They measure the cost per unit of an output or outcome. Examples of efficiency indicators include: cost per meal delivered; cost per ton of garbage collected; cost per pupil educated.

Productivity indicators. These measures,according to Ammons(1), combine the dimensions of efficiency and effectiveness in a single indicator. For example, the number of meals delivered per hour measures the efficiency; the number of meals delivered on time (and warm) measures the

effectiveness. The unit cost (labor-hours) per on time (and warm) delivery measures productivity.

Designing a good performance measurement system may be challenging to public managers not accustomed to measuring or setting performance targets. However, this effort should not be dismissed as too complex or too difficult. Developing a system involves an understanding of what the program is trying to accomplish, who the main users or customers are, and a basic knowledge of the current level of service. The Government Performance Results Act of 1993 in the United States represents an effort to institutionalize performance measurement systems at the federal level(8a). The act requires federal agencies to:

- establish top-level agency goals and objectives, including annual program goals;
- define how they intend to achieve those goals,and;
- demonstrate how they will measure agency and program performance in achieving those goals.

A good evaluation and performance improvement system should include the following seven-step system developed by the National Center for Public Productivity at Rutgers University-Campus at Newark:

1. IDENTIFY THE PROGRAMS TO BE MEASURED

To start with, programs to be measured must be clearly defined. Programs are groupings of routine activities aimed at providing support for specific public services. Groupings of individual activities make up a program. For example, the following activities-street resurfacing; street patching; seal coating; and curb repair-constitute a program that is traditionally called street maintenance. Usually, programs are defined by governments and are generally listed on an organizational chart contained in the operating budget. Programs relate directly to the organizational structure and the managerial areas of responsibility.

Choosing what programs to measure is a matter of judgement. On the one hand, programs should not be too few, so that only a small portion of services are covered or the information collected is insufficient. On the other hand, too much reporting can be excessively costly, or overwhelming and impractical. Generally, performance measurement systems work best when they concentrate on collecting limited but essential information about basic programs that need the most managerial oversight and where accountability reporting is most important.

2. STATE THE PURPOSE AND IDENTIFY THE DESIRED OUTCOMES

Typically, a government ministry, department or agency initiates a strategic planning process to clarify its mission, goals, and objectives. Through this process, the agency can identify the outcomes or results it wants to achieve through its programs. A manager can only measure the performance of a program if its purpose is clearly stated. Preparing a well-articulated statement of purpose is a very important first step. Ideally, a clear mission statement is the starting point. If that is not available, a thorough program description is a good place to begin. For example, GASB (12) offers the following statement of purpose for a public transportation system:

The basic purpose is to provide safe, dependable, convenient,and comfortable transportation services at minimum cost to the citizens,including special clients groups such as the handicapped and the elderly.

3. SELECT MEASURES OR INDICATORS

A good system uses a few selected indicators to measure outcomes and performance. Most government programs that have established performance measurement systems incorporate the indicators described above: input, capacity, output, outcome, efficiency and productivity. For example, in a sanitation department the input measures might be the amount of labor hours; the operating budget; the number of vehicles on the road. Output measures could include tons of refuse collected; number of households served; number of missed collections. Efficiency measures could include labor-hours per ton of garbage collected or household served;dollars spent per 1,000 households; and productivity indicators could include measures such as the cost per mile of clean streets; the cost per household of twice weekly refuse collection.

4. SET STANDARDS FOR PERFORMANCE AND OUTCOMES (TARGETS FOR ACCOMPLISHMENT)

Under this step, public managers should specify the conditions under which program goals and objectives are met. Managers need to determine what service effectiveness and quality means for a particular program and explicitly state how they are going to determine whether the stated terms of

effectiveness and quality have been met. This involves comparing actual program outcomes or results against some agreed-upon standards,including:

- previous performance (the percentage of residents who feel safe in their neighborhoods this year compared to last year);
- performance of similar organizations (the percentage of residents who feel safe in their neighborhood compared to the percentage in a neighboring community);
- performance of the best organizations (the percentage of residents who feel safe in their neighborhoods compared to the percentage of residents who feel safe in their neighborhoods in the recognized "safest communities" in the country);
- pre-set targets, (next year 85 percent of all residents will feel safe in their neighborhoods, and in three years 95 percent will feel safe.)

5. MONITOR RESULTS

Each accomplishment target should be monitored on a continuous basis. Monitoring provides the manager with the results needed to decide whether or not the performance targets have been met. Systematic and periodic monitoring gives the manager an opportunity to keep track of the operation of the program, and take corrective action when necessary. Usually, monitoring will vary depending on the program and target accomplishments. For the most important programs and services, monthly data collection and reporting systems that indicate results will be necessary.

6. PERFORMANCE REPORTING

A good performance measurement system reports program results on a regular basis. The report focuses on what was accomplished and what it cost the public. Reports should be brief, convey information graphically, and include minimal explanatory information. Information should be presented in such a way as to facilitate comparisons over time, comparisons with similar organizations, comparison with best programs nationwide and against pre-set targets.

7. USE OUTCOME AND PERFORMANCE INFORMATION

Information from an effective performance measurement system is regularly used in program planning to reevaluate goals and objectives and to adjust priorities. Another use is in managing for results where outcome information is used to promote continuous improvement in program operations and results. In terms of analysis for action, a well-developed performance measurement and evaluation system will enable managers to spot weaknesses and challenges to program delivery, as well as program strengths and opportunities for improvement.

PART II: PERFORMANCE MANAGEMENT FOR IMPROVEMENT

It is useful to distinguish between performance measurement and performance management for improvement. Performance management for improvement is broader than performance measurement. Performance management for improvement incorporates the results of performance measurement into general management practices. In the U.S. the National Performance Review (25a) defined performance management as "the use of performance measurement information to help set agreed-upon performance goals,allocate and prioritize resources, inform managers to either confirm or change current policy or program directions to meet those goals, and report on the success of meeting those goals."

It is essential that performance measurement be considered an inherent and indispensable part of the management for improvement process. Measurement for measurement's sake alone is insufficient. Managers can have stacks of data at their fingertips, but unless they use that data to improve policy and procedures it is virtually useless. Osborne and Plastrik write in their book "Banishing Bureaucracy"(27) "We have not listed performance measurement as an approach because we do not believe the act of measuring itself has enough power to force fundamental change in public organizations. It is critical competent organizations need it... But some public organizations have measured performance for years, with virtually no impact."

As a means of achieving measurable impacts, we have found the following system to be an effective integration of measurement and improvement factors. This conceptualization, or "roadmap," is based upon our research has to best practices, and in particular the experiences of award-winning Exemplary State and Local (EXSL) programs identified by our National Center for Public Productivity (19). Today, to produce public services,

the best public organizations have developed multiple, reinforcing capacities. In particular, government agencies which have been formally recognized as high achievers, as state-of-the-art:

- apply quality management principles;
- use measurement as a decision making tool;
- work hard to motivate employees;
- adapt new technologies; and
- develop public-private partnerships.

PART III: GETTING POLICY MAKERS AND CITIZENS INVOLVED

Policy-level and citizen involvement in performance measurement may help public managers stay focused on what really matters in their communities. Citizen involvement can increase the impact of performance measurement by encouraging managers to look beyond traditional output measures (that often have little meaning to citizens) and instead focus on quality of life issues and community goals. When you think about it, why do managers care about government performance? They want to improve the quality of service delivery and ultimately the quality of life in the communities they manage.

The overall goal of involving policy makers and citizens in performance measurement is to build lasting processes that involve citizens in assessing municipal performance so that government policies and services reflect the needs of the community. Policy makers, citizens and public managers, together, can establish performance measures that are meaningful to both citizens and managers. So, for example, instead of just calculating the tons of refuse collected, public managers might also ask policy officials and citizens to rate the cleanliness of their streets. Instead of counting the number of health workers in specific neighborhoods, public managers could also ask citizens how well they feel in their own neighborhoods.

The relevance of performance measures increases when managers incorporate citizens' perceptions. A powerful management tool results when public managers combine or compare traditional output measures with outcome measures that reflect citizens' perceptions.Meaningful measures that the average citizen can understand provide citizens with the opportunity to assess and improve government, and influence how government services can be made more responsive to community needs and priorities.

Although citizen participation can ultimately improve the level and quality of municipal service provision by making services more responsive to the needs of citizens,it is often difficult to achieve. While citizen involvement typically results in effective policy and meaningful public sector, (32) active citizenship is often perceived as burdensome, costly and time consuming (25, 33, 34). The traditional,top-down,hierarchical model of public administration limits the role of citizen participation in performance measurement. Performance measurement systems may be easier to design and implement when citizens are excluded and managers are the ones determining what will be measured and how it will be measured. Although easier to implement, such performance measurement systems fall short in terms of measuring what matters most in a community. Citizen involvement increases the social relevance of indicators by combining facts—hard data—with value-how citizens feel.

Tangible benefits can be derived from effective citizen involvement and participation not only in performance measurement, but in public decision making as well. Thomas (33) indicated in "Public Participation in Public Decisions" that public involvement can increase the effectiveness of public managers and the decisions they make. For example:

a) Decision quality may improve as citizens and citizen groups add to the information available for making decisions. That information might prevent repetitions of many ill-advised public decisions.

b) With citizens involved in making decisions, acceptance of the decision may increase, enhancing the likelihood of successful implementation.

c) If citizens assist in service delivery, services may become more effective and more effiscient.

d) As involvement increases,citizen understanding of governmental operations may increase and criticisms of governmental agencies may lessen, improving the plight of the beleaguered bureaucrat.

If we were to revisit the seven-step performance measurement system developed by the National Center for Public Productivity, but this time incorporate citizen involvement in each of the steps, the process would look something like this (italics indicate changes):

1. IDENTIFY THE PROGRAMS TO BE MEASURED

To start with, programs to be measured must be clearly defined by citizens. Determining what programs should be measured is the first step in identifying community priorities. In some communities citizens will identify education and recreation as essential programs. In other communities it might be public safety and public transportation that top the list of citizen priorities. Choosing what programs to measure is a matter of judgement and citizens should have an equal voice in determining what programs are to be measured. Generally, performance measurement systems work best when they concentrate on collecting limited, but essential information, about basic programs that need the most managerial oversight and where accountability to the citizens is most important.

2. STATE THE PURPOSE AND IDENTIFY THE DESIRED OUTCOMES

Typically, a government department or agency, in collaboration with the citizens they serve, initiates a strategic planning process to clarify its mission, goals, and objectives. Through this process, the agency and clients/citizens identify the outcomes, or results, they want to achieve through agency programs. Citizens and managers can only measure the performance of a program if its purpose is clearly stated.

3. SELECT MEASURES OR INDICATORS

A good system uses a few selected indicators to measure outcomes and performance. Most government programs that have established performance measurement systems incorporate the indicators described in the previous section: input, capacity, output, outcome, efficiency and productivity. Citizens and managers should determine the most appropriate measures for each program, rather than rely exclusively on the data collected by individual programs. Good program measurement systems develop citizen surveys to assess citizen satisfaction and perceptions. For example, in a sanitation department the input measures might be the amount of labor hours; the operating budget; the number of vehicles on the road. Output measures could include tons of refuse collected; number of households served; number of missed collections. Outcome measures might include the percentage of citizens who are satisfied with refuse collection; the percentage of citizens who rate the cleanliness of their street as above average.

4. EFFICIENCY MEASURES

Set standards for performance and outcomes (targets for accomplishment).

Under this step, citizens and public managers should specify the conditions under which program goals and objectives are met. Citizens and managers need to determine what service effectiveness and quality means for a particular program and explicitly state how they are going to determine whether the stated terms of effectiveness and quality have been met.

5. MONITOR RESULTS

Each accomplishment target should be monitored on a continuous basis by the management and staff delivering the service and the citizens receiving it. Monitoring provides the citizens and manager with the results needed to decide whether or not the performance targets have been met. Systematic and periodic monitoring gives the mana ger and citizens an opportunity to keep tabs on the operation of the program, and take corrective action when necessary.

6. PERFORMANCE REPORTING

A good performance measurement system reports program results on a regular basis. Reports are more than an internal management tool and are made public. They are shared with elected officials, citizens, the media and other government watchdogs. The report focuses on what was accomplished and what it cost the public. Reports should be brief, convey information graphically, and include minimal explanatory information. Information should be presented in such a way as to facilitate comparisons over time, comparisons with similar organizations, comparison with best programs nationwide and against pre-set targets. Performance data should be presented in a way that is meaningful to citizens so they can understand what is happening in their neighborhood as well as the community as a whole.

7. USE OUTCOME AND PERFORMANCE INFORMATION (ANALYSIS AND ACTION)

Information from an effective performance measurement system is regularly used in program planning to reevaluate goals and objectives and to adjust priorities. Another use is in managing for results, where outcome information is used to promote continuous improvement in program

operations and results. A well-developed performance measurement system will enable citizens and managers to spot weaknesses and challenges to program delivery, as well as program strengths and opportunities for improvement.

FINDING ASSISTANCE

The field of public sector performance evaluation is changing rapidly. Fortunately, state-of-the-art knowledge and best practices are now more readily available. The World Wide Web offers access to dozens and dozens of helpful sites. Of course, all such information must be adapted to specific national and local needs.

At the National Center for Public Productivity at Rutgers University we maintain web pages devoted to the improvement of public performance, as well as links to best practices worldwide. Those pages and links may be accessed at: http://www.newark.rutgers.edu/~ncpp

CONCLUSION

While it may be costly and time-consuming to include citizens or their representatives in the measurement of government performance, ultimately the performance measurement system developed will be more useful and meaningful. The data collected will have an impact on policy and program administration. The performance measurement system, rather than focusing on managerial accomplishments and administrative achievements,will address quality of life issues and community goals and aspirations. Governments will be measuring to improve services, to make government more responsive to the needs of the citizens they serve.

REFERENCES

1. Ammons, David N. *Municipal Benchmarks: Assessing Local performance and Establishing Community Standards.* Sage,1996.
2. Colwell, W. L., & Koletar, J. W. (1984). "Performance Measurement for Criminal Justice: The Federal Bureau of Investigation (FBI) Experience" *Public Productivity Review* VIII (3): pp. 207-224.
3. *Conference Proceedings of the 1993 Managing for Results: Performance Measures in Government.* Austin, Texas, October 27- 29.
4. Epstein, Paul D. (1992). "Measuring the Performance of Public Service," in M. Holzer (ed.), *Public Productivity Handbook.* Marcel Dekker: NY, (pp.161-194).
5. Epstein, Paul D. *Using Performance Measurement in Local Government: A Guide to Improving Decisions, Performance and Accountability.* National Civic League Press,1988.
6. Fischer, R. J. (1994). "An Overview of Performance Measurement." *Public Management* 76: pp. S2-S8.
7. Glover, Mark. (1992). *A Practical Guide for Measuring Program Efficiency and Effectiveness in Local Government,* Tampa, FL: Innovation Group Publication.
8. *Government Accounting Standards Board, Service Effort and Accomplishment Measures.* 1997. Norwalk, CT.

8a. *Government Performance Results Act of 1993.* Washington, D.C.: Government Printing Office. 1993.

9. Halachmi,Arie, and Marc Holzer, eds. "Productivity Measurement and Management." (*Public Productivity Review,* 44) San Francisco, CA: Jossey-Bass,1987.
10. Harvey, J. (1987). "Measuring Productivity in Professional Services." *Public Productivity Review* (44): pp. 29-38.
11. Hatry, H. P., Blair, L., Fisk, D., Greiner, J. M., Hall, J. R. Jr, & Schaenman, P. S. (1992). *How Effective Are Your Community Services?* The Urban Institute and ICMA: Washington,DC.
12. Hatry, H. P., Fountain, J. R., Jr., Sullivan, J. M., & Kremer, L. (1990). *Service Efforts and Accomplishments Reporting: Its Time Has Come.* Governmental Accounting Standards Board (GASB): Library of Congress Catalog Card Number: 90-80879.
13. Hatry, Harry P., et al. *Efficiency Measurement for Local Government Services.* Washington,DC: The Urban Institute,1979.
14. Hatry, Harry P., et. al. *How Effective Are Your Community Services?* Washington,DC: The Urban Institute,1977.
15. Hatry, Harry P., et. al. *Measuring the Effectiveness of Basic Municipal Services.* Washington,DC: The Urban Institute,1974.
16. Hatry, P. & Fisk, D. (1992). "Measuring Productivity in the Public Sector". In M. Holzer (ed.). P*ublic Productivity Handbook.* Marcel Dekker: NY, (pp.139-160).
17. Holzer, M. and Halachmi,A. (1996). "Measurement as A Means of Accountability." *International Journal of Public Administration,* 19 (11&12),(pp. 1921-1944).
18. Holzer, Marc, (ed.). *Public Productivity Handbook.* Marcel Dekker, 1992.

18a. Holzer, Marc. "Focus: Productivity Improvement in New York State—the Science and Art of Capacity Building," in *Managing New York State,* Number 1,1988. Albany, New York: Governor's Office of Employee Relations. The ten steps contained in this article are based, in part, on Holzer, et.al., *Managing for Productivity,* including previous work by Constance Zalk.

19. Holzer, Marc and Kathe Callahan, *Government at Work,* Thousand Oaks,CA: Sage Publications,1998.
20. Holzer, Marc and Vatche Gabrielian. "Performance Measurement and Improvement." Newark, N.J.: National Center for Public Productivity, 1998.
21. Kymlicka, W. and Norman, W. (1995). "Return of the Citizen: A Survey of Recent Work on Citizenship Theory," In Beiner, R. (ed) *Theorizing Citizenship.* Albany,NY: SUNY Press.
22. *Managers Guide for Improving Productivity.* Washington, D.C.: Office of Intergovernmental Personnel Programs, United States Office of Personnel Mangement,1980.
23. McGowan,R. P.,& Poister, T. H. (1985). "Impact of Productivity Measurement Systems on Municipal Performance," *Policy Studies Review,* Vol. 4,No. 3: pp. 532-540.
24. Milakovich, Michael E. "Total Quality Management for Public Service Productivity Improvement." in Marc Holzer, ed., *Public Productivity Handbook,* New York: Marcel Dekker, 1991.
25. Mitchell, J. (1997). "Representation in Government Committees." *Public Administration Review.* March/April, Vol. 57, No. 2, p. 160-167.

25a. *National Performance Review.* Washington, D.C.: Government Printing Office,1997.

26. Osborne, David and Ted Gaebler, 1992. *Reinventing Government.* Reading, MA: Addison-Wesley.

27. Osborne and Plastrik, 1992. *Banishing Bureaucracy*. Reading, MA: Addison Wesley.

28. *Performance Measurement: A Guide for Local Elected Officials.* Washington, D.C.: The Urban Institute,1980.

29. Poister, Theodore H., James C. McDavid and Anne Hoaglun Magoun. *Applied Program Evaluation in Local Government.* Lexington,MA: Lexington Books,1979.

30. Romzek, Barbara S. and Dubnick, Melvin J. "Accountability." *International Encyclopedia of Public Policy and Administration.* Jay Shafritz, Editor-in-Chief. New York: Westview Press, 1998. pp. 6-11.

31. Schacter, H. L. (1997). *Reinventing Government or Reinventing Ourselves: The Role of Citizen Owners In Making a Better Government.* Albany, NY: SUNY Press.

32. Stivers, C. (1990). "The Public Agency as a Polis: Active Citizenship in the Administrative State," *Administration and Society,* 22 (May),86-105.

33. Thomas,J.C. (1995). *Public Participation in Public Decisions.* San Francisco: Jossey-Bass.

PART III

A) Monitoring Evaluation in a Results-Based Management Approach

B) Evaluation Initiatives in Science and Technology

10 EVALUATION PARTNERSHIPS

Niels Dabelstein, Head of Evaluation Secretariat DANIDA, OECD/DAC WP

Donor Government financed aid is expenditure of taxpayers money outside national borders.In addition, aid is often spent through partners or co-operating governments over which the donor government has no formal control. This is a feature distinguishing aid evaluation from evaluation of most other publicly funded programmes. For that reason alone, the accountability function may very well be the most important for aid evaluation. Not the least in periods of *aid fatigue* and pressure on government budgets, continued public support for aid expenditure is contingent upon aid agencies' accountability and their ability to document aid results.

Generally, aid evaluation performs several functions: it provides lessons for managers and staff of ongoing aid programmes thereby leading to improvements in implementation; it establishes information and documents experience that can be used in planning and designing future aid programmes;it is used by managers, strategy- and policy makers for revising existing — and for devising new — aid strategies and policies; and finally aid evaluation delivers information on efficiency and effectiveness of aid programmes, and thus ensures accountability towards politicians and the public.

While these functions apply to both donors and partners, the scope differs substantially.

Donors are primarily interested in evaluation the effectiveness, efficiency and impact of "their" investment. But even the best evaluations face difficulties when it comes to establishing causal linkages between the individual donor's efforts and developmental changes. The individual donor's evaluations, be they of projects, programmes, or sector support, run the risk of ascribing positive development to a limited intervention, when the reality is that development is the result of the synergetic effect of all interventions by the developing countries themselves as well as the support provided by donors, enhanced or hampered by external factors.

Our partners – the developing countries – have a wider scope: they would focus on overall performance of projects,programmes, and sectors, and even the entire economy. Evaluation is thus an aid to resource allocation and prioritisation, to sound management, and to accountability towards politicians and the public.

In short you could say that donors evaluate development aid while developing countries evaluate development.

Evaluation institutions exist in many developing countries, but most have little impact on policy and management decisions due to a number of barriers: Poor demand; lack of a culture of accountability (often related to ethics or corruption); absence of evaluation, accounting, or auditing skills; and lack of feedback mechanisms into decision making processes.The strongest barrier appears to be the lack of demand: Credible evaluation is a function of good governance i.e. demand for accountability more than of institutionalisation of evaluation and/or professional capacity.

How, then, can donors contribute to enhancing evaluation capacity in our partner countries.

Support to building evaluation capacity in developing countries has been on the agenda for several years, and the DAC Working Party on Aid Evaluation has promoted the concept actively, through several regional seminars organised jointly with the regional development banks (Africa in 1990, Asia in 1992, Latin America in 1993, and in Africa again in 1998) and through the members' direct assistance, progress towards establishing credible and capable evaluation functions in developing countries has been slow. In recent years the multilateral institutions have been particularly active, as witnessed by this seminar.

In 1996 the Working Party conducted a survey of donor support for and experiences from evaluation capacity building. The Key findings of the study, which I believe are still valid were:

- Sustainable evaluation institutions need support at the highest policy and management levels and should be able to demonstrate its use to these levels. The design of evaluation systems must also take into account the specific government and administrative culture in the country/organisation.

- Political advocacy and senior management demand should be preconditions for ECB supply activities and must be linked to the issue of governance.A long-term strategy is needed for effective interventions in both cases.

- Sustainable and effective evaluation systems must have a legal foundation or a firm statutory organisational regulation.

- An evaluation unit's independence from line management is important, as is the security of career possibilities for evaluation staff and managers.

- The scopes of national-level performance evaluation and performance auditing systems are moving closer to each other, although the former is likely to be more closely integrated in the planning process, while the latter tend to focus more on accountability to the policy-making level. The choice of approach may, however, depend on other factors, such as political commitment, the legal framework and institutional capabilities.

- Development policy and aid tend to shift from a project/programme to sector/policy focus, setting new demands for host-country evaluation institutions.

- Regional, sectoral and programme/project evaluations become more useful if they are based on a co-ordinated approach linked to a national evaluation system, particularly with respect to methodologies and data needs.

These findings generally apply to both donor and recipient agencies.

While the donors can promote and facilitate evaluation capacity building through collective activities, the main support has been, and will continue to be, channelled through multilateral and bilateral donor agencies. The Study identified several areas where donor agencies could strengten and mainstream ECB work within their agencies and in collaboration with development partners:

- Promote an agency ECB support policy or strategy, particularly in view of new aid forms being introduced,including programme assistance for institution and capacity building as part of good governance initiatives at national and sectoral levels.

- Advocate and stimulate the evaluation issue in country dialogues and sector programme assistance.

- Provide technical advice to operational units responsible for ECB support activities and advise on training facilities and materials on evaluation issues.

- Support the establishment of twinning arrangements between other domestic evaluation institutions and host country institutions.

- Assist in securing consistent evaluation methodologies and terminology in the ECB support activities of the agency.

- Co-ordinate their evaluation programmes with host countries and other donors in order to optimise the use of resources and the constrained capacity of recipient countries' evaluation systems.

- Arrange joint-evaluations with a genuine participatory approach, where the needs of both parties are incorporated from the start and where the capacity building element is taken into account specifically.

I will focus on the very last of these possibilities: Joint-evaluation – or in current lingo: Partnership in evaluation.

In line with the emphasis on development partnership, local ownership, and good governance, the donors should use the joint-evaluation as a vehicle to assist in developing an "evaluation culture". Thus the value of evaluation as a management tool as well as an instrument for shared lesson learning and accountability could be demonstrated. To succeed in this, donors need to co-ordinate evaluations and ultimately let evaluations be co-ordinated by the recipient countries.

Although joint evaluations do take place they are still infrequent, and have primarily concerned donors'jointly financed programmes,or been evaluations of multilateral agencies by groups of donors. Common to most of these evaluations is that the developing countries have played a minor - if any role in their planning and execution. Rarely are the recipient countries involved until an evaluation scheduled by the donor is initiated and mostly the recipient government is involved in providing information to the donor but not in the analyses and final assessment of performance. Evaluation programmes are prepared in response to agency needs for lesson learning and for accountability, and is geared to the planning and programming cycle of the agency, not to the planning cycles of partner countries.

Despite these difficulties the donor community considers it crucial to more fully engage with developing countries in a partnership strategy for evaluation.By jointly analysing opportunities both donor and partner countries enhance their sense of ownership and feel responsible for results. Then, evaluations of development assistance programmes can be effectively carried out with the participation of donor and partner countries, as both would share the interest in accountability for results,lessons learned and improved programme effectiveness.

Evaluations carried out with the participation of partner country representatives are more effective as they generally lead to a fuller use of the findings. In fact, joint-evaluations can generate multiple benefit, including: i) a better understanding of the results;ii) greater acceptance or sense of ownership of the results; iii) enhanced credibility and validity of the findings; iv) increased sense of responsibly and incentives to follow the recommendations formulated by the evaluation;v) possibly improved partner country's evaluation capacity.

DAC Member countries also use joint evaluations to improve partner countries' evaluation capacity. The deeper partner country representatives are involved in evaluation process, the more they are exposed to new evaluation methods and approaches, and the learning process is smoother as a result.

The degree of partnership between partner and donor countries will vary according to the objectives and uses of the evaluation. In some cases, DAC Members have involved partner country representatives in the evaluation process by including them in the evaluation team to define terms of reference (i.e. objectives, questions, scope, methods, uses, etc.). On other occasions, they took part in the evaluation process by assuming different roles such as that of key informant, a member of a focus group, a researcher, an interviewer, a country regional expert, or simply by gathering data.

In most cases, partner country representatives have been involved in joint evaluations to gather and analyse data. Generally the International Finance Institutions (IFIs) have not included partner countries in the planning and design phases of the evaluation cycle, while bilateral donors have done so more often. Some [For example, Denmark, the Netherlands, New Zealand, Switzerland and the European Commission] have involved partner representatives more fully in the evaluation process by including them in the following phases: planning and designing; gathering and analysing data; identifying evaluation findings, conclusions and recommendations.

CONSTRAINTS

A1998 review of the DAC Principles for the Evaluation of Development Assistance already identified some of the difficulties encountered when implementing partnership in evaluation. These included recipients'unwillingness,or lack of interest to participate in joint evaluations; time constraints and higher costs; communication problems and delays; increased complexities and, occasionally, political obstacles. The fact that partner countries have inadequate knowledge of donor countries' evaluation policies may also be one of the possible impediments to partnership in evaluation.

Lack of partner government commitment to the process;discrepancies between the donor and partner country's objectives or reasons for carrying out the evaluation; and insufficient evaluation capacity within the partner country have been identified as

the most important impediments to effective partnership in evaluation.

On the demand side, although many partner countries have created an evaluation unit attached to the planning or finance ministry, they are not yet strongly committed themselves to evaluation activities. Governments are not yet fully results oriented. The need for accountability, participation, transparency, rule of law has only recently been considered necessary for improving governance practices. Moreover, the demand for evaluation comes more often from the donors community (which has limited influence), than from the partner countries' Parliament or taxpayers.

On the supply side, differences in donors' evaluation and management requirements have hindered the development of evaluation capacity of partner countries.

Public sector reforms can open new opportunities to effectively stimulate demand and increase supply of evaluation and performance management systems. SIDA has suggested that, in order to develop evaluation capacity effectively, development agencies could also provide financial and technical resources to conduct programme evaluation separately from donors'projects.This could also create demand for evaluation independently from the donor needs.

THE FUTURE

It is important to keep in mind that to have successful partnership in evaluation partner and donor countries should, at least partially, share the same objectives. This normally calls for joint elaboration of the goals, and thus partner representatives' participation during the programming phases. For these reasons, evaluations should shift their primary objective away from the control and accountability functions, to become more a learning instrument to improve development assistance performance and to learn lessons.

Donor and partner countries' governments should be committed at senior level to a joint evaluation process. At the same time, Parliament,stakeholders, beneficiaries and civil society in both donor and partner countries should be kept informed about the results of joint evaluations as so to enhance the sense of ownership not only of the evaluation findings but also of the development programme.

As participatory processes usually need longer time frame, it is crucial to plan sufficient time to carry out a joint-evaluation. It is necessary to plan during the design phase, the number of evaluators and financial resources needed, bearing in mind that joint evaluations generally involve more people and call for more co-ordination.

When evaluation capacity is judged to be insufficient, it is suggested to carry out some actions to support its development in partner countries simultaneously to the evaluation. Donors' agencies dispose of different tools to promote the development of evaluation capacity, notably "learning by doing" training (e.g. their involvement in the definition of questions, data collection, methods); or technical assistance to personnel, as part of monitoring programme/project, or other kinds of training.

In sum, it is rare that donor and partner counties undertake joint evaluations in full partnership. The most important factors impeding effective joint-evaluations are lack of governments' commitment to partnership in evaluation.Different conceptions of the purpose of evaluations and limited time and resources availability.

I have attempted to point out the advantages of joint-evaluations, to identify the main obstacles or constraints,and indicated a few ways in which the partners in development may enhance evaluation capacity through joint-evaluations.

I hope that during the next 45 minutes we can identify practical ways of moving joint-evaluation a few steps further.

11 EVALUATION TOOLS AND METHODOLOGIES: Case Studies

Jacques Toulemonde, Director, Centre for European Evaluation Expertise

THE INTERNATIONAL MODEL OF PROJECT EVALUATION

Donor institutions have disseminated the practice of evaluation worldwide through a systematic review of the projects that they fund. This has given birth to a "model of project evaluation" that is typified hereafter. The evaluation follows the logical framework of the project,which means that it looks at activities, outputs, outcomes and the problem to be solved. All evaluation criteria are systematically applied: rationale, relevance, effectiveness,efficiency, and sustainability. Unintended effects are also valued.

Donor institutions play the leading role in the evaluation system. They set up the evaluation agenda and the rules of the game. This does not prevent participatory methods to become more and more familiar. Participation typically extends to national authorities responsible for the project,to the project management and to beneficiary groups on the project site. Project evaluation and monitoring go hand in hand. Project managers periodically report on the progress of implementation, which generates monitoring data. As far as these data are properly filed, project evaluators can rely upon a sound basis of field data. Evaluation therefore consists of a quick external review. Additional collection of field data is typically limited to a few interviews and a site visit.

The evaluation system follows the project cycle. It may include three different exercises: ex ante, on-going and expost. These three types of evaluation respectively correspond to the phases of project design, implementation and termination.

The description above is a deliberate caricature. In fact, donor institutions have a more complex practice including joint evaluation with national authorities, country evaluations, thematic evaluations, coordination of donors' evaluations, etc. Is it useful to typify the model of project evaluation in such a simplistic way? For the sake of this presentation, the answer is yes since my purpose is to set up a benchmark that everybody is familiar with.

In order to make the above model workable the project under evaluation must fulfil several conditions. First, it must be reasonably simple. This means that it delivers a single type of output to one target group and that it aims at solving one major problem. Second, the project is supposed to involve one single line of stakeholders,namely the line, which connects headquarters of the donor institution to the recipients on the project site. Third, the project is under the responsibility of an identified management team that produces periodic reports and monitoring data. Finally, the project runs from A to Z with a beginning and an end. Its objectives should be made explicit from the very beginning.

A EUROPEAN MODEL OF POLICY EVALUATION

This paper builds upon a very different evaluation model, which applies to a large policy of the European Union. Through a contrasted view of these two models, I will describe some problems that arise when practice shifts from project evaluation to policy evaluation. I will also propose methodological recommendations for solving these problems. The Cohesion Policy is a major stake for the European Union. It principally aims at developing those regions that lag behind in terms of social and economic wealth. Many other goals are also pursued like reducing unemployment, promoting an environmentally sustainable development, etc. This policy is the second biggest in terms of cost for the European taxpayer. It is implemented through; hundreds of programs are managed and funded in partnership with national and regional authorities.

In this context, the term "program" applies to: (1) a coordinated set of different types of projects (e.g. training courses, advice to enterprises, subsidies to employers,etc.), of which (2) the objectives are clearly stated (e.g. placement of the long-term unemployed) without implementation being defined entirely in advance and, finally (3) which has a limited budget and duration (generally five-years).

Within the framework of the European Cohesion Policy, programs are particularly complex. They typically encompass dozens of different types of activities. Their global objective is like an umbrella for several operational objectives and priorities. In most European regions, programs follow each other with incremental changes every five years.In this case, the concept of program cycle with a beginning and an end is somewhat artificial. For these reasons,I tend to consider that what is under evaluation has many features of a policy and I will use this term hereafter in order to be better understood by non European readers.

The evaluation system of the European Cohesion Policy was created in 1989 and it is now ten years old.The system is one of the most ambitious and comprehensive in the world. It makes evaluation compulsory in every European region at ex ante, interim and ex post stages. It puts much emphasis on independent external evaluators. It establishes an obligation to evaluate that applies to the Union itself, but also to the fifteen member states and to hundreds of regional authorities. The least to say is that all these authorities were not that much familiar with evaluation ten years ago. There was therefore a need to support the development of an evaluation culture across Europe. This was the rationale for launching the MEANS program in the early 90s.

MEANS is a French acronym for "Methods for Evaluating Actions of a Structural Nature" ("Structural" because the policy aims at improving social and economic structures). This program, in which I have been closely involved, allowed for several years of research to be undertaken on evaluation practice in the field of social and economic development. The main activities of the MEANS program were bottom up. It consisted of identifying problems and of designing solutions, not of giving lessons.

The outputs of the MEANS program comprise a body of methodological experiences, ideas and innovations which were presented in seven 50-60-page handbooks. The programme also published methodological reports and quarterly newsletters. It organised training seminars and three European conferences (the most recent, in Seville, was attended by just under 600 practitioners). An effort was made to take full advantage of existing experience and recent innovations, particularly in the following areas: measurement of employment effects, development of systems of indicators and evaluation of macro-economic effects. Through the evaluations initiated at the European level and the work that has gone into building on these experiences, sufficient knowledge has now been acquired to produce a comprehensive publication accessible to a wide public. The result is a collection covering all aspects of the evaluation of socio-economic development policies (European Commission, 1999).

At the end of my long tour into this large European policy, I conclude that it requires an evaluation model that is very different from that of project evaluation. Expressing my conclusion in a provocative way (that would certainly not be endorsed by European authorities),I would say that in the case of the European Cohesion Policy: (1) systematic evaluation along the logframe technique is not workable, (2) systematic evaluation on every sites is not workable, (3) systematic relying upon monitoring data is not workable, (4) systematic following of the policy cycle is not workable.

I develop these points in the next sections,together with a series of recommendations that build on the European practice and that help making policy evaluation workable.

SCREEN THE POLICY WITH A LOGICAL DIAGRAM OF EXPECTED IMPACTS

One of the first steps of any evaluation is to understand the cause-and-effects assumptions that connect the activities and their expected impacts. This is easily done by using the logical framework if the activities and expected impacts are in small numbers. This is not so easy in the case of the European Cohesion Policy. One single evaluation is often supposed to assess thirty or more different kinds of activities and ten or more different kinds of impacts. Such a complex setting is not surprising for those who evaluate national policies. How many logframes should be developed for describing such a policy? Certainly too much for this technique to be workable.

In order to tackle this problem,European authorities have first recommended that evaluators and program designers establish a hierarchy of objectives as represented by an objective tree. This technique consists of a diagram in which the most global objective is placed on the far left, and the most detailed objectives on the far right. Each objective can thus be broken down into several sub-objectives which, in turn,are broken down into objectives on a more and more specific level. In practice, neither the objectives nor the relations between them can be articulated so simply. In almost all cases, a given activity may contribute to achieving several objectives of general interest. This is contradictory with the structure of a "tree" and this calls for something resembling more a "network" or a "web" of objectives and activities.

In order to address this problem, an alternative technique has been proposed in the MEANS collection. It is called: "logical diagram of expected impacts". The diagram sketches the main causal links that connect the activities, situated on the left, and their successive results, outcomes or impacts, situated on the right. Contrary to an objective tree which draws a simplified structure of objectives,the logical diagram of impacts attempts to mapping the causal links between the activities under evaluation and their ultimate objectives.

In the European case, this alternative technique has provided a workable means for clearly describing the logic of a complex policy. In addition, the diagram provides a means to quickly screen the whole policy before deciding to scope some parts in more depth.

SELECT KEY ISSUES FOR IN-DEPTH EVALUATION

Evaluation experience of the European Cohesion Policy shows a wide variety of ways of addressing evaluative questions. Fairly often, the evaluation team is expected to treat ten or more questions within a few months. Other evaluations, by contrast, devote a whole year to the in-depth analysis of a single question.

The first option is in line with the model of project evaluation in which the donor institution tends to ask every possible question. In this case the evaluator is asked a list of questions like: did the various activities produce their intended effects (effectiveness)? Were the effects obtained at reasonable cost (efficiency)? Was it worth trying to obtain this or that effect (relevance)? And what about unintended effects? Since the European Cohesion Policy typically encompasses many activities and many effects, the number of questions is multiplied in such a proportion that evaluators cannot convincingly answer all of them. They often admit that "available data do not allow to assess the impacts",or, even worse, they answer the questions without any evidence basis.

On the contrary, the European practice has shown that sound evaluation works tends to address a limited number of questions. This allows evaluators to derive their conclusions from an in-depth analysis, including collection of field data when necessary.

In the following example, the evaluation steering group considered several questions before deciding to analyze three of them in more depth. The story started in 1997 when public authorities in the French region of Franche-Comté had to evaluate an industrial regeneration program partly funded by the European Union. The evaluation steering group devoted its first two meetings to the identification and selection of evaluative questions. Seven fields of questioning were identified. These were then discussed in order to formulate, for each of them, a realistic and useful evaluative question.

These questions drafted by the steering group were as follows: (1) Which categories of projects have the best impact on the attractiveness of the territory? (2) Has support to enterprise helped to create or maintain jobs that are firmly rooted in the area? (3) Have services for enterprises contributed sufficiently to the diversification of economic activity? (4) Does training serve to raise qualifications or rather to make them more relevant? Which of

these two expected effects is the most necessary? (5) Is the probable impact of higher education and research projects sufficient? (6) What is the effectiveness of the different categories of projects in terms of creation and maintenance of employment? (7) Which categories of projects guarantee or improve equal opportunities between men and women?

An in-depth study of all seven questions would have been too costly. It was therefore decided to limit the number of questions to study in depth to three. Certain questions were eliminated so that there would not be duplication with other evaluations and studies underway in the region. Finally, questions (1), (4) and (5) were judged priorities.

The first type of in-depth analysis applies to a cluster of similar activities (as for questions 2,3,4,5 above). In this case, evaluators can easily implement surveys of beneficiary groups. Primary data obtained in this way provide indications relative to gross effects. The analysis of causality may also be carried out, confounding factors may be identified, deadweight may be estimated, evaluation design may include comparison with non-beneficiary groups. This type of analysis focuses on the success or failure of a given activity through its many different impacts.

Such an in-depth analysis deserves to be undertaken for expensive activities in terms of budgetary appropriations (assuming the stakes involved in decision-making increase in proportion to the budget). It is also relevant for new or innovative activities (considering that they are experiments, which need to be validated by an evaluation). The second type of in-depth analysis focuses on a particular group of recipients, such as the long-term unemployed,creators of new enterprise, small farmers or women (as for question 7 above). The evaluation is performed from the point of view of these recipients. It analyzes their needs, it uses their judgement criteria and it tries to understand their interaction with the policy. In this situation, the evaluation produces conclusions on the comparative impact of different activities that directly address the group under study. It can also show how these different activities reinforce or contradict one another with regards to the group under study (synergy effects). The third type of in-depth analysis concerns an expected impact or a category of impacts (as for questions 1 and 6 above).The question then applies to every activity under evaluation. It focuses on their respective contribution to the production of the desired impact. In this situation,the analysis concludes on the comparative strengths or weaknesses of the different activities in terms of contributing directly or indirectly towards a given objective.

Such analyses deserve to be selected when a given objective is considered as strategic (assuming that the corresponding issue is most likely to gain from an informed debate), or when an objective is controversial (assuming that the corresponding issue is most likely to gain from the clarification of judgement criteria). The notion of thematic evaluation calls to mind that of in-depth evaluation. In the framework of the European Cohesion Policy, some thematic evaluations examine how the same impact (e.g. the competitiveness of SMEs) is obtained by various activities (type 3 of in-depth analysis) or how interventions financed in a specific domain such as Research & Development produce several kinds of impact (type 1 of in-depth analysis).

In the framework of the Cohesion Policy, thematic evaluations have proven to be the most useful ones for policy-makers at European level. This confirms that policy evaluation should focus on a few strategic issues instead of asking all possible questions.

INVOLVE ALL RELEVANT PARTNERS IN AN EVALUATION STEERING GROUP

In those European regions where the culture of public managers was old fashioned, the evaluation practice has first started as a bureaucratic exercise. A typical behavior was to hastily contract out evaluation works at the lowest possible cost (sometimes 0.01% of the expenditures under evaluation) and with the lowest possible ownership. A current practice was to use standard terms of reference, scarcely orientated towards regional particularities, badly suited to the policy-makers expectations and hardly motivating for the managers. As long as the writing up of the terms of reference has been considered an office job, it has been difficult to make them properly adapted to the needs of evaluation users.

This poor state of practice has quickly changed and evaluation becomes a managerial exercise. More time is invested in drafting the terms of reference of exercises that cost up to 0.8% of the expenditures under evaluation. In the best cases, the commissioner of the evaluation officially mandates a steering group from the outset and for the entire duration of the process. This group is no longer a close shop that gathers funding institutions,

managers and operators. It is increasingly open to other partners who hold various stakes (e.g. somebody speaking with the voice of "environment" in a policy that is mainly geared at economic development). The steering group is then given an active role in the preparation of the terms of reference, which implies the organization of one or two or even three meetings before the external evaluator is chosen. This makes it possible to draft precise evaluative questions, to check that the essential (or embarrassing) questions have not been forgotten, to reject questions which are more a matter of auditing or control, to select priority questions, and to specify the judgement criteria and norms. An extreme but salutary choice may be the postponement or cancellation of the evaluation, when the steering group foresees that its results will not be used.

This preliminary stage does not involve additional workload. It is rather an alternative way of organizing the evaluation process.In fact,the first meetings of the steering group provide an opportunity to structure the evaluation process in a clear, fair and useful manner, a task that external evaluators often undertake with great difficulties.

DISTINGUISH POLICY EVALUATION FROM PROJECT EVALUATION

In the framework of the European Cohesion Policy, evaluation teams tend to view complex policies as a series of components, and therefore analyze each of these components separately. Components (or activities) tend to be evaluated in terms of their own success criteria. For example an activity which consists of improving roads is evaluated in terms of the total number of hours saved annually by vehicles using the road. The time saved can be globally related to the public investment so that comparisons can be made with road improvement activities in other similar regions.

The indicator mentioned above is specific to a particular activity (improving roads), just as the approach adopted is that of an evaluation specifically for the road transport sector. Specific evaluations have been performed for all sorts of activities: transport infrastructures,training of the unemployed, rehabilitation of run-down neighborhoods,support to innovation networks, etc. Volume 2 in the MEANS Collection presents numerous examples of activities, with a description of their results and specific impacts and suggestions for appropriate indicators.

A standard procedure for policy evaluation consists of performing specific evaluations of its components, and then of synthesizing the conclusions at the policy level. Evaluation by synthesis has faced serious methodological shortcomings when it has applied to heterogeneous activities. Since the evaluation of each component uses its own indicators,analyses distinct impacts, and applies different criteria, the final synthesis can only be the juxtaposition of partial evaluations. The very notion of synthesis therefore loses its meaning. An evaluation report can be produced in which a long chapter is devoted to each activity, but in which the overall conclusion is nothing more than a few lines with very poor or tenuous contents.

In these circumstances it is impossible to make relevant recommendations for policy-makers. For example, it is not possible to justify recommendations for the reallocation of budgets between activities because there is no argument showing that one is more effective or successful than another. To produce genuine synthetic conclusions, the evaluators should apply common criteria at policy level. This recommendation, which is simple in theory, has proven to be hardly workable in the European model of policy evaluation. In most cases, evaluators tend to use the criterion of budgetary consumption, which has even been termed in Italy as the "capability to spend public funds". The exact opposite of an evaluation culture!

Much attention is presently paid to measuring performance. Performance is often translated into a ratio of actual vs. expected outputs, which is also termed "completion rate". The fact that an activity rapidly achieves its objectives may be an argument in favor of its effectiveness or success, but this argument is necessarily weak. Some activities have very cautious objectives while others have very ambitious ones. When such activities are compared, the conclusions are distorted and the synthesis at policy level is impossible.

Finally, it has become clear that policy evaluations must address questions that are of policy interest. Such questions are not related to the outputs of specific activities, or to their immediate outcomes. Evaluative questions of policy interest are closely related to the ultimate policy goals, i.e. macro-economic or macro-social development. These questions are difficult ones and they require long and careful evaluation studies before being credibly answered. It has proven impossible to deal with such difficult questions for every activity on every site.

Consequently, a good practice is to distinguish two levels of evaluation. Evaluations of specific activities address questions related to immediate outcomes and apply judgement criteria that make sense for managers. These evaluations apply separately to all sites and activities. At the policy level, evaluations address questions related to ultimate outcomes and apply judgement criteria that make sense for policy-makers. Policy evaluation cannot reasonably apply to every activity and every site. On the contrary, they typically involve case studies and field surveys in a limited number of sites.

EVALUATE WITHOUT MONITORING DATA WHEN NECESSARY

The task of evaluators is considerably easier when they can rely upon a set of sound monitoring data related to inputs and outputs. In the case of the European Cohesion Policy, much effort has been made to develop evaluation together with monitoring. After ten years, it must be said that the state of practice is far from perfect. In most sites and for a majority of activities, it can be considered that outputs are properly monitored, both quantitatively and qualitatively.

In a near future, this may also be the case for job creation, an impact that is of strategic interest at policy level. Nevertheless most impacts are unlikely to be subject to systematic monitoring within the next decade. Unfortunately, these impacts give rise to many evaluative questions of policy interest. Here comes the problem of evaluating without monitoring data. Examples from the European practice show this is not impossible.

For instance, this achieved in the case of the British TEC program (TEC for Training and Enterprise Council). This success story of useful evaluation took place in the instance of an absolute lack of monitoring data. The TEC were financed by the British Department of Trade and Industry. They offered SMEs management consultancy, training, and information services. The impact of this programme could not be expressed in jobs created, since it was aimed at enhancing productivity which, in the short term, led to the destruction of jobs. It was therefore necessary to evaluate the programme in terms of other criteria such as productivity, profitability or investments induced. None of these impacts had been monitored.

An initial questionnaire was sent to the 75 operators in the Councils. This questionnaire concerned the consultancy services provided, and the categories of businesses targeted. Individual interviews were then carried out with 20 operators. These interviews were used to describe in greater detail the strategies developed by the operators and to establish a typology of operators. Six operators were then selected for deeper case studies. A sample of 300 beneficiary firms was selected among the "clients" of these six operators. Another sample of 200 non-beneficiary firms was also constituted, in keeping with the matching pairs technique.

A second questionnaire was addressed to the beneficiary firms and to those in the matching group. A statistical comparison between the firms in the two groups provided quantitative indications on the effects of the programme, in terms of turnover, productivity and investments. In order to avoid "biased" statements, the questionnaires comprised numerous and partially redundant questions (triangulated questionnaire).The survey incorporated a search for deadweight effects.

This example shows that evaluation can (and sometimes should) be implemented in the absence of monitoring. It shows that public authorities should strengthen their capacity to undertake both exercises, but that poor monitoring should not be accepted as an excuse for cancelling policy evaluation.

MERGE EX ANTE, ON-GOING AND EX POST EVALUATION

In the evaluation literature, contradictory opinions are expressed on the timing of evaluation. Some authors consider that evaluation is only a retrospective exercise while others also attribute a prospective dimension to it. In the context of European Cohesion Policy, evaluation is clearly defined as an activity, which takes place before, during and after the implementation period. These three phases are related to the policy cycle and consist of ex ante evaluation,interim evaluation and ex post evaluation.

Given the uninterrupted sequence of policy cycles, the three exercises have been increasingly overlapping as the European evaluation system developed. Evaluations performed at different stages and for different cycles tend to collide at the same time. The sequence of three evaluation phases in successive cycles creates overlaps that have to be organized as efficiently as possible to avoid any duplication of work.

The relative continuity of activities from one period to the next makes it possible to use conclusions from the recent past to judge the relevance of the new activities proposed. In the case of the European Cohesion Policy, ex ante evaluations that prepare the 2000-2006 policy cycle had to be undertaken by regional authorities in 1998-1999. By this time, the interim evaluations of the period 1994-1999 were just finished. Some of the ex post evaluation for the period 1989-1993 took place at the same time.

In so far as evaluation must draw conclusions from the experience of preceding activities to improve future activities, an interesting solution is to establish an evaluation plan that covers several years and that more or less merges the three phases. The idea is to identify the different possible evaluations and to establish their date and content in relation to the agenda of policy-makers and to the needs of other evaluation "customers"at various levels.

TRANSFERABLE LESSONS FROM EUROPEAN PRACTICE

Lessons and recommendations that emerge from the European practice can be summarized as follows:

(1) Systematic evaluation along the logframe technique has proven not to be workable, but an alternative technique (the Logical Diagram of Expected Impacts) has been successfully used for describing and quickly screening a complex policy.

(2) Systematic evaluation on every site is not workable.A model of policy evaluation should better involve all relevant partners in a steering group, require them to select key issues for in-depth evaluation and to ask evaluative questions that clearly are of policy interest.

(3) Systematic relying upon monitoring data is not always workable. Policy evaluators can and must work without monitoring data when necessary.

(4) Systematic following of the policy cycle is not workable. Policy evaluation should be organized through evaluation work-plans that merge ex ante, on-going and ex post exercises.

My personal opinion is that these lessons,learnt in the context of a large European policy, are of general value for other large policies in other continents. Are the corresponding recommendations transferable to other policy settings? This is a matter for discussion.

REFERENCES

European Commission (1999) *Evaluating socio-economic programmes,* MEANS Collection,Luxembourg: Office for Official Publications of the European Communities.

12 EVALUATION INITIATIVES IN SCIENCE AND TECHNOLOGY PROGRAMS

Rustam Lalkaka, President,Business & Technology Development Strategies

1. BACKGROUND

Science is intended to understand the world while **technology** is to help change it – hopefully, for the better. From a development standpoint, distinctions between high and low technology are irrelevant. What is important is that the product or service is useful, it is 'friendly' to society and the environment, and it contributes to the overall priorities of economic growth with equity, social justice, basic needs and employment. Further, knowledge at the root of science may be value free, but technology is conditioned by context and culture.

A SHARED HERITAGE

Our knowledge of inner and outer space is changing our understanding of the universe and our place in it. At the same time, we are witnessing the decline of the natural environment and the bio-diversity on which we depend for life itself. While the cold war may have ended, weapons proliferate and struggles continue by the world's poor and its ethnic minorities for a better existence. What then does the coming millennium hold for human civilization? And what can the industrializing countries in particular now do to enhance their national innovation systems, in order to apply technology to the task of accelerating economic growth with fairness and stability?

The historical perspective should be kept in view: For about 2,000 years, from say 300 BC to 1770 AD, countries which are today called 'developing' were at the forefront of applied science and many significant innovations moved from east to west. For the next 200 years, following the industrial revolution, countries today called 'developed' began to pull ahead, technologically and thereby economically and militarily. And over the last 50 years, the pace and pattern of technical change has altered sharply and the industrial countries have out-paced the rest of the world. Nevertheless, a dozen Newly Industrializing Countries now have the technical infrastructure for major innovation; and for **all nations** the advanced technologies – adapted, applied,and absorbed – can help improve their lives.

This paper looks at the main dimensions of the innovation system and indicates some of the special considerations involved in making evaluations of S & T programs. Two case studies illustrate appropriate techniques. Such evaluation has to become integral to the a country's governance, as the taxpayer has the right to ensure that the large sums of money for building the S & T establishment are being spent wisely and with maximum impact.

2. MAIN COMPONENTS OF NATIONAL INNOVATION SYSTEM

For the purpose of reviewing national Monitoring and Evaluation capacity, the innovation system can be considered in terms of six sub-systems:

1. **S & T policy** comprising the setting of priorities and allocation of resources for basic research, advanced science-based technologies, competitiveness in all economic sectors, and for other national priority concerns.

2. **Innovation strategy,** both short and long-term, based on the nation's competitive advantage, the management of research and its commercial utilization, the roles of publicly-funded and corporate research, targeting the special needs of large rural and disadvantaged communities, the cost-effective acquisition of technologies, and the 'deepening' of technical competencies for adaptation, application, dissemination and, progressively, local innovation.

3. **Technical human resources** including science education starting in primary school and its progression through technical universities and continuous learning, vocational training, new disciplines such as the management of technology and technological entrepreneurship development. The computing and communication technologies have a crucial role in developing new ways to eradicate illiteracy.

4. **Technical support services** required to enhance quality, raise productivity, lower production costs, and develop an integrated marketing-research-production system, especially the support services for strengthening technology-based small enterprises.

5. **International cooperation** through strategic alliances,design and production sharing in the supply chain, linkages to donor agencies and development banks, and technical and financial collaboration while safeguarding the intellectual property of the inventor.

Underlying the above sub-systems is the overarching problem of **mobilizing financial resources for S & T** (which is much more than R & D), particularly in poor countries where programs for the survival of its people must itself be the highest priority.

Each country has to design its own national innovation system based on its endowments and constraints, its culture, climate and historical legacy. Table 1 indicates some key knowledge indicators in selected countries. The expenditures on R & D are a proxy of the commitment to innovation but must be interpreted carefully as they do not indicate the productivity, the prioritization and inter-sectoral allocations of research efforts. More relevant today are the proportions of high-tech products in total manufacturing exports and the electronic infrastructure in a nation.

TABLE 1: KNOWLEDGE-RELATED INDICATORS – SELECTED COUNTRIES

	GNP PPP US$ 1997	Military Expend. % GDP 1995	TVs per 1,000 1997	PCs per 1,000 1997	Internet Hosts per 1,000 1997	R&D Personnel per 1 mil. 1985-95	R&D Expend. % GNP 1985-95	High-tech. Exports % mfg exports 1997	Royalty Receipts US $ mil. 1997	License Payments US $ mill. 1997	Patent Appls filed by Residents 1996	Patent Appls filed by Non-Residents 1996
Argentina	10,100	1.7	289	39.2	15.92	671	0.4	15	6	240	n/a	n/a
Brazil	7,430	1.7	316	26.3	9.88	168	0.6	18	32	539	2,655	29,451
China	3,070	2.3	270	6.0	0.16	350	0.5	21	55	543	11,698	41,106
Hong Kong	24,350	n/a	412	230.8	108.20	98	0.3	29	n/a	n/a	41	2,059
Egypt	3,080	5.7	127	7.3	0.33	458	0.5	7	54	365	504	706
India	1,660	2.4	69	2.1	0.11	149	0.8	11	12	150	1,660	6,632
Jordan	3,350	7.7	43	8.7	0.79	106	0.3	26	n/a	n/a	n/a	n/a
Korea	13,430	3.4	341	150.7	37.66	2,636	2.8	39	252	2,413	68,446	45,548
Malaysia	7,730	3.0	166	46.1	18.38	87	0.4	67	0	0	n/a	n/a
Mexico	8,110	1.0	251	37.3	8.75	213	0.4	33	130	501	389	30,305
Poland	6,510	2.3	413	74.4	45.34	1,299	0.7	12	27	175	2,414	24,902
Singapore	29,230	4.7	354	399.5	187.98	2,728	1.1	71	n/a	n/a	215	38,403
S. Africa	15,690	2.2	125	41.6	34.02	938	0.7	n/a	73	258	n/a	n/a
Turkey	6,470	4.0	286	20.7	4.30	261	0.6	9	n/a	n/a	367	19,668
Germany	21,170	n/a	570	255.5	140.58	2,843	2.4	26	3168	4694	56,757	98,338
Japan	24,400	1.0	708	202.4	107.05	6,309	2.9	38	7303	9620	340,861	60,390
USA	29,080	3.8	847	406.7	975.94	3,732	2.5	44	33676	9411	111,883	111,536

Some major considerations in formulating an effective evaluation for the above sub-systems are outlined in Annex-1.These then form the basis for developing the needed evaluation capabilities.

1. S & T POLICY

At the national level, an appropriate strategy for developing countries starting the process of technology transformation could be a cascade of first, importing equipment and know-how, acquiring higher level design and operating experience, to be adapted, improved and applied, and then on to indigenous innovation and exports, both of technology-based goods and know-how itself.

In core technologies a nation (and a company) can make good progress (and money) by using other peoples' innovations and purchasing embodied research in capital equipment, with a balance between payments for intellectual property rights *and* investments in adaptation, improvement and reverse engineering. Consider the case of Japan which increased the number of technical collaborations and fees paid by eight-fold in the period 1960–70; concurrently in this period, it spent nine times more on assimilating and upgrading this technology, to build manufacturing capability for the subsequent export of products and technology.

However, with the galloping pace of technical change all countries need to move more rapidly from investment-led growth to innovation-driven development. This calls for investment in science and technology – one's own or spill-offs from others, the efficient import of know-how and equipment, and foreign investment that brings these in. The building of a knowledge society requires the continuous commitment of civil society and the deployment of significant resources in scientific, educational and social fields.

Technology for national priority concerns

The majority of the world's population still depends on the traditional knowledge of its flora and fauna for food and medicine.This knowledge — and the related bio-diversity — have to be preserved and supplemented by modern technology. Rural folk have a shrewd sense of what helps or hurts them, and governments as well as international development agencies have not built upon their knowledge, with poor results. In many situations, the traditional can be upgraded by blending with the advanced, At the same time, technology for national security, defense preparedness and national harmony has to be the foremost priority.

The evaluations of giant national undertakings such as the Three Gorges Dam in China require inter-disciplinary skills due to their enormous environmental, social and financial implications for the future.

2. INNOVATION STRATEGIES

For the **longer term,** the nation has to move from imitation to innovation, more so as the advanced proprietary technologies become difficult to buy. This requires: Analyses of the strategic implications of global change and competitive advantage for niches in regional and international markets; Realistic assessments of current capabilities and resources, to identify the gaps and take action on technical support services, human resource development, special financing instruments;

Formulation of new legislation providing fiscal incentives for entrepreneurial activity and small business promotion, for protecting intellectual property and the environment; Identification of strategic, generic technologies suited to local endowments, and 'prospecting' for concepts and research results capable of rapid patenting, prototyping and market entry; and Promotion of the culture of informal net-working, information sharing and risk-taking. It may also involve a university linkage, some engineering, production and marketing capabilities, and almost always require a committed entrepreneur or champion to mobilize resources, start the venture, survive, and (with luck!) thrive.

Research management

Perhaps the greatest invention of this century is the invention process at **corporate industrial laboratories,** starting with Edison's in New Jersey. These big establishments of GM, Xerox and IBM were downsized, but are now being rejuvenated to again become the prime movers of innovation. Even so, the R & D budgets of a single large multinational corporation are larger than the total national budgets of say China or Korea or India. Indeed, the research expenditures of the developing countries are only around 5 percent of the whole world. In this situation, these countries have to increase budgets where possible, enter into research collaborations where appropriate, and importantly, improve the productivity of their own research.

For governments in developing countries, the responsibility is to restructure the publicly funded **research and technology organizations,** improving

the facilities and the incentives for researchers, making them more accountable in their performance and more responsive to the needs of business. Institutes in industrializing countries operating for a decade or more should now recover at say, three-quarters of their annual operating expenses through contract research and services. Many now claim to do this, some using creative accounting.

In most industrializing countries, the bulk of the formal R & D is done in publicly-funded laboratories, much less in the technical universities and corporations. Nevertheless, the extent of improvisation and manufacturing improvements taking place on the shop floor are impressive.

A case example is presented in Annex-2 of the **Benchmarking of Research and Technology Organizations,(RTOs)** towards improving their performance. With financial generous support from the Danish and Canadian Governments,The World Technological and Research Organization organized the 'Best Practices Project' to evaluate and draw lessons from 60 RTOs, 20 each in the Americas,Europe-Africa, and Asia. The methodology and results can be of interest to all in the pursuit of improved research performance.

Corporate innovation

Everywhere, managements have to be receptive to new ideas, to listen. Corporations in countries such as Japan and US are adopting unconventional innovative ways to stimulate creativity in their personnel. Further, as researchers better understand the physiological, psychological and biochemical roots of the creative process, the left-right-front-back functions of the brain, they now expect to be able to instill more creativity in more persons

The traditional wisdom is that the product must be based on the needs of the market, but where there was no existing market — as for the now ubiquitous 'Post-It', paper clip, stapler and photo-copier — it takes extraordinary persistence to succeed. In promoting innovation, the large enterprises are mimicking the small, by out sourcing supplies and services and breaking out into small intra-preneurial groups, while the small are acting like the big through alliances and virtual consortiums. The flexibility and creativity of an entrepreneurial techno-venture may lead to more break-through innovations than can be generated by larger sized firms in many sectors.

China has had success in promoting innovation through its SPARK program for rural enterprises and TORCH program for advanced technologies. Now it is investing Rmb 4.8 billion in the 1998-2000 period on the 'Knowledge Innovation Project', with the objective of raising its S & T system to rank among the world's top ten before 2010. At the same time, it has much work ahead in revitalizing its Village and Township Enterprises and re-structuring its State Owned Enterprises.

3. BUILDING TECHNICAL HUMAN RESOURCES

In countries at the leading edge such as Korea 20 percent of tertiary enrollments are in sciences and engineering while the Scandinavian and former soviet Union countries spend up to 8 percent of their GNP on education. Countries with poor technological capabilities enroll and spend less than half as much. In some cases, there is the misallocation of resources resulting in high unemployment among engineering graduates along side high illiteracy rates.

What has changed in this decade is the function of the university: It is no longer teaching alone. It must be restructured to take responsibility for an active role in the more complex field of economic development,covering a portfolio of applied and basic research,consulting and community services,specialized training, and distance learning and tech-based venture formation. While it moves towards becoming an **'entrepreneurial university',** the corporation moves towards becoming a **'learning enterprise'.**

Within the research universities, the Media Laboratory at MIT, Robotics Institute at Carnegie-Mellon,and the electrical engineering department at Stanford have been the creators of continuous innovations. Some such as artificial intelligence, expert systems, fuzzy logic and biochemical computers will become the industries of the 21st century.

Also being changed the world over is the education curricula. New courses being introduced are on the **management of technology,** which provides knowledge on key issues at the interfaces of science, engineering, business and civil society. Young and old are being exposed to **entrepreneurship development,** which seeks to transform the nascent entrepreneur into a successful enterprise-owner. And every person needs to be proficient in the **English language and computing,** because without these you cannot work, play, or participate in the expansion of Internet and e-commerce.

Interestingly, high school students in the US fare poorly in mathematics and sciences,but they seem

to acquire the questioning attitudes and computing skills, which produce good results at the university level. The research universities have become seedbeds for innovations, with the University of California system earning $ 61 million through royalties and its 528 patents in 1997, while Stanford and M.I.T. each created 15 start-up companies.

4.TECHNICAL SUPPORT SERVICES

These cover the technical infrastructure of institutions and capability to support the innovation system. They include productivity centers, metrology and total quality management, indigenous consultancy organizations and services to strengthen small enterprises, on which large proportions of employment, incomes and exports depend.

Business development services

While there is consensus that small enterprises will be the prime creators of employment and growth in the future, these ventures need special help. Various financial and non-financial mechanisms are converging in a synergistic system. There is a current debate on the outreach, impact and financial sustainability of alternative means to strengthen new ventures. Such services have to be initially subsidized,which the developing country governments can ill afford and donors usually resist (although such support receives major state funding in OECD countries).

The ability of a BDS (such as an incubator) to replace the resources it consumes and become *financially sustainable* can be shown by an analysis of the flow of funds in and out of the system over at least 5 years. *Sustainability* implies the ability to continue achieving positive outcomes and the durability of the benefits achieved. *Effectiveness* can be expressed in terms of all the benefits derived at the whole system in relation to the use of all resources and the overall satisfaction of those involved. *Outreach* depends on the replicability of the embodied concept and the means of reaching larger numbers of enterprises. The metrics and criteria of assessing BDS performance require common understandings by donors and governments as well as by the businesses they serve.

Starting in the 1980s China has built impressive networks of tech parks and incubators. Their performance has now to be assessed and enhanced. **Annex 3** presents a case example on **Evaluating the Performance of Technology Business Incubators in Brazil.** A similar approach to evaluating the effectiveness of China's large public investment in facilities such as technology parks and business incubators would point to interesting lessons for future developments in China as well as in other developing countries.

5. INTERNATIONAL COOPERATION

In a world shrinking rapidly with the advent of instantaneous, low cost computing and communications, the S & T system — and those who operate and evaluate it — have to think and act globally.

A major effort to promote such international cooperation and strengthen endogenous technical capacities was the UN Fund for Science & Technology for Development. Independent evaluations of UNFSTD-supported technical cooperation programs in China (eg the National Remote Sensing Center and the Beijing Center for Computer Software and Research) showed that with catalytic funding significant capacities were built.The main lessons: the success depended on the commitment and discipline of the Chinese institutuions and on vigorous monitoring of progress.

International cooperation requires a renewed, continuing dialogue between the poor and the rich, without arrogance, with the purpose of finding joint solutions and sharing the benefits of human knowledge fairly.

6. EVALUATION CAPACITY

In developing national capacity for science and technology related programs, their special characteristics have to be noted.

Whether it is in the speed of computing and communications or the unraveling of the human genome, technologies are advancing at exponential rates. While the 19th century ushered in the telephone, electricity and automobile, this century has brought us near-costless and ubiquitous information flows.The evaluators of such systems have to be abreast of current practices and coming trends. They have to be sensitive to the political, cultural, ethical and environmental consequences, the possible impacts and outcomes of the programs being evaluated.

These call for inter-disciplinary teams of professionals with objectivity, impartiality and utmost honesty. As S & T programs typically have long time horizons and high investments,the judgements made in the

evaluation process also have far-reaching consequences and potentials for high costs as well as benefits.

EVALUATION OF WORLD BANK PROJECTS

The World Bank has been recently involved in loans for industrial technology development (ITD) projects in China, India, Brazil, Korea and Turkey. As these are significant loans they justify serious efforts on loan appraisals, monitoring progress, evaluating results and follow-up. In turn,the national and Bank M and E capacity has to be strengthened.

Evaluation lessons on policy lending for ITD projects indicate the characteristics of successful projects as:

- Competitive environment and demand-led;
- Private sector participation and use of strong specialized institutions and financial intermediaries; Need for integrated approaches;
- Unhindered access to international sources of technology and protection of intellectual property; Priority of financing for tech-based small and medium enterprises as well as for non R, D & E support institutions;
- Further, the Bank should support a broad range of lower-income developing countries, not only the advanced ones.

Measuring performance requires agreement on methods and data collection among those being measured and those wanting the results.The tasks are more complex in science and technology projects as the consequences cross national boundaries and there is seldom a consensus on the interpretation , whether it relates to global warming or biologically-engineered produce. The donors have the additional problems of being responsible to their own taxpayers and responsive to the real needs of aid beneficiaries.

REFERENCES

Golden, W. T. Ed. *Science and Technology Advice to the President,* Congress and Judiciary, Pergamon Press,1988

Goldman,M. and Ergas,H., *Technological Institutions and policies: Their role in developing technological capability in industry,* World Bank Technical Paper No. 383,19997.

Lalkaka, R. *Applying technology for enhancing competitiveness, in Human Resource Needs for Change in R & D Institutes,* WAITRO-NISTADS,New Delhi,1996.

Lalkaka, R. and Wu Ming Yu, *Managing Science Policy and Technology Acquisition: Strategies for China and a Changing World,* Tycooly, Dublin and UN Fund for Science & Technology, New York,1984.

Lalkaka, *Manual on Planning and Operation of Technology Business Incubators,* UNESCO, Paris,1999,under publication.

Mackay, K., *Evaluation capacity development,* World Bank, ECD Working Paper No. 6,1997.

Nelson, R. R. Ed. *National Innovation Systems,* Oxford University press,New York,1993.

WAITRO, *Best Practices for the Management of Research and Technology Organizations,* Special Report,1999.

World Bank, *Knowledge for Development,* Oxford University Press, 1998/99.

World Bank, *World Development Indicators,* Washington DC,1999.

UNDP. *Human Development Report, 1999,* Oxford University Press, New York.

UNDP-SIDA, *Measuring and managing results: lessons for development cooperation,* 1997.

ANNEX 1
Issues in evaluations of innovation sub-systems

CHECKLIST FOR SUB-SYSTEM-1: S & T POLICY

1. What promotional measures are being taken to build a national consensus on the role of science-based development in strengthening the economy and improving the lives of its people?
2. What are the country's short, intermediate and long term goals to which the innovation system must be organized to contribute?
3. How can the processes of formulating the national S & T strategy be made more effective through better involvement of all stake-holders in rural and urban communities, civil society and defence establishment, particularly women, youth, ethnic minorities and other disadvantaged groups?
4. How are the legislative bodies that formulate the policy instruments and regulations being prepared for the content and implications of their actions?
5. What are the present arrangements for providing advice on S & T issues to the executive and legislative authorities in government?
6. What methodologies are appropriate to make candid evaluations of the effectiveness of a particular type of intervention?
7. What are the experience and qualifications of persons on inter-disciplinary teams for specific evaluation tasks and how can their competencies be strengthened?
8. How, by whom and when will the basic information and data needed for evaluation purposes be collected and analyzed?
9. What measures are available for dealing with complaints and disagreements with the findings and recommendations of the evaluation teams?

CHECKLIST FOR SUB-SYSTEM-2: INNOVATION STRATEGY

1. What is the overall allocation for research and development in the national budget, and how is this distributed among the industry, social services,agriculture, health,education,defence and other sectors?
2. What are to be the roles of basic and applied research, scientific break-throughs and incremantal innovation,public,corporate, university research?
3. What is the organization system in public laboratories and the means for prioritization of the research portfolios to deal with the special needs of small and large enterprises?
4. What are the incentive systems to promote research productivity and increased research activity in private and state laboratories?
5. What needs to be done to introduce a program of bench-marking the performance of research institutes? see Annex 2.
6. How does the policy and process of international technological and financial collaborations stimulate (or restrict) the acquisition of selected technologies?
7. If a specific research program is under evaluation, how do the outputs and outcomes compare to the original design?
8. How will the research results be scaled-up and taken to market? Will the benefits be replicable and sustainable?
9. What are the mechanisms at the public, university and private research institutes for assisting the commercialization of research?

CHECKLIST FOR SUB-SYSTEM-3: TECHNICAL HUMAN RESOURCES

1. What is the Government priority and support for science and technology related education and training at all levels,in schools, university, continuous and distance learning, vocational training?
2. How are the new computing, CD-ROM and Internet technologies being applied in preparing students, formulating training materials and methods?
3. What are special measures to improve proficiency in English language and computer literacy at all age levels, including senior citizens?
4. How is tertiary education being prepared for the new millennium?
5. What programs are underway to create the 'entrepreneurial university' and the 'learning enterprise'?
6. What special programs and methods are being used to develop new curricula specific to local needs and conditions, such as 'Management of technology' and Entrepreneurship development'?
7. What are the special programs targeted at education of women, ethnic minorities and disadvantaged groups?
8. What structural changes are being made in the research establishment, services and private sectors to motivate technically competent persons to work in the country rather than migrate? What is being done to attract expatriate nationals to return to their country of origin?

CHECKLIST FOR SUB-SYSTEM-4: TECHNICAL SUPPORT SERVICES

1. What are the main objectives and outputs of the service being evaluated? Importantly, what are the longer-term outcomes expected?
2. What provisions have been made in the original design and subsequent implementation for the collection of the data needed for measuring effective performance, see Annex 3.
3. Have the financial and technical resources been provided to achieve the desired results? Ample finance can be a disincentive while scarce resources can mean that the project managers must continuously mobilize additional money to the neglect of their substantive tasks.
4. How do the operations of the TSS fit in to the larger strategy for the development of the concerned sector? What are the political or other hidden agendas which the sponsors/ financiers of this service seek and why?
5. What are the levels of competence and training of the managers? What are the levels of autonomy and accountability?
6. If the service is in the public sector and receives public subsidy, does this affect private sector or non-governmental providers of the same service?
7. How are national engineering and management consultancy capabilities being promoted?
8. How can the national metrology, standards and quality system be strengthened for international competitiveness?
9. How are the national capabilities being developed for preserving the environment?

CHECKLIST FOR SUB-SYSTEM 5: INTERNATIONAL COOPERATION

1. As a matter of state policy, is the S & T activity such as can benefit significantly by linkages to other countries or donor agencies?
2. What will be the main purposes of international cooperation: developing local skills, acquiring special equipment? Bench-marking and learning from others?
3. In pursuit of better international relations, would the project experience and results obtained be shared with other international agencies, other developing countries?
4. What are the benefits and costs related to forming an international association or other mechanism for collaborative efforts in the future?
5. What mechanisms are in place to attract technical and financial collaborations in selected S & T sectors? To better adapt, apply and improve the technology acquired?
6. How can the S & T system be enabled to share its experiences with other countries, developed and developing? Is there a national technical assistance program to assist other countries?

Annex 2
Case Example: Benchmarking the effectiveness of Research & Training Organizations (RTOs)

Scientific research and technological development account for the bulk (up to two-thirds) of total annual S & T expenditures in many countries. The productivity of research and the performance of research institutes, largely supported by public funds in developing countries, are of importance to the national economy. As technological change is moving at exponential rates, these RTOs can provide major support on training, research and consultancy services for small enterprises, which need to upgrade their total operations.

A comprehensive study has been undertaken by the World Association of Industrial and Technological Research Organizations (WAITRO) on **benchmarking of RTOs,** that is, identifying and measuring those practices which enabled RTOs to enhance their operative and strategic efficiency. Benchmarking is a systematic process whereby comparisons are made on productivity, quality and good practices between a chosen comparable set of RTOs.

Benchmarking requires: an **evaluation** of the research process, the **systematic collection** of information, the **continuous measurement** of performance, and the **long-term commitment** to improve RTO effectiveness.

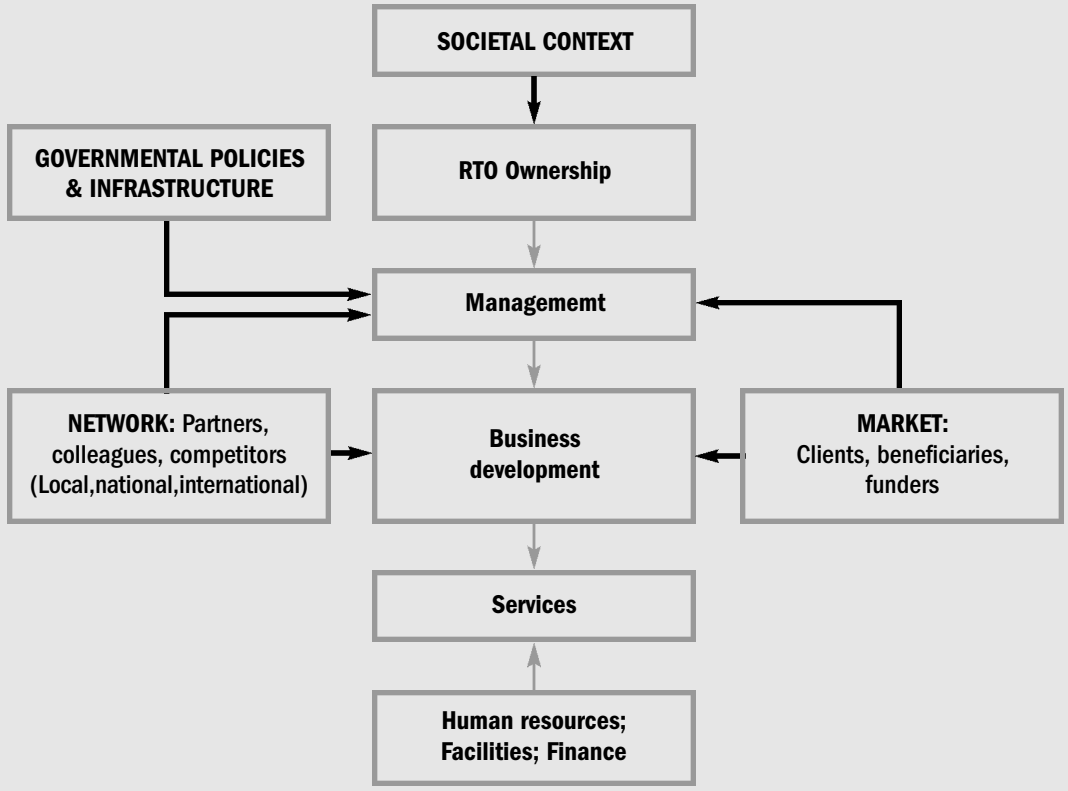
The RTO model can be conceptualized as in Figure A2-1 below.

In the WAITRO project, funded by DANIDA/ Denmark and IDRC/Canada, 60 RTOs were examined (6 in Europe, 10 US/Canada, 20 Asia, 10 central/south America and 14 in Africa) by an international project team consisting of experts from Canada, Denmark and India.

The following processes were benchmarked:

1. RTO governance (ownership, legal structure, governing board's constitution, size and powers,mission and vision,level of autonomy, mandate on sectors or regions).
2. Financial management (funding methods,

FIGURE A2-1: THE RTO MODEL



level, dependence on state support, flexibility, retention of surplus/loss).
3. Services (types such as basic/applied, testing, training, consulting, identification of needs, quality).
4. Business development (activities,effectiveness in attracting business, rewards, promotion, costing, affordability, prioritization).
5. Organizational and project management (structure, team selection, unit responsibility to meet RTO goals, management and follow-up)
6. Capability building (staff development, capacity-building methods, funding of training/equipment).
7. Personnel management (Recruitment process, promotion, compensation, motivation, staff evaluation, internal communication).
8. Networking (relations with other technology providers and with industry to better serve them)
9. Policy and programs (Role in formulating national S & T and industrial policies, leveraging government programs).

As precise information was not available to quantitatively measure performances,broad macro-level indicators had to be used as proxies. These included average growth of income from clients, ratios of grants to total income, expansion of the RTO in services and territory served, and recognition in terms of national and international patents and publications. The above indicators were given weighted ratings, for a total of 100 points.

The main lessons learned regarding good RTO practices are summarized as follows:

- Despite differences in culture, there is a commonality of problems such as poor patronage from industry, declining support from government and difficulties in commercializing research.
- Reliable management information systems are prerequisites for implementing good practices.
- While RTOs from North America,Europe and the industrialized Asian countries were more successful overall,many effective practices were found in RTOs from developing countries.
- The best practice emerged when the RTO was focused on the needs of their clients.

The WAITRO team has developed a consultancy manual of diagnostic tools for transformation efforts at RTOs,including check-lists and procedures on strategic planning, change management, self assessment,technology and industrial audits. RTO transformation excercises have been conducted in several countries based on the WAITRO findings. Their success depends in good measure on the explicit administrative and financial support given to the change process by all the stakeholders.

The process follows the trajectory [visualized in Figure A2-2]. below

FOLLOW-UP ACTIONS

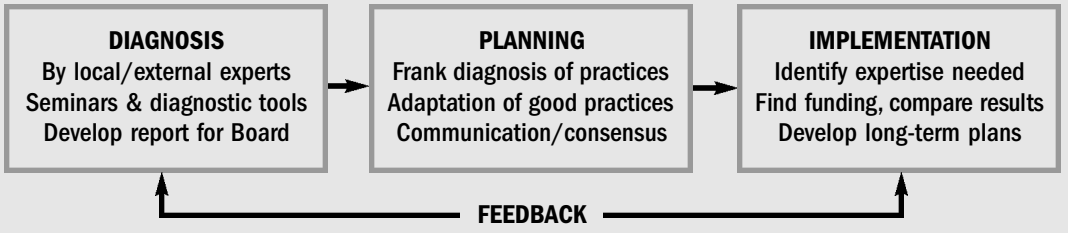
Implementation of a bench-marking process requires the candid analysis of current practices and the progressive move towards an organizational transformation. This is not easy. People resist change, particularly if the change is imposed on them. Real Change happens only when the researchers from the bottom-up and the leadership can both recognize the need for change, can jointly implement a constructive process rather than assign blame, and the leaders can motivate and communicate a vision for the future in which everyone wins.

Concurrently, the research managers need special orientation on new techniques of business management under conditions of global change. This is best done at special training courses focused on the needs and conditions of RTOs.

An RTO benchmarking network calls for sharing of information which in turn requires confidentiality and trust among the parties concerned. A data base of best-practice indicators (with the secret codes rather than actual names of RTOs) can then be made available to those participating in the network.

The next step would be to develop cooperative projects between small enterprises and the RTOs, where enhanced research performance can be applied to the real needs of the small industry sector.

FIGURE A2-2: THE RTO PROCESS



Annex 3
Case example: Evaluating the performance of Technology Business Incubators (TBIs)

FRAMEWORK FOR PERFORMANCE ASSESSMENT

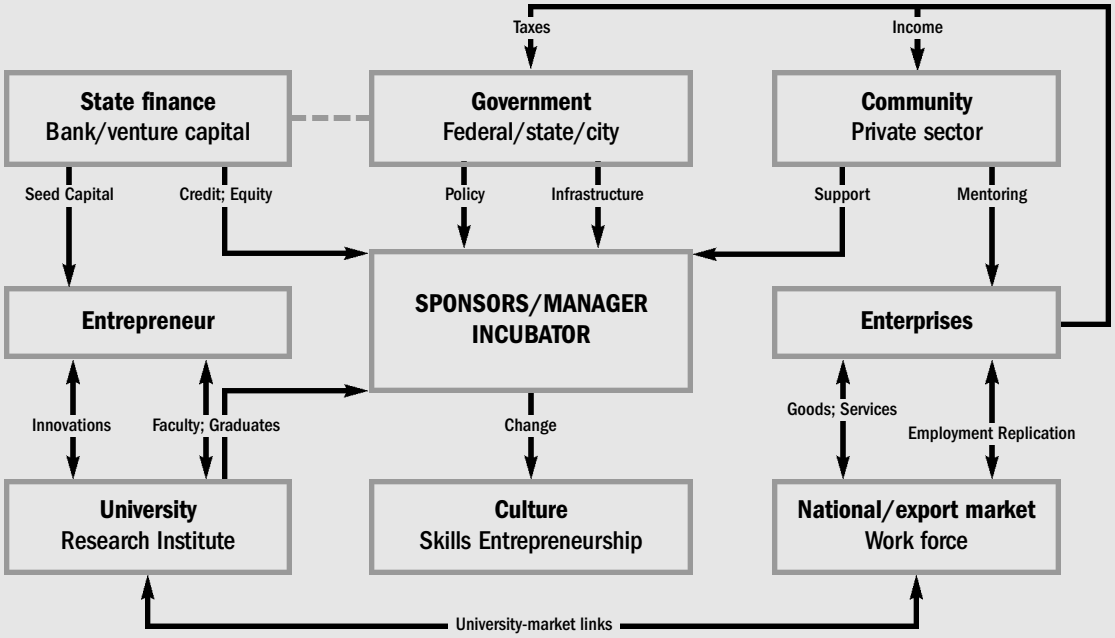
The majority of incubation programs worldwide can be characterized as "public-private partnerships" in which initial (and often continuing) financial support is received from the state bodies. Many governments consider them as part of the business infrastructure, and the evidence indicates that the annual taxes and other benefits from regional economic development more than offset the capital and operating cost subsidy. Private sector participates when it sees that the program will lead to greater business opportunities and promote spin-offs. Good measures of performance of an incubation system are the medium-term benefits accruing to small businesses, sponsors, local community, region and nation, Figure A3-1.

The overall system assessment requires that donors make provision for — and pursue — the collection of the needed information by the management team, on firms in the facility and those leaving as well as other parameters. While some of the coefficients can be calculated readily, others require complex social benefit-cost estimations.

INSTITUTIONAL ANALYSIS OF BIOMINAS AND PARQTEC

The location, implementation, governance and management factors of the Biominas and ParqTec

FIGURE A3-1: ASSESSMENT OF INCUBATOR IMPACTS, EFFECTIVENESS AND SUSTAINABILITY



I. Impact/Outreach	II. Effectiveness	III. Sustainability
1. Enterprises created 2. Survival rate of enterprises 3. Jobs generated (6 years) A. in incubated/affiliated firms B. in graduated firms C. indirect jobs 4.Enterprises reached 5. Replication of "pilot" model 6. 'Extra-curricular' activities	7. Employment per net $ subsidy 8. Taxes paid per net $ subsidy 9. Income, sales & exports 10. Research commercialized 11. Disadvantaged groups 12. Incubator expansion	13. Revenue surplus (6 years) 14. Services cost recovery 15. University-business links 16. Stakeholder satisfaction 17. Tenant/graduate satisfac. 18. Changes in culture 19. Enhancement of skills 20. Leveraging state policies 21. Enhanced self-esteem

incubators are reviewed below. The main characteristics are summarized in Table A3-1.Incubators need to be sited where there is a strong business infrastructure with availability of scientific talent, good living conditions, and positive government and community support. On these counts, both the Biominas and ParqTec incubators have good locations. Both plan major technology parks linked to the incubators.

Biominas: Minas Gerais (MG),the second largest industrial state with its capital at Belo Horizonte, has traditional strengths in minerals, mechanical and automotive manufacture, and now in biotechnology. MG has an impressive technical infrastructure, with universities such as UFMG, Vicosa, Ouro Preto and Uberlandia; research institutes Rene Rachou, and FUNED; support agencies such as FAPEMIG and SEBRAE; and a vibrant private sector with BIOBRAS as a world-class insulin producer. Biominas incubator occupies a prime site of 10,000 sq m land adjacent to the publicly-funded, research laboratory — CETEC. It is a custom-built building with excellent biochemical laboratories.

ParqTec: The city of Sao Paulo is Brazil's major business hub. Some 230 km to the north-west is the city of Sao Carlos, population 200,000, which has the distinction of having the highest density of PhDs in science/engineering — one for every 230 inhabitants. Industry is concentrated on consumption goods and mechanical appliances. It has two public universities and private learning centers, laboratories for cattle and animal protein development, and over 70 enterprises in aero-

TABLE A3-2: PRO FORMA INCOME AND EXPENSE AT PARQTEC AND BIOMINAS INCUBATORS

All figures in US$,000. Rate R/$ =0.84

	PARQTEC			BIOMINAS	
	1996	1997	1998	1997	1998
REVENUES					
Rental Income	24.4	38.6	39.4	28.9	81.7
Fees from Tenants	2.1	2.1	2.5	3.1	12.9
Utilities Cost Recovery	9.2	18.5	18.5	–	–
Partnership with SEBRAE	139.4	62.2	239.4	186.0	186.0
TOTAL REVENUE	**175.1**	**121.4**	**299.8**	**218.0**	**280.6**
EXPENSES					
Manager	27.7	38.6	32.9	36.0	54.0
Admin Assist/tech adviser	0.0	0.7	4.5	10.0	5.0
Lab coordinator	–	–	–	7.0	13.4
Receptionist/Secretary	0.8	2.5	4.2	7.3	14.3
Other Professionals	22.8	26.8	33.4	37.9	48.4
Fringe Benefits	5.9	8.4	10.9	–	–
SUB-TOTAL STAFF	**57.3**	**77.0**	**85.9**	**98.2**	**135.1**
Bldg. Maintenance/Cleaning	31.9	70.6	70.6	8.0	17.5
Utilities/Telephone (net)	18.5	24.4	23.5	30.8	48.4
Travel & Promotion	1.3	0.8	20.2	3.2	3.2
Supplies	0.7	1.3	0.7	4.5	9.3
Audit & legal	9.1	9.1	20.2	3.2	4.9
Insurance	5.0	3.8	4.7	–	–
Publications			56.3	–	–
Interest/debt repayment	–	–	–	30.4	39.7
TOTAL EXPENSES	**123.7**	**187.0**	**282.1**	**178.3**	**258.1**
OPER. SURPLUS (DEFICIT)	**51.4**	**(65.6)**	**17.7**	**39.7**	**22.5**

nautics, informatics, new materials and robotics. Fundacao ParqTec de Alta Tecnologia houses the CINET and SOFTNET incubators together with related facilities.

FINANCIAL VIABILITY AND OUTREACH

Estimates of income and expenses

Estimates of income and expenses for the two incubators are shown in Table A3-2 below. Such data is hard to collect in most countries. Annual revenue from services (in relation to total) is considered low. Support through the partnership with SEBRAE constitutes more than half of revenue, on average.

Both Foundations plan to achieve greater financial self-sufficiency through the development of Technology Parks. Biominas has initiated the feasibility analyses and business planning for a biotech related park adjacent to the existing incubator-CETEC complex.
The São Carlos Science Park is on a 172,000 square meter property that it owns in a prime industrial location. Planning and design for the first 3,500 square meter building has been completed along with a master plan for the balance of the property. The first structure will house the ParqTec headquarters as well as incubator modules for 64 additional tenant enterprises. The master development plan also includes industrial sites for lease to technology-based enterprises as well as two multi-tenant buildings and a convention center, designed to establish

Cost effectiveness and stakeholder satisfaction

The evaluation methodology in this case example uses a combination of qualitative description, quantitative analysis, and stakeholder perceptions. The stakeholders interviewed for this purpose were the public and private sponsors as well as the incubated enterprises themselves.

Businesses incubated and jobs created

An approach to evaluating the effectiveness of incubation programs is to look at the number of businesses incubated, the success rate, and the number of jobs created by incubated firms. As noted, both incubators have to aggressively recruit more tenants and affiliates as well as increase the throughput of graduating businesses.

The figures in Table A3-3 below should be considered as preliminary; as it is often difficult to get data from privately held firms on sensitive topics such as sales, payroll and taxes. At ParqTec, the tenant firms have 69 employees while 17 (of the 21) graduated firms have 168, making a total of 237 direct jobs.

As ParqTec has been in operation at its present location since 1990, it has more results to evaluate in compared to Biominas,which has been operating only since July 1997 in its new permanent facilities. The 1997 estimate of public capital and operating subsidy for ParqTec and the personal and corporate taxes payable would be approximately as follows, based on the limited data available:

Total jobs (with employment multiplier of 1.5*)	357
Capital cost subsidy per year (20-year straight line depreciation)	$ 19,150
Operating subsidy per year (average of last 3 years)	$147,000
Capital and operational subsidy per year	**$166,150**
Total subsidy over 7 years	$1,163,050
Subsidy cost per job (excluding jobs in affiliates)	**$ 3,258**
Estimated payroll & corporate taxes by tenants & graduated firms	$1,054,320
Return on public investment as taxes per year	**$6.34 per $ subsidy**

* *Indirect employment multiplier based on ranges for similar economic activities in the U.S.*

TABLE A3-3: JOBS AND TAXES, 1997(APPROXIMATE), US $

	ParqTec	Biominas
Jobs (tenants and graduates)*	237	92
1997 payroll	$1,854,000	$1,030,040
1997 sales	$9,846,990	$2,558,320
1997 payroll taxes payable	$ 463,500	$ 258,510
1997 corporate taxes payable	$ 590,820	$ 153,500
Total taxes	$1,054,320	$ 412,010
Initial Investment in incubator	$ 383,000	$1,940,000

* *This includes current tenants plus the one graduate tenant at Biominas and 17 graduated firms at ParqTec for whom information is available. Taxes are estimated at 25% on payroll and 6% on sales.*

The subsidy per job should decline at ParqTec as more firms graduate and continue to expand, and as additional incubator space becomes available. For mixed-use incubators, which typically have much larger areas and less services for tenants, the subsidy cost per job can be much lower. A point to note is that while the investment is made once, the jobs continue, and it is useful to think in terms of "job-years" in the stream of benefits.
1997 taxes realizable from sales and payroll of ParqTec tenants and graduates could be about six times the subsidy.

Performance evaluation by incubator tenants and graduates

Biominas: Representatives of the present enterprises and the one graduated were asked to evaluate the effectiveness of the incubator as well as the advantages and disadvantages in being tenants. All persons interviewed felt that the program is of value to them. The major benefits expressed were as follows: help in dealing with bureaucracies resulting in faster permits; valuable assistance in marketing and faster time to market for new products; excellent infrastructure and labs; interaction with other tenants; and legal assistance.

ParqTec: The incubator tenants and graduates interviewed expressed satisfaction with their experiences. The major benefits cited were its: good location for a startup venture, access to facilities such as labs. telephone, internet, and fax service, valuable marketing assistance received, legal assistance for incorporation and patent development, and business training on site.

LESSONS LEARNED

To summarize, the ParqTec and Biominas incubators studied have had **positive impacts and outcomes** on their respective city and state economies in nurturing entrepreneurs and creating sound enterprises with good survival rates. ParqTec has generated employment with public subsidy of around US$ 3,258 per job, *without* including jobs in affiliates. The estimated return in the form of taxes could be about $ 6 per dollar of public subsidy.

The linkages to universities and research institutes have resulted in commercialization of some technologies. The sponsors and tenants at both incubators have expressed satisfaction with the results achieved, particularly the help in marketing, business planning, and securing government permits. Both are helping their government sponsors in promoting technological development together with other social aspects such as reinforcing the cultures of entrepreneurship and university-research-business cooperation.

That being said, Biominas and ParqTec have the major challenges ahead of enhancing their operational effectiveness through innovative activities and creative financing, increased occupancy and higher fees for quality services, with more affiliate companies and anchor tenants, in order to reduce dependence on state subsidies.

13 REVIEW OF DRAFT OF SCIENCE & TECHNOLOGY EVALUATION STANDARD OF CHINA

Chen Zhaoying, Vice President,NCSTE, China

1. BACKGROUND

The science and Technology Evaluation Standards of China (hereinafter referred to as the Standards) is developed in accordance with the following demands:

- **Demands from evaluation performers (the people who "do" evaluation)**
  Evaluators/evaluation organisations need unified standards in order to regulate their evaluation practice, to standardise the planning and regulate their evaluation, to raise level of science and implementation of evaluation, to raise the overall level of science and technology evaluation, to ensure the evaluation quality and to reduce and avoid disputes.

- **Demands from evaluation users (the people who "use" evaluation results)**
  Through the Standards the evaluation consignors and users can better understand science and technology evaluation in order to utilise the evaluation results.

- **Demands from government agencies**
  Mrs. Zhu Lilan, Minister of Science and Technology pointed out at the National Technical Innovation Conference that we should establish a scientific and impartial evaluation system, implement it according to law and gradually standardise our science and technology evaluation.

2. CHARACTERISTICS

A SELF-DISCIPLINARY DOCUMENT

The Standards is a professional self-disciplinary document. The formulation of the Standards comprises an important part of the management of science and technology evaluation. The Evaluation Standards Committee is responsible for the interpretation and revision of the Standards, but has no power to enforce the implementation. In accordance with the Standards, member organisations of science and technology evaluation, such as federations and associations, may conduct self-management and self-control of the evaluation development and development disciplinary actions against the violation of professional conduct and code of ethics.

THE ACCEPTED EVALUATION STANDARDS WITH A BINDING FORCE FOR ENFORCEMENT

According to international practice, evaluation standards can be approved by relevant government agencies as an accepted evaluation standard with a binding force for enforcement.

The initiative and development of science and technology evaluation in China have been supported by the government. The evaluation on government investment in science and technology will occupy an important place in the overall science and technology evaluation for a period of time in the future. The administrative department of the government should approve the standard or part of it for use of relevant evaluation organisations as a reference.

THE STANDARDS FOR PROFESSIONAL PRACTICE

As a standard of professional performance, the Standard is a condensation of the theories and practices of science and technology evaluation, reflecting the appraisers'understanding, viewpoints and experiences of professional evaluation. Evaluation work requires both creativeness and strict compliance with the standards, and a combination of professional responsibilities and morality.

SCOPE

The Standard is a standard for science and technology evaluation profession. Any evaluation activity relating to science and technology including planning, achievements,projects,organisations,policy and personnel can be covered by the Standard.

EFFECTIVE TIMEFRAME

The Standard is used not only to address the problems in present evaluation activities, but will also play roles for a long period to come. Therefore, pertinence, continuity should be adhered to during the formulation of the Standard.

3. MAJOR REFERENCE SYSTEMS FOR FORMULATION OF THE STANDARDS

UNIFORM STANDARDS FOR PROFESSIONAL APPRAISAL PRACTICE (USPAP) OF THE UNITED STATES

USPAP is a professional standards influential not only in the United States, North America, but all over the world. It is a standard document with a self-disciplinary force for the profession formulated by professional appraisal association and recognised by appraisers, appraisal organisations and appraisal clients. The effectiveness and influences of the standards surpass the ordinary standards formulated by other professional associations for it was legislated and approved by the government as an accepted appraisal standard Almost all the customers in the United States require that the appraisal assignments should comply with USPAP. The authority of USPAP has been widely acknowledge by government agencies and companies in the United States. It has become the basis for formulating an international appraisal standard.

Since the promulgation of USPAP on April 27, 1987, it was revised 13 times. Significant changes have been made while the major structure remains.

INTERNATIONAL PROPERTY APPRAISAL STANDARDS (VOL.94,95)

International Property Appraisal Standards was formulated and promulgated by the International Evaluation Standard Committee. The framework is designed in accordance with the objectives of property appraisal activities and the derived values.

EVALUATION STANDARDS OF OECD COUNTRIES

"Evaluation Standards for Science, Technology and Innovation Projects" OECD-1990 (excerpt).

GOVERNMENT AUDITING STANDARDS (UNITED STATES AUDITING ADMINISTRATION)

- Financial audit
- Audit relating to financial affairs
- Performance audit
- Comprehensive audit (contract audit)

FRAMEWORK OF CHINA'S ASSET APPRAISAL STANDARD SYSTEM

The Chinese Evaluation Association is now making research on and formulating the assets appraisal standards in the light of USPAP.

4. FRAMEWORK AND CONTENTS

The differences of evaluation activities in different counties are reflected in their market economy conditions, legal and management systems while the basic evaluation methods and techniques have little differences all over the world (refer to the International Assets Evaluation Standards Vol. 94&95). The Standard takes the internationally adopted open frame structure, which can ensure the overall stability of the Standards and accommodate the changing conditions of evaluation so that the Standards will be improved step by step.

FRAMEWORK AND CONTENTS OF SCIENCE AND TECHNOLOGY EVALUATION STANDARDS

Part I. The main body

The first level:basic standard

This part is composed of basic principles, codes of ethic and terminology of science and technology evaluation. The basic standard shows the features of science and technology evaluation and serves as a basic and general standard for science and technology evaluation profession. It is the basis for guiding and regulating all kinds of technical evaluation activities. The basic standard sticks to the basic rules of evaluation profession, and at the same time, has its own characteristics of technical evaluation. The main contents include:

- Standard of professional conduct.
- Principle of independence and impartiality, principle of avoidance and misleading, confidentiality, and fees.
- Professional morals (activities violating the professional morals).
- Capabilities and qualification of the appraisers/ evaluation organisations, the structure of appraisers' competency and capabilities,continuing education and training of the appraisers.
- Alternatives (requirements for using the alternatives).
- Basic concepts and terminology of science and technology evaluation.

The second level:technical standard

Composed of the major evaluation procedures, and the obligations and rights of the related parties of the evaluation, the technical standards are the basic part of the Evaluation Standards. It describes the main principles, key issues and points to be noted in the planning and implementation of the evaluation, and stipulated in detail the requirements and contents of the evaluation report. The main contents of the technical standard include the evaluation report. The main contents of the technical standard include:

- The roles,obligations and rights of the parties involved in the evaluation
- Evaluation procedures
- Preparation of the evaluation
- Evaluation planning

The preconditions/presupposition of evaluation, evaluation standards, evaluation index, reliability of evaluation, rating system of evaluation

- Information collection, data reviewing and sorting.
- Selection of evaluation methods.
- Multi-index comprehensive evaluation model, case study, comparative study.
- Requirements for evaluation report
- Utilisation of Evaluation results
- Quality control of evaluation activities
- Archiving of evaluation activities.

The Third level:statements on standards

Statements on standards forms as an organic part of the Standards,which further identify, intemperate and explain the evaluation standards and have the same force as the text of the standards. As the statements are flexible, they should be stipulated as required and after being well conceived. The statements can be upgraded into part of the standards according to the practical situation and maturity level during the process of the revision of the standards. The first edition of the statements on standards contains the following:

- How to maintain independence, objective and impartial? (three elements endanger the independence).
- How to justify that the appraisers capabilities meet the requirements of the evaluation assignment?
- Different requirements and characteristics of different types of evaluation activities.
- Major concepts and methods of classification for science and technology evaluation in foreign countries.
- When to use the alternatives regulations in the process of evaluation? What are specifically required for the evaluation report when the alternative regulations are used?
- How to make use of comments from the evaluation results caused by the limits?

Part II. References

Guidelines for the implementation of typical science and technology activities

The following is some references, cases and advises which have no binding forces for appraisers in light of some typical science and technology evaluation activities.

- Evaluation on science and technology programs
- Evaluation on project relevance
- Evaluation on project acceptance
- Quality control of individual evaluation

Explanation on the background,concepts, methods and processes of the development of the Science and Technology Evaluation Standards

Introduction of some representative evaluation standards and norms of both domestic and abroad

5. METHODS AND PROCESSES OF THE DEVELOPMENT OF THE STANDARDS

STRESS ON FUNDAMENTAL WORK

The development of American Uniform Standards for Professional Evaluation Practice (USPAP) over the past ten-odd year has given us much enlightenment. First and foremost, USPAP is a condensation and summary of the theories and practice of evaluation. A complex system engineering, USPAP is developed on the basis of sound theories and practice. In the process of formulating the Evaluation Standards for Science and Technology of China, much attention is given to the fundamental work, especially the research on evaluation theories and the summing-up of experiences and cases, in order to lay a sound basis for the development of the Standards.

DEVELOP THE STANDARDS ON THE BASIS OF INTERNATIONALLY ACCEPTED EVALUATION PRINCIPLES; TAKE IN THE NEW CONCEPTS AND PROGRESS OF THE INTERNATIONAL EVALUATION PROFESSION

The practice of international evaluation profession shows that the basic evaluation principles in all the countries,mainly the code of conducts and professional morals, are almost the same, although different countries have their own national conditions and different types of evaluation activities. Base on the internationally accepted basic evaluation principles, the Standards generalised the experience and research results of the evaluation practice in China, and took in the new concepts of and the latest progress of the theoretical research on international evaluation profession.

Generally speaking, standards are developed by two ways:inductive and deductive. The inductive way is to form specific rules, first according to experiences and general practice, and then formulate general evaluation standards. From specific to general,the standards formulated this way are descriptive. Making standards in the way of deduction is to identify the generally accepted concepts first and then formulate the specific evaluation standards. From general to specific, standards formed this way are normative. Judging from the evaluation standards formed this way is normative. Judging from evaluation standards of different countries and regions,the standards would not be consistent if they are formulated just on the basis of experiences and general practice. Based on the conceptual structure of evaluation standards of other countries, the Standards are formulated with methods of combining deductive and inductive ways. The experiences and lessons learned in the science and technology evaluation practice in China are fully related in the Standards.

PARTICIPATION AND TRANSPARENCY OF THE STANDARDS FORMULATION

Importance has been given to the participation of the appraisers and clients in the process of the formulation and revision of the Standards. It is planned to set up an evaluation standards committee at proper time and promulgate the draft standards to the public for comments and suggestions so as to improve it to a top level of the profession. The early participation of people working in this profession and other related fields enabled the universality of the standards and laid a good foundation for the understanding and implementation of the standards after it is formulated and revised.

Publicity should be given to the standards after it is formulated and revised, especially the updated information. The revised Uniform Standards for Professional Evaluation Practice should be published each year. The revised part should be notified in the preface in order to allow the appraisers and customers to understand the latest development and make a better use of the standards.

REVISE THE STANDARDS REGULARLY AND IMPROVE IT STEP BY STEP

As the evaluation standards need to be revised and reprinted with the deepening of the evaluation practice and the changing of the situation, it should be formulated to meet the basic demands at first rather than covering everything. New contents should be added to the Standards and revision should be made step by step. In this way we can eventually produce an evaluation standards for science and technology with Chinese characteristics, which is influential and has a strong binding forces. As evaluation theories are not perfectly established at present, many basic definitions and principles are not clearly identified. We should not finalise the present research results of evaluation. Instead, only the matured theories and experiences can be absorbed in the evaluation standards.

The Evaluation Standards Committee will make timely revision according to the situations of practice and the implementation on the basis of reviewing the evaluation theories and practices. The revision will be illustrated in news media and professional journals in order to obtain sufficient attention and understanding of the public. The standards,then, will always reflect the latest theories of the evaluation profession and remain proper consistency to serve as the guidelines of evaluation practice. In summary, China is still at the first stage in formulating the evaluation standards for science and technology. Therefore the standards will not be made perfect to cover everything at one step.

PART IV
ECD Opportunities for National Initiatives and International Cooperation

14

GUIDE TO DIAGNOSING ECD NEEDS: Video Speech

(Full Presentation is available in booklet form)

Keith Mackay, Senior Evaluation Officer, OED World Bank

DEVELOPMENT OF M&E SYSTEMS WITHIN A GOVERNMENT – DIAGNOSING ECD NEEDS + READINESS

- Danger if we take too simple an approach and assume M&E a "good thing" and its merits should be obvious to all development assistance.

- World Bank experience (and other dev't asst agencies) is that most forms of capacity building are difficult to achieve, and require long-term efforts if they are to be sustainable.

- This also applies to ECD.

- A key lesson from experience is that to be successful/sustainable we need to understand broader institutional environment.

- Another key lesson is that there is no single, best model for ECD.
  - ECD has to be tailored to each country's circumstances.
  - One motivation for ECD is that many of the most advanced economies in the world typically have some type of national system for M&E.
  - Countries such as Australia, Canada,USA, + various European countries for instance.
  - But when you compare different countries, you find that their approaches differ considerably.

- Increasingly, a national M&E system is not being seen as a stand-alone activity, but as part of sound governance:
  - performance measurement is seen as counterpart to performance management;
  - the multiple uses of a national M&E system can provide support to a number of different public sector reforms.

  In Ghana for instance, there are a number of related and mutually-supporting reforms:
  - there is a medium-term expenditure framework (MTEF) a – e this sets out funding allocations for each type of activity for every ministry over next 3 years. The MTEF also focuses explicitly on planned government outputs,and this necessitates the accurate measurement of actual outputs;
  - there are performance improvement plans a – e prepared by each ministry and setting out particular initiatives designed to help build a client-oriented service culture in the civil service;
  - and there are also employment contracts for senior officials a-e these ensure a close focus on what results (what performance in terms of promised outputs) are expected from senior officials over the coming year;
  - and a big push to decentralize government to the district level a – e this creates

a greater need to be able to measure government performance at the regional and district levels, to provide greater support to district government management.
- Ghanaian government (GOG) officials recognize that national + sectoral M&E systems can support all these reforms.
- actually, GOG officials increasingly understand that M&E does more than support the reforms: they are coming to realize that a number of the reforms will not be fully successful unless the government's performance is measured more systematically and rigorously.

- So in your country, helps if you identify the different uses + different users of M&E info, and the different types of M&E

- Different types fill different roles:
  1. Basic financial data on government spending on different activities, projects, sectors
     - these data are complementary to basic socio-economic statistics or development indicators
  2. Performance monitoring/indicators on government inputs, processes and outputs
  3. Cost-benefit analysis
  4. Formal program evaluation
  5. More qualitative reviews
  6. Performance audits a – e typically done by a National Audit Office

- Different uses of M&E findings include:
  1. Policy formulation and planning to help clarify the pros and cons of alternative types of policy intervention;
  2. A second is resource allocation in budget processes so M&E findings can be *one* input to Government decision-making, along with a number of other influences on government;
  3. To support management via greater results orientation — the learning function;
  4. Accountability—this is the quid pro quo in response to devolution of authority to managers.

- The need to understand which of these different uses of M&E likely to be most important in your country.

- It also helps to understand which different potential users of M&E actually want this info.

- A mistake which we evaluators make is to be supply-driven: we may think that all we need do to make a national M&E system is a Government decree, plus some training
  - but really, our experience tells us that this will not be enough

- But another key lesson we have learned is that there has to be real demand æ real commitment — by a Government to using M&E if a system is to be developed in a sustainable manner
  - An example of this is our work with the Government of Benin to develop performance indicators to support budget decision-making — as part of an MTEF
  - the Finance Minister in Benin is strongly committed to this use of performance indicators

- Consideration of types of M&E — and of the different users and uses of them — has big implications for how you structure M&E functions and implications for who plans or commissions M&E, who conducts it, and who uses it.

The World Bank has prepared an ECD Diagnostic Guide[1]—to help Governments and analysts in addressing these issues and in developing an action plan for ECD.

- There are eight key steps :
  - the Guide provides checklists of questions and issues to help you
  - it's really what we call a SWOT (strengths, weaknesses,opportunities and threats) analysis

  1. Identify key ministries and their formal relationships
     - ministries with major role in performance management — i.e., in resource allocation decisions and in ongoing management of sectors + projects
     - e.g., key central ministries and large line ministries

  2. Diagnose public sector incentives and the unwritten "rules of the game"
     - it is these unwritten rules + incentives, not formal decrees and procedures,which determine the behavior of ministries, line managers and other civil servants.
     - so what is extent of autonomy + authority of line managers?
     - what are rewards + incentives for good performance?

[1] *Keith Mackay,* Evaluation Capacity Development: A Diagnostic Guide and Action Framework, *ECD Working Paper Series No 6, January 1999,The World Bank.*

- what are sanctions for poor performance?
- so, this Step helps answer the question: "does a performance culture exist?" If not, "what are the roadblocks to achieving it?"

3. Find out how budget resource allocation and line management decisions are actually taken a – e the reality, not the rhetoric
   - this Step looks at actual systems,structures, roles and info flows

4. Next step is to investigate the extent of influence of M&E on budget and line management decisions
   - again, "what is the reality?", not the rhetoric

5. Map out the evaluation activities and capabilities of ministries and other organizations
   - this step focuses on supply of M&E,and on processes for making M&E findings available

6. Map out the evaluation activities and contribution of development assistance agencies
   - development agencies often focus on building M&E for donor-funded projects, and this may have contributed to national M&E capacities
   - and they might have been involved in supporting other relevant capacity-building work in the past, in areas such as building national statistical systems

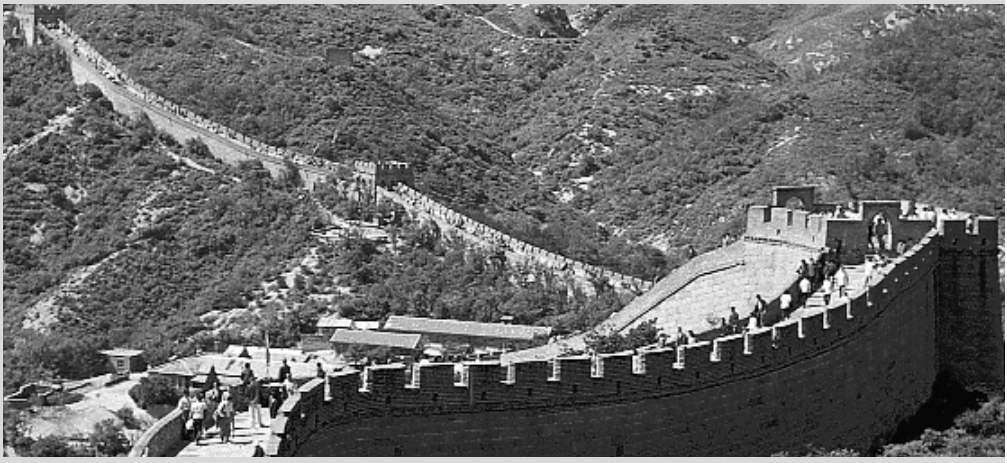
7. Understand the major public sector reforms underway or completed in recent years æ these may well provide synergies with ECD a – e as Ghana, for example, illustrates

8. Map out an action plan for ECD, focusing on key issues and dimensions:
   - demand and supply
     - where is existing demand for M&E, and how can it be consciously built up by "winning hearts + minds"?
     - where is existing supply of M&E,and again, how can it be built up?
   - different types of M&E;
   - evaluation infrastructure and systems
     - what planning systems would be desirable to help decide which areas of Government should be evaluated?
     - what mechanisms are necessary to ensure evaluations are completed in timely manner?
     - what mechanisms would be set up to provide M&E findings to users of them?
   - possible support from development assistance agencies
     - this could include technical assistance and advice, training, twinning arrangements; support for ECD diagnoses, etc;
   - think of timelines for an action plan,and sequencing and speed of implementation
     - need some flexibility in approach
     - but helps to set indicative targets for 1,3, 5 and 10 years in the future; and
   - sustainability a – e how sustainable do you think your ECD efforts might be?
     - what are some of the risks and threats to it?

- This is a detailed and systematic approach; it helps ensure that your ECD work takes full account of opportunities and possible barriers

- I regret not being with you in person to discuss these issues —I'm sure your debate will be interesting and I wish you every success with your ECD efforts.

PART V
Meeting Wrap-Up



15 MEETING WRAP-UP

Christine Roth, Evaluation Advisor, UNDP

Presentations and discussions during the 1999 Beijing Conference on Evaluation Capacity Development gave us a glimpse of the ever-evolving environment we are living in. At the dawn of the 21st century, in a world of increasing globalization, competition, and reduced public resources, scrutiny is growing for governments and development agencies to demonstrate value for money in public services.

As a consequence, the field of evaluation is changing rapidly to deal with the new demands placed on it. First and foremost, evaluation is being recognized as an essential part of good governance and public sector reform. With the advent of public sector reforms, governments and aid agencies have begun to revisit their monitoring and evaluation (M&E) arrangements, looking for ways to develop effective learning systems and promote transparency and accountability.

One of the key elements of this reform has been the adoption of results-based approach to management (RBM). The Conference focused particularly on the interface between RBM and monitoring and evaluation in connection with public sector reform and good governance. The shift to RBM requires a focus on development results, or outcomes, rather than inputs. In this paradigm, the continuous assessment of whether intended outcomes are being achieved is critical. M&E becomes a central function in a results oriented environment. Moving to outcomes also requires the development of strategic partnerships and the concerted efforts of many actors in society. Methodologically, this approach is raising several issues, such as attribution and accountability inter alia.

During the past two days, we have engaged in lively discussions on some of these methodological issues as well as broader topics including:

- The Independence of the evaluation function
- The necessary link between evaluation and decision-making
- The balance between learning and accountability
- The challenges facing evaluation
  - Political environment
  - Systemic problems
  - Internal capacities (human and financial resources)
- The role of partnerships
  - Joint evaluations with governments, aid agencies and development banks
  - Involvement of stakeholders and civil society in evaluation exercises

Of the many constraints hindering the establishment of effective monitoring and evaluation systems at national level, we have collectively identified the following:

- Weak demand for evaluation
- The lack of political commitment and appropriate institutional support for evaluation
- The lack of standards and criteria for evaluation
- The lack of local expertise and resources for evaluation

As a first step to lifting these constraints, the concept of networks has emerged in our discussions as a powerful instrument to mainstream the culture of evaluation in the public sector by allowing: (1) exchange of experiences, (2) access to best practices, (3) sharing databases.

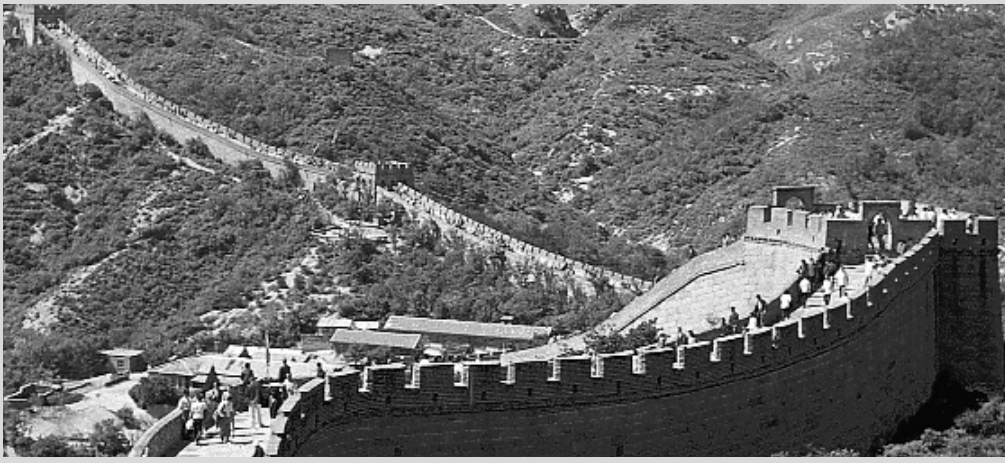
Concretely, several countries have committed to establishing evaluation networks at the national and regional level. Sri-Lanka and China are two examples and support will be provided as needed by the international community.

At the international level, both UNDP and the World Bank, in agreement with several other donors are promoting the establishment of an international development evaluation association. This association will help to form links between evaluators engaged in development related activities and will support the national efforts in strengthening evaluation capacity development.

Last,an opportunity appeared during this meeting for the development agencies in general,UNDP in particular, to maximize the synergies between the work of development banks – the Asian Development Bank in this particular case – and of national entities to promote the role of evaluation in good governance and public sector reform.

As we move forward in the field of ECD, Beijing will represent one of the milestones that has helped furhter refine our understanding of issues and contributed in shaping the international community's game plan to meet the challenges ahead.

PART VI
Annexes



ANNEX I:
List of Participants

CHINA

Mr. Sun FUNQUAN
Professor
NCSTE – National Center for Science and Technology Evaluation
P.O. Box 3814
Beijing, People's Republic of China100038
Tel: (86) 10-685 14841/685 15544-3805 Ext.3825
Fax: (86) 10 68514854
Email: tac@public3.bta.net.cn

Mr. Li XINNAN
Professor, Vice-President
NCSTE – National Center for Science and
Technology Evaluation
P.O. Box 3814
Beijing, People's Republic of China100038
Tel: (86) 10-685 18087/685 15544-3805 Ext.3818
Fax: (86) 10 68514854
Email: lxin@public.bta.net.cn/tac@public3.bta.net.cn

Ms. Chen ZHAOYING
Professor, Deputy Director
Member of Executive Committee, International Association for Technology Assessment and Forecasting Institutions
P.O.Box 3814, Beijing, P.R. China 10038
Tel: (86)10 68515544/66515544
Fax: (86)10 68514854
Email: tac@public3.bta.net.cn

Mr. Fang YAN
Head, Evaluation
NCSTE –National Center for Science and Technology Evaluation
P.O. Box 3814
Beijing, People's Republic of China100038
Tel: (86) 10-685 18087/685 15544-3805 Ext.3818
Fax: (86) 10 68514854
Email: tac@public3.bta.net.cn

DEMOCRATIC PEOPLE'S REPUBLIC OF KOREA

Mr. Kim SEON-BIN
Director
Public Welfare Technology R&D Division
Ministry of Science and Technology
Government Complex-Gwacheon
Seoul, Democratic People's Republic of Korea, 427-760
Tel : (82-2) 503 7606
Fax :(82-2) 749 1417
Email : kimsb@mostws.most.go.kr

FIJI

Ms.Litia MAWI
Director of Management Improvement
Public Service Commission
29 Greene Street, Howell Road
Suva, Fiji Islands
Tel: (679) 386 001
Fax: (679) 301 714

Mr. Yuxue Xue
Assistant Resident Representative
UNDP Fiji
Private Mail Bag
Suva, Fiji
Tel: (679) 312 500
Fax: (679) 301 718
Email: registry@undp.org.fj

INDIA

Dr. S.P. PAL
Advisor, Planning and
Evaluation Organization
Planning Commission
Yojana Bhavan
Parliament Street
New Delhi - 110001
Tel: (91 11) 3710162/91 11 221 3653
Fax: (91 11) 371 7681/91 11 372 5491

KAZAKHSTAN

Ms. Svetlana CHERIMISSINOVA
Deputy Head State Programmes Division
Regional Policy Department
Ministry of Economy
33 Pobedy Prospect/52 Abai Str,
Astana, Republic of Kazakhstan
Tel: (7-3172) 11 81 34/29
Fax: (7-3172) 11 81 36

KYRGYSTAN

Miyegombo ENKHTSETSEG
Head, Programme Support Unit
UNDP, Kyrgystan
90 Chuykov Street
Bishkek 720000, Kyrgystan
Tel: (906-312) 666100
Fax: (996-312) 66055
Email: root@undp.bishkek.su

Mr. Chorobek IMASHEV
Deputy Minister of Finance
The Kyrgyz Republic
58 Erkindik Blvd., #305
Tel: (996-312) 22 74 55
Fax: (996-312) 66 16 45

Mr. Leonid Nukhimovich KOMAROVER
Head, Economic Analysis and
Prognosis Department
Office of the Prime Minister of
the Kyrgyz Republic
205 Chuy Avenue #401
Tel: 22 29 24
Fax: 66 13 82/66 25 77

Mr. Sabyrbek MOLDOKULOV
Deputy Chairman, State Committee on
Foreign Investments
58a Erkindik Bvld, #210
Tel: (996-312) 22 32 92
Fax: (996-312) 66 10 45

Ms.Elzira SAGYNBAEVA
National Programme Officer
UNDP Kyrgystan
90 Chuykov Street
Bishkek 720000, Kyrgystan
Tel: (906-312) 666100
Fax: (996-312) 660557
Email: root@undp.bishkek.su; Fo.kgz@undp.org

MALAYSIA

Dr. Selva RAMACHANDRA
Assistant Resident Representative
UNDP Malaysia
Wisma UN
Block C, Damansara Offices Complex
Damansara Heights
50490 Kuala Lumpur, Malaysia
Tel: (60 3) 253 1020/ (60 3) 255 9122
Fax: (60 3) 255 2870

Mr. Ab.Alim BIN ZAKARIAH
Principal Assistant
Director Economic Planning Unit
Prime Minister's Department
Blok B6, Aras 2
Kompleks Jabatan Perdana Menteri
Pusat Pentadbiran Kerajaan Persukutuan
62502 Putrajaya
Tel: (603) 88882767
Fax (603) 88883768

MONGOLIA

Mr. Rentsen BATMEND
Principal Officer, Cabinet Secretariat
Room 374
Government House,
Ulaanbaatar 12, Mongolia

Dr. Dungersuren SUKHJARGALMAA
Assistant Resident Representative
UNDP Mongolia
7 Erhuu Street
Sukhbaatar Region
Ulaanbaatar, Mongolia
Tel: (976 1) 327 585
Fax: (976 1) 326 221
Email: fo.mon@undp.org/registry@undp.org.mn

NEPAL

Mr. Devendra Raj Panday
Chairman
Nepal South Centre
Tripureswar
Khathmandu
P.O. Box 8242
Tel: (977) 1 247 656
Fax: (977) 1 240 685
Email: nesac@npl.healthnet.org

PAKISTAN

Ms. Shaheen MASUD
Joint Secretary, Economic Affairs Division
Ministry of Finance and Economic Affairs
Address: **MISSING TEXT**
MISSING TEXT
Tel: (051) 9201 706
Fax: (051) 920 5971 / (051) 921 0734

PHILIPPINES

Mr. Ernesto BAUTISTA
Assistant Deputy Resident Representative
UNDP Philipppines
NEDA Sa Makati Building
106 Amorsolo Street
Legaspi Village
1229 Makati City
Metro Manila, Philipppines
Email: ernesto.bautista@undp.org

SRI LANKA

Mrs. Nanda ALAHAKONE
Director, Post Evaluation Unit
Ministry of Plan Implementation &
Parliamentary Affairs, 3rd Floor
"Sethsiripaya"
Battarumulla
Sri Lanka
Tel: 863 954
Fax: 862 485

Mr. Kithsiri ASOKA KASTURIARACHCHI
Assistant Resident Representative
UNDP Sri Lanka
202-203 Bauddhaloka Mawatha
Colombo, Sri Lanka
Tel: (94 1) 582630
Fax: (94 1) 581 116
Email : asoka.kasturiarachchi@undp.org
THAILAND

Ms. Netnarumon SIRIMONTHON
Assistant Resident Representative and
Programme Management Specialist
UNDP Thailand
United Nations Building
12th Floor
Tel: (66 2) 280 0519
Fax: (66 2) 280 0556
Email: fo.tha@undp.org

VIETNAM

Mr. Nguyen DINH CUNG
Director, Macro Economic Policy Department
Central Institute for Economic Management
68 Phan Dinh Phung
Honoi, Vietnam
Tel: (84 4) 845 8241
Fax: (84 4) 856 6795

THE UNITED NATIONS DEVELOPMENT PROGRAMME (UNDP)

Mr. Khalid MALIK
Director, Evaluation Office
UNDP, Room DC1-478
One United Nations Plaza
New York, NY 10017
United States
Tel: (212) 906 5095
Fax: (212) 906 6627
Email: Khalid.Malik@undp.org

Mr. Antonio MOLPECERES
Learning Resource Advisor
304 E 45th Street
New York, NY 10017
Room FF-1104
Tel: (212) 906 6724
Fax: (212) 906 5120
Email: Antonio.molpeceres@undp.org

Ms. Christine ROTH
Evaluation Advisor
UNDP, Room 476
One United Nations Plaza
New York, NY 10017
United States
Tel: (212) 906 5722
Fax: (212) 906 6627
Email: Christine.roth@undp.org

THE WORLD BANK OPERATIONS EVALUATION DEPARTMENT/ WORLD BANK INSTITUTE

Mr. Keith MACKAY
Senior Evaluation Officer
OED Room G 3212,The World Bank
1818 H Street, N.W. Washington, DC 20433
U. S. A.
Tel: (202) 473 7178
Fax: (202) 522 1655
Email: kmckay@worldbank.org

Mr. Ray RIST
Evaluation Advisor
WBI Room G 3192,The World Bank
1818 H Street, N.W. Washington, DC 20433
U. S. A.
Tel: (202) 458 5625
Fax: (202) 522 1655
Email: rrist@worldbank.org

Ms.Jody ZALL KUSEK
Evaluation Advisor
WBI Room G 3192,The World Bank
1818 H Street, N.W. Washington, DC 20433
U. S. A
Email: Jkusek@worldbank.org

OECD

Mr. Niels DABELSTEIN
Head of Evaluation Secretariat
DANIDA/OECD/DAC WP
Danish International Development Assistance,
Ministry of Foreign Affairs
2, Asiatisk Plads, DK 1448 Copenhagen,
Denmark
Tel: 45.33.92. 00.39/92.00.00
Fax: 45.33.92.16.50
Email: Niedab@um.dk

Mr. Shoko KIMURA
Evaluation Division,
Economics Cooperation Bureau,
Ministry of Foreign Affairs, OECD/DAC, Japan
2-2-1, Kasumigaseki, Chiyoda-ku
Tokyo 100, Japan
Tel:81 3 35 92 83 42
Fax:81 3 35 93 80 21

Ms. Fiona SHERA
Economic Advisor
DFID, United Kingdom
Tel: (44) 171 917 0545
Fax: (44) 171 917 0561
Email: F-Shera@dfid.gov.uk

ASIAN DEVELOPMENT BANK (ADB)

Mr. Peter C. DARJES
Operations Manager
Asian Development Bank
Address:
Tel: (63-2) 632 4111
Fax: (63-2) 636 2161
Email:pdarjes@mail.asiandevbank.org

CONSULTANTS

Ms.Annette BINNENDIJK
P.O. Box 146
Basye, Virginia 22810 USA
Tel: (540) 856 2624
Email: annetteb@shentel.net

Mr. Leonard J. EARLY
Department of Finance, Australian Public Sector
33 Bringley Cr. Fraser ACT, 2615, Australia
Tel: (61 2) 6259 5390
Fax: (61 2) 6259 5382
Email: learly@pcug.org.au

Mr. Marc HOLZER
Executive Director, National Center for Public Productivity, Rutgers University, NJ, USA
7th floor Hill Hall 360 King Blvd
Newark, NJ 07102
Tel: (973) 353 1351/5504 Ext.23
Fax: (973) 353 5907
Email: mholzer@pipieline.com

Mr. Adil KHAN
UNDP Senior Advisor
Taj Samudra Hotel
25, Gale face Centre Road
Colombo 3
Sri Lanka
Tel: (00941) 889190
Fax: 94.1. 446348/330509
Email: akhan@sltnet.lk

Mr.Rustam LALKAKA
President, Business & Technology Development Strategies
1441 Third Avenue, Suite 5A
New York, NY 10028, USA
Tel: (212) 452 1678
Fax: (212) 535 0060
Email: rlalkaka@earthlink.net

Mr. Jacques TOULEMONDE
Director
Center for European Evaluation Expertise
C3E, 13 bis Place Jules Ferry-F69006
Lyon, France
Tel: 33 (0) 4 72 83 78 80
Fax: 33 (0) 4 72 83 78 81
Email: toulemonde@c3e.fr

ANNEX II: Conference Agenda

International Conference On Evaluation Capacity Development
Beijing, People's Republic Of China,27-28 October 1999

Wednesday 27 October 1999

8:00- 9:00 **Registration**

MORNING SESSION

09:00- 10:00 **Opening Ceremony:**

Session Chair:Mr. Li Xinnan, Vice President, Chinese National Center for Science and Technology Evaluation, RPC (NCSTE)

Opening Speech: Ms. Deng Nan, Vice Minister, Ministry of Science & Technology

Address:

- Mr. Khalid Malik, Director, UNDP Evaluation Office
- Ms. Ma Dexiu, Director, Industry Department, State Commission for Development & Planning
- Mr. Ray Rist, Senior Advisor, World Bank Institute
- Mr. Zhan Jingtao, Director, Public Expenditure Department, Ministry of Finance
- Mr. Yu Liegui, Director, Industry & Transportation Department, Ministry of Finance
- Mr. Wang Jianzeng, Director, Science & Technology Department, State Commission for Economy & Trade

10:00 - 10:15 **Launching of the Handbook "Result-Oriented Monitoring and Evaluation"**

Mr. Qi Rang, Director, Development & Planning Department, Ministry of Science & Technology

Mr. Khalid Makik, Director, Evaluation Office, UNDP

10:15 - 10:30 Break

THE EVOLVING CONTEXT FOR EVALUATION CAPACITY DEVELOPMENT
Session Chair:Mr. Ray Rist,Advisor, World Bank

10:30 -12:15 *Statements:*

The New Dimensions in ECD in International Development
Mr. Khalid Malik, Director, UNDP Evaluation Office

Background paper: **The Asian Perspective in Evaluation Capacity Development: The Challenges of the New Millenium**
Mr. Adil Khan, Senior Advisor, M&E UNDP/UNOPS, Sri Lanka

Background paper: **The S&T Evaluation in China: Practices and Roles**
Ms. Chen Zhaoying, Vice President of NCTSE

Background paper: **Aid Evaluation/A donor perspective**
Mr. Niels Dabelstein, Head of Evaluation Secretariat DANIDA, OECD/DAC WP:

Questions and Answers

12:15-12:30 **Objectives of the Conference**
Mr. Antonio Molpeceres, Learning Resource Advisor, UNDP

12:30-14:00 Lunch

AFTERNOON SESSION: PERFORMANCE MANAGEMENT IN THE PUBLIC SECTOR
Session Chair:Representative of Ministry of Finance, PRC

14:00-14:45 *Background papers:* **Results Based Management:A Donor Agency Perspective**
Ms. Annette Binnendijk, Consultant to the Evaluation Office, UNDP

Respondents: Mr. Fang Yan, Chief of Project Evaluation, NCTSE

National Delegations: Mongolia, India

Questions and Answers

14:45-15:30 *Background paper:* **Linking Evaluation to Policy Formulation and Budget Processes**
Mr. Ray Rist, Senior Advisor, World Bank Institute

Respondents:
Mr. Len Early, Department of Finance & Administration, Australian Public Sector
Ms. Xu Yaoling, Vice President, NCSTE

Questions and Answers

15:30-15:45 Break
Announcement for reception

Session Chair:Mr. Wang Jianxin,Director of Evaluation Division, Planning Department,Ministry of Science & Technology, RPC

15:45-17:30 *Background paper:* **Evaluation Capacity Building in the People's Republic of China**
Mr. Peter C. Darjes, Asian Development Bank

Respondents: Mr. Zhao Lu, Director of Science Division, Public Expenditure Department, Ministry of Finance

National Delegation: Sri Lanka, Fiji

Background paper: **Public Performance Evaluation and Improvement**
Mr. Mark Holzer, Executive Director, National Center for Public Productivity, NJ, USA

Respondents:
Ms. Jody Kusek, Advisor, World Bank Institute
Mr. Gu Wenxing, President, Shanghai Science Research Institute

Questions and Answers: Wrap up UNDP Evaluation, Fang Yan, Lalkaka

18:30 - 21:00 Reception given by UNDP Evaluation Office

Thursday 28 October 1999

MORNING SESSION :
A) MONITORING AND EVALUATION IN A RESULTS-BASED MANAGEMENT APPROACH
B) EVALUATION INITIATIVES IN SCIENCE AND TECHNOLOGY
Session Chair:Mr. Khalid Malik,Director, UNDP Evaluation Office

09:00-9:45 *Background paper:* **M&E and RBM: Balancing Learning and Accountability Concerns,the UNDP Experience**
Ms. Christine Roth, Evaluation Advisor, UNDP Evaluation Office

Respondents: Mr. Dai Guoqing, Director, Finance Division, Conditions Finance Dept., Ministry of Science & Technology

National Delegations: Kyrgyzstan

Questions and Answers

9:45-10:30 *Background paper:* **Evaluation Partnerships**
Mr. Niels Dabelstein, Chairman Working group on Aid Evaluation, OECD/DAC

Respondents: Br. Yang Sai, Director, Beijing Kehuan S&T Development Center

National Delegation: Vietnam, Pakistan

Questions and Answers

10:30-10:45 Break

Session Chair:Ms.Li Maoming, Senior Advisor, NCSTE

10:45-11:30 *Background paper:* **Evaluation Tools and Methodologies:Case Studies**
Mr. Jacques Toulemonde, Director, Center for European Evaluation Expertise

Respondents: Mr. Du Zhanyuan, assistant Director, Development & Planning Department, Ministry of Development & Planning Department, Ministry of Science & Technology

National delegation: Russia, Malaysia

Questions and Answers

11:30-12:30 *Background paper:* **Evaluation Initiatives in Science and Technology Programs**
Mr. Rustam Lalkaka, former Head of UNFSTD

Respondents: Mr. Wang Jianxin, Director, Evaluation Division, Development & Planning Dept., Ministry of Science and Technology

National delegation : Korea

Background paper: **Review of Draft of Science & Technology Evaluation Standard**
Ms. Chen Zhaoying, Vice President, NCSTE

Respondents: Mr. Zhang Zheng, Director, Wuhan Science & Technology Evaluation Center

National delegation: Kazakstan

Questions & Answers

12:30-14:00 Lunch

AFTERNOON SESSION: ECD OPPORTUNITIES FOR NATIONAL INITIATIVES AND INTERNATIONAL COOPERATION
Session Chair:Mr. Ray Rist,Advisor, World Bank Institute

14:00-14:45 **ECD opportunities for National Initiatives**
Video Speech: Guide to Diagnosing ECD Needs:
Mr. Keith McKay, Senior Evaluation Officer, World Bank

Questions and Answers

14:45-15:45 *Group Work:* **National needs and planned steps for ECD**
National delegations, UNDP, World Bank and Asian Development Bank as facilitators

15:45-16:15 **Current trends and future steps for ECD at the national level**
Presentations by rapporteurs of working groups

16:15-17:00 Break

Session Chair:Ms. Chen Zhaoying, Vice President,NCSTE

17:00-17:15 **Meeting wrap-up:** Ms. Christine Roth, Evaluation Advisor, UNDP Evaluation Office

17:15-17:45 **Concluding remarks:**

- Mr. Ming Tinghua, Chairman, Invention Association of China (former Director of Chinese Patent Office)
- Mr. Ray Rist, Senior Advisor, World Bank Institute
- Mr. Yang Qiwen, Director, Basic Research Department, Ministry of Science & Technology
- Mr. Khalid Malik, Director, UNDP Evaluation Office
- Mr. Li Xinnan, Vice President of NCSTE, Vice President of NRCSTD

Friday 29 October, 1999

18:30-21:00 **RECEPTION HOSTED BY THE NCSTE: CLOSING CEREMONY, GREAT HALL OF THE PEOPLE**
Chair:Ms.Li Maoming, Senior Advisor, NCSTE

Presence of senior officers
Mr. Zheng Guoan, Director, Ministry of Science & Technology

UNITED NATIONS DEVELOPMENT PROGRAMME

© Evaluation Office, 2000

Evaluation Office
United Nations Development Programme
One United Nations Plaza
New York,NY 10017, USA
Tel. (212) 906 5095, Fax (212) 906 6008
Internet: http://intra.undp.org.eo
or http://www.undp.org/eo