unicef

From policies to results
Developing capacities for country monitoring and evaluation systems

In partnership with:

 DevInfo

IDEAS

International
Labour
Office

THE WORLD BANK

UN
D
P

United Nations Development Fund for Women

UNIFEM

WFP

wfp.org

unicef

From policies to results
Developing capacities for country monitoring and evaluation systems

In partnership with:

DevInfo

IDEAS

International Labour Office

IOCE

THE WORLD BANK

UNDP

United Nations Development Fund for Women
UNIFEM

WFP
wfp.org

ISSUE #1:   New trends in development evaluation, 2007. Published jointly with IPEN.

ISSUE # 2:   Bridging the gap: The role of monitoring and evaluation in evidence-based policy making, 2008. Published in partnership with the World Bank, IDEAS, DevInfo and MICS

ISSUE # 3:   Country-led monitoring and evaluation systems. Better evidence, better policies, better development results, 2009. Published in partnership with the World Bank, IDEAS, IOCE, UNECE, DevInfo and MICS

ISSUE # 4:   Country-led monitoring and evaluation systems. Watch and listen international keynote speakers, 2009. DVD published in partnership with IDEAS, IOCE, WFP, OECD/DAC Network on development evaluation and DevInfo

Disclaimer

From policies to results

Developing capacities for country
monitoring and evaluation systems

EDITOR

Marco Segone

AUTHORS

Vaidehi Anushyanthan
Michael Bamberger
Angela Bester
Shanelle Boyle
Manuel Fernando Castro
Sekouba Diarra
Diego Dorado
Osvaldo Feinstein
David Fetterman
Oscar A. Garcia
Enrico Giovannini
Jon Hall
Billy C. Hamilton
Philomen Harrison
Caroline Heider
Douglas Hsu
Oumoul Khayri Ba Tall
Jody Zall Kusek
Alexey Kuzmin
Saraswathi Menon

Linda G. Morra Imas
Yakob Mudesir Seid
Rosetti Nabbumba Nayenga
Indran Naidoo
Robert Ndamobissi
Joe Nuwamanya Kanengere
Finbar O'Brien
Rita O'Sullivan
Kris Oswalt
Hallie Preskill
Nicolas Pron
Jean Quesnel
Michael Quinn Patton
Ana Carolina Ramirez
David Rider Smith
Ray Rist
Craig Russon
Belen Sanz Luque
Marco Segone
Velayuthan Sivagnanasothy

Contents

Part 2:
Developing country monitoring and evaluation capacities

Strengthening institutional capacities and an enabling environment for country-led monitoring and evaluation systems

Strengthening individual monitoring and evaluation capacities

Part 3:
Good practices and lessons learned from country-led monitoring and evaluation systems

Annexes

Preface
by Director of Evaluation, UNICEF

As Director of Evaluation at UNICEF, it is my pleasure to write a preface for the third publication in this important series that addresses country-led monitoring and evaluation systems.

Within the United Nations system, there is a growing recognition that national capacity development for monitoring and evaluation systems is an essential part of the broader support to policy reform and to promoting national ownership of evidence-based policy making. The United Nations also recognizes that its own efforts at country level are best assessed by a country's own evaluation system and is encouraged to see a series of country-led evaluations taking place that assess how the United Nations is Delivering as One at country level. Supporting country-led evaluations is, in itself, an important way of building national capacity for evaluation.

UNICEF continues to be committed to supporting national capacity development and is encouraged to see that a task force of the United Nations Evaluation Group is currently focused on this issue. We look forward to continuing collaborations and work with major partners at global, regional and country levels to help achieve the goal of developing strong national systems that will enhance sustainable development results.

The previous publications in this series, *Country-led monitoring and evaluation systems* and *Bridging the Gap – the role of monitoring and evaluation in evidence-based policy making* have already made important contributions to the debate on how monitoring and evaluation can enhance evidence-based policy making. This publication furthers the debate by providing new analysis on strengthening country evaluation capacities, and in moving from policies to results through country-led monitoring and evaluation systems.

Finbar O'Brien

Evaluation Office
UNICEF

Preface
by IDEAS President

I am pleased to be able to write this Preface for this book and on this topic. UNICEF is to be commended for its continued emphasis on development evaluation in general, and on evaluation capacity development (ECD) in particular. Indeed, in 2009, the International Development Evaluation Association (IDEAS) held its last Global Assembly in Johannesburg, South Africa, on just this topic. ECD is a pressing concern as governments seek to build the knowledge and capacity to start to track their own performance, and to strengthen their governance efforts.

What has become clear in the past decade, and is also reflected in this book, is that capacity building needs to take place at both the organizational and individual levels. They are both necessary and both are required if quality information is to be collected, analyzed, and used for evidence based decision-making. Strong organizations need strong staff, if good monitoring and evaluation (M&E) systems are to become a reality.

Good evaluation data are different from, and more than, monitoring data. Indeed there is often confusion between monitoring and evaluation data. Both are important, but indeed, within organizations, monitoring often crowds out evaluation. The emphasis here, in this book, is as it should be, on building evaluation capacity. It is perhaps for another book in this series to address the fundamental relations (strengths and weaknesses) between the "M" and the "E".

Again, UNICEF is to be congratulated for its continuing efforts to bring such materials to the evaluation community. I am personally pleased to see this book being published.

Ray C. Rist

President
International Development
Evaluation Association
(IDEAS)

Preface
by IOCE President

Evaluation is getting increased attention and interest around the world. This is because it is believed to have the potential to provide sound answers to 'enduring questions' facing humanity such as, "How to make policies work better?", "What interventions lead to success?" and "How can success be identified, sustained and replicated?"

The intention of this issue, building on the previous series, is to explore further how monitoring and evaluation are expected to lead to results. "Results" implies improvements to the lives of people, as that is the 'normal' expected outcome of a public policy. A successful policy would be an 'intelligent' one in the sense that it is based on or influenced by 'quality' information, just as a tree is nurtured by 'quality' water and ingredients to provide healthy leaves and fruits. Policy design and evaluation processes are therefore intimately related, to enable the latter to feed back into the former. As development questions are context related, so are policies, and hence their evaluation.

Monitoring and evaluation of policies would help understand why changes do not always happen as expected, and/or what happens instead. This ambitious aim is possible only if the evaluation process is capable of integrating the 'multiple' lenses needed to understand what happened, and to tell the story in a meaningful way to allow learning and corrective action to follow on. This apparently simple statement implies methodological challenges in the evaluation design and process, within different contexts. It is therefore fundamental that strong national evaluation systems exist to question national development policies, though that is hardly the case in most parts of the developing world. Contributions in this issue illustrate the importance of strengthening national evaluation capacities and of exploring strategies for doing that. Without having strong national evaluation systems in place, development results will not be sustainable, and evaluation will mainly serve external needs – as evidenced by several studies.

IOCE, like other evaluation networks, is dedicated to promote the use of evaluation to improve the effectiveness of public policies, and to help strengthen national evaluation capacity where it is needed. For that, we aim to provide an international platform where evaluators from all over the world multiply the opportunities to network, learn from each other, access evaluation resources, and engage in debates and research to advance evaluation theory and practice. More ambitiously, and despite our limited resources, we also aim to build a global community capable of channelling a diversity of voices, views and knowledge in the ways important questions are being tackled by the international development stakeholders.

I would like to extend my warm regards to the authors, with a special mention to Marco Segone and UNICEF for their enormous support to evaluation.

Oumoul Khayri Ba Tall

President
International Organization
for Cooperation in Evaluation
(IOCE)

Editorial

This publication aims to contribute to the international debate on moving from policies to results through country-led monitoring and evaluation systems, by providing new analysis on strengthening country monitoring and evaluation capacities.

It does so by offering a number of strong contributions from senior officers in institutions dealing with national monitoring and evaluation capacity development. These are: UNICEF, the World Bank, UNDP, WFP, UNIFEM, ILO, OECD, the International Development Evaluation Association (IDEAS); and the International Organisation for Cooperation in Evaluation (IOCE), as well as senior Government representatives responsible for the national monitoring and evaluation systems in countries in Asia, Africa and Latin America.

From policies to results through country-led monitoring and evaluation systems

In the last decade, several countries have reformed their social policies. However, at times these policy reforms did not deliver the expected results to the target population, including children and women. It seems that the main challenge resides in implementing policy reforms, rather than in designing and adopting them.

This publication suggests that country monitoring and evaluation systems can play a central role in producing the relevant evidence to enhance the capacity of national social protection systems to implement policy reforms, by identifying bottlenecks and providing insights into the effectiveness of the strategies used to implement policy reforms.

However, in certain cases countries do not have the needed capacity to produce relevant and sound evidence. National monitoring and evaluation capacity development is therefore a central strategy to enhance country capacity to assess more effectively the progress of policy reform implementation, through use of evidence generated by country-led monitoring and evaluation systems.

Segone introduces the theme of this publication describing how country-led monitoring and evaluation systems can provide support for addressing the policy implementation challenge. He states that there is wide consensus among countries, donors and international agencies, that systems to monitor and evaluate national development strategies, including policy reforms implementation, should

be led and owned by countries, while international agencies and donors should support the strengthening of country monitoring and evaluation capacities. The second part of Segone's article proposes a conceptual framework for country monitoring and evaluation capacity development, explaining that it should be context-specific, address supply as well as demand capacities, and be comprehensive by strengthening individual and institutional capacities while enhancing an enabling environment.

In this context, **Quinn Patton** presents the major trends which he expects will shape evaluation over the next quarter century, including: increased international and cross-cultural expansion of evaluation with greater globalization and diversity; increased political interest in accountability, performance indicators and transparency; and, growing emphasis on evaluation capacity-building and skill development, among others.

Giovannini and **Hall** outline the 'measuring progress of nations' approach which encourages the development of statistical measures which go 'beyond GDP' to provide a comprehensive picture of societal wellbeing. Giovannini and Hall argue that better measures of 'progress of nations' will lead to more effective policy. Some of the potential implementation benefits of a country-led 'measuring progress' initiative include: fostering 'whole-of-government' cooperation; helping with reform challenges; developing an outcome-focused governance culture; and strengthening accountability mechanisms.

Hamilton and **Kusek** explain that public programmes may fail because of poor decision making, as well as lack of effective tools for monitoring the results of decisions made in response to sudden and unexpected needs. Potentially, in any public programme, many things can go wrong. Inevitably, some things *will* go wrong. However, the programme need not to fail if the problems are recognized and dealt with before they begin to coalesce and cascade.

Heider suggests three key evaluations questions to be addressed in policy evaluations: understanding the quality of a policy, the results it has achieved, and how these results can be explained. These questions generate evaluation evidence that is concrete and usable by policy makers and implementers. It discusses the implications of policy evaluations for the evaluation principles of independence, credibility and utility, as well as the challenges embedded in addressing the three key questions.

Russon focuses on policy analysis systems as a special case of country-led monitoring and evaluation systems. He suggests that such systems can help overcome the policy implementation challenge in two ways. First, these policy analysis systems can conduct analyses *for* policies, which will address the implementation challenge. The second way they contribute is by conducting analyses *of* policies, which will attempt to determine the impact of the policy proposals.

Developing country monitoring and evaluation capacities

This chapter demonstrates how to strengthen country monitoring and evaluation capacities, focusing on institutional capacities and the enhancement of an enabling environment for country-led monitoring and evaluation system, while enhancing individual capacities.

Strengthening institutional capacities and enhancing an enabling environment for country-led monitoring and evaluation systems

Menon argues that there has been greater investment in the development of analytical methods and approaches from the perspective of development cooperation rather than in recognizing fully the potential of evaluation as part of governance within countries. Menon explains that this recognition must be rooted on a fuller understanding of national ownership and capacity in evaluation than is normally the case. If so, this recognition will transform the understanding of evaluation itself.

Sanz stresses the role of evaluation to achieve gender equality and women's empowerment. She calls for the need to rethink and enhance the approaches and methods we use so that evaluations contribute to social transformation processes, including through policy reform implementation. The article highlights the key policy/normative frameworks within the United Nations which establish the relevance of human rights and gender equality, and explores emerging evaluation paradigms which mirror some of those frameworks within the evaluation practice. In concluding, she highlights the key parameters of what constitutes the emerging approach of gender equality and human rights responsive evaluation.

Pron, Oswalt and **Hsu** explain that duty-bearers at the international, national and sub-national levels are, more than ever before, seeking access to relevant, reliable and timely data. It's in this

context that the use of DevInfo at the country, sub-national and regional levels, in supporting country-led monitoring and evaluation systems, is increasing around the world. Across Africa, Asia and Latin America, many countries are implementing and institutionalizing DevInfo, within their national monitoring and evaluation frameworks, to provide high quality data for policy-making to duty-bearers across all sectors of human development. This reflects a major drive of duty-bearers towards strengthening their data monitoring, evaluation and assessment capacities, with the goal of ensuring sustained development for all rights-holders.

Quesnel explains that, due to an exponential demand for high quality evaluations, there is a worldwide growing trend toward professionalization in evaluation. Quesnel presents the rationale for professionalization, its key components, the profile of a professional evaluator and the different levels of professional recognition.

Morra focuses on the issue of competencies in relation to the evaluation of development interventions. She explores why competencies are attracting so much attention. Some of the efforts being made by national and regional professional evaluation organizations to delineate evaluator competencies within their specific contexts are briefly described, while focusing on the issue of *global* competencies. She describes an effort to develop international standard competencies in monitoring and evaluation and the challenges and specific issues which that presents.

Bamberger describes the progress that is being made in the transition from individual impact evaluations to building a systematic approach to identifying, implementing, and using evaluations at sector and national levels. When this is achieved, the benefits of a regular programme of impact evaluations, as a tool for budgetary planning, policy formulation and implementation, begin to be appreciated. Although this article focuses on impact evaluation, it is argued that the institutionalization of impact evaluations can only be achieved within the framework of a comprehensive monitoring and evaluation system that provides a menu of evaluations, to cover all the information needs of managers, planners, and policy-makers.

Feinstein analyses the implications of the growing partnerships for development, the challenges of evaluating partnerships and the importance of partnership in evaluation. He explains that, in the evaluation field, partnerships have been pursued in at least two different ways: to complement resources among development evaluation entities, combining for example the knowledge,

experience and/or reputation of some organizations, with the financial resources of other institutions; and to conduct joint evaluations, establishing "partnerships for evaluation" by several partners pooling financial resources and expertise.

Strengthening individual monitoring and evaluation capacities

Preskill and **Boyle** make the argument that evaluation capacity building should be done in an intentional, thoughtful manner in order to make the best use of resources and maximize the impacts of such efforts. The definitions, examples and guiding questions provided in the article are intended to help evaluation capacity building facilitators not only decide which strategies make the most sense, given the organization's existing infrastructure and resources as well as cultural context, but also to emphasize that the design and implementation of each strategy is critical for successful capacity building.

Kuzmin's article takes a look at the notion of utilization-focused evaluation training, and identifies five groups of factors which affect the use of evaluation training to strengthen evaluation capacities in organizations. These are: factors which cause involvement of organizations in evaluation training; factors related to the utilization focus of the training; factors related to the training itself; factors related to complementary learning that enforces the use of training; and factors related to the sustainability of the training outcomes. Finally, he argues that a good utilization-focused evaluation training should contribute both to the development of evaluation competence as well as to the use of evaluation competence in the ongoing practice.

Quinn Patton explains how being engaged in the process of *utilization-focused evaluation* can be useful quite apart from the findings which may emerge from those processes. If, as a result of being involved in an evaluation, primary intended users learn to reason like an evaluator and operate in accordance with evaluation's values, then the evaluation has generated more than findings. It has been useful beyond the findings in that it has increased the participants' capacity to use evaluative logic and reasoning. "Process use," then, refers to using the logic, employing the reasoning, and being guided by the values that undergird the profession of evaluation.

Fetterman presents *empowerment evaluation* as an evaluation approach designed to enable people to help themselves by internalizing and institutionalizing evaluation as part of their normal day-to-day planning and operations. He explains how empowerment evaluation

helps to build capacity and enhance self-determination, leaving a legacy of evaluation practice behind which contributes to sustainability.

O'Sullivan explains that *collaborative evaluation* systematically invites and engages stakeholders in programme evaluation planning and implementation. Unlike "distanced" evaluation approaches, which reject stakeholder participation as members of evaluation teams, collaborative evaluation assumes that active, on-going engagement between evaluators and programme staff result in stronger evaluation designs, enhanced data collection and analysis, and in results which stakeholders understand and use.

Good practice and lessons learned from country-led monitoring and evaluation systems

The last part of this book documents selected good practice and lessons learned in designing and implementing country-led monitoring and evaluation systems.

Naidoo describes how monitoring and evaluation has evolved in the Public Service in South Africa, highlighting the linkages between monitoring and evaluation and democracy. He describes his experience in managing the monitoring and evaluation unit at the Public Service Commission, which is one of the key players in monitoring and evaluation in the South Africa's Public Service. He also considers the growth of monitoring and evaluation at the continental level.

Castro, Dorado and **Ramirez** provide an overview of the Colombian monitoring and evaluation system's most salient features. They highlight success factors and obstacles encountered during its implementation, and identify a number of lessons which might be of interest to evaluation practitioners and officials implementing monitoring and evaluation systems in developing countries.

Sivagnanasothy and **Anushyanthan** explain how institutionalization of 'management for results' has been instrumental in creating monitoring and evaluation systems and practices in Sri Lanka, by putting in place the foundation for the emergence of the thriving results-focused monitoring and evaluation community in the Government of Sri Lanka. Although still work-in-progress, the initiative has achieved significant progress towards a government-wide shift to results based monitoring and evaluation practices.

Diarra and **Ndamobissi** present the national system Mali has designed and implemented to monitor and evaluate the Poverty Reduction Strategy. The article describes the governance of

the system, including the contribution done by the UN system, as well as good practices, challenges and opportunities.

Rider Smith, Kanengere Nuwamanya and **Nabbumba Nayenga** analyze the political economy of designing and implementing a country-led monitoring and evaluation system. They explain that in order to understand where the investment in monitoring and evaluation has had the biggest impact on public policy and accountability, it is necessary to examine the relationships between policy, institutions and individuals in the public sphere. Evidence from Uganda suggests that the linkage between measurement and policy change has been productive only when the environment in each of these spheres is conducive and interfaced.

Garcia and **Bester** share the lessons learned from the joint country-led evaluation of the role and contribution of the UN system in South Africa, so that partner countries and donors maximise the benefits they can derive from joint evaluations. The lessons learned from the South Africa experience can inform principles and good practice for joint country-led evaluations in other countries.

Mudesir presents the Ethiopian perspective regarding the role of National statistical offices in country-led monitoring and evaluation systems. He advocates for an effort to improve the national monitoring and evaluation system and the generation of statistical data to sustain development processes. This needs the provision of a 'due concern' from the political leadership of the respective country, the development partners, academia and the professionals involved in the system and the statistical data production. This again involves continuous capacity building in the field of statistics; proper funding for the statistical data production; utilization of new technologies; and above all, mainstreaming of the statistics work in the country's development agenda.

Last but not least, **Harrison** highlights how a sub-regional strategy in CARICOM was instrumental to strengthen national capacities to use DevInfo for evidence-based decision making.

I wish you an interesting and inspiring read.

Marco Segone

Editor

Acknowledgement

The editor would like to thank the 40 authors for their excellent contributions to this publication. In sharing their wisdom and experience, and the good practices and lessons learned, they have shown how to overcome the policy reform implementation challenge, by enhancing national capacities for country-led monitoring and evaluation systems. The richness of partners in this publication, which includes the World Bank, UNDP, WFP, UNIFEM, ILO and DevInfo, testifies to the importance and commitment of international organizations to the theme of this book. The partnership with the International Organization for Cooperation in Evaluation (IOCE) and the International Development Evaluation Association (IDEAS), the two global evaluation professional associations, as well as the prefaces by the respective Presidents, Oumoul Khayri Ba Tall (IOCE) and Ray C. Rist (IDEAS), are especially appreciated.

The editor would like to thank Finbar O'Brien, Director, Evaluation Office, UNICEF, for his continuous support to foster the evaluation function within UNICEF and UNEG, and to strengthen national evaluation capacities though country-led evaluation systems. The identification of good practices at country level has been possible thanks to the active engagement and knowledge of the UNICEF Regional Chiefs of monitoring and evaluation in different regions, notably Ada Ocampo, Bastiaan Van't Hoff, Philip Jespersen, Christina Bierring, and Pierre Ngom. Also critical was the support of Ross Smith and Mekonnen Ashenafi, monitoring and evaluation specialists in Uganda and Ethiopia; Violet Speek-Warnery, Deputy Representative in Barbados; and Soma de Silva, Chair of the Sri Lanka Evaluation Association.

Part 1
From policies to results through country-led monitoring and evaluation systems

MOVING FROM POLICIES TO RESULTS BY DEVELOPING NATIONAL CAPACITIES FOR COUNTRY-LED MONITORING AND EVALUATION SYSTEMS[1]

by Marco Segone,
Systemic Management, UNICEF Evaluation Office,
and Former Vice-President and Senior Advisor
to the Executive Board, International Organization
for Cooperation in Evaluation (IOCE)

Introduction

Within the social policy reform debate occurring in several countries, much attention has been given to policy advice and formulation, as well as policy (and budget) decision-making. However, it appears that the real challenge is in implementing policy reforms to "translate" policy statements into development results for vulnerable populations, including children and women. Strengthening national social systems to implement policies is therefore paramount. For this, a strong country monitoring and evaluation system is crucial to provide essential information and analysis. A strong system will help to ensure such policies are being implemented in the most effective and efficient manner; to review policy implementation and design; and, to detect bottlenecks and to inform on adjustments needed to enhance systemic capacities to deliver results.

While more and more countries are designing and implementing national monitoring and evaluation systems, technical capacity to develop monitoring and evaluation systems that meet international quality standards is often weak. Therefore, national strategies to strengthen monitoring and evaluation capacities are needed. These

strategies should be comprehensive and integrated, addressing both the technical and political side, as well as the three different levels of capacity development: individual, institutional and the enabling environment.

This article is a contribution to strengthen country capacities to deliver results by enhancing national social systems through monitoring and evaluation capacities. The article is divided into two parts. The first one focuses on describing how country-led monitoring and evaluation systems can support action to address the policy implementation challenge; the second part presents a conceptual framework for strengthening national monitoring and evaluation capacities.

Monitoring and evaluation systems to address the policy implementation challenge

While the importance of designing good social policies is unquestionable, implementing them effectively may prove challenging. More specifically, although formal adoption of policy reforms is a critical milestone in the overall reform process, real change does not occur until policies are implemented. Given the nature of politics, actual implementation of key reforms is often a lengthy, multi-stage process. Certain groups that benefit from the status quo may actively oppose reform implementation and try to halt the process. The organization charged with carrying out a given policy reform is often different from the one that has adopted the policy through legislative or executive action. The implementing organization may lack the necessary funds; human resources; authority; motivation; capacity; or, experience to implement the new policy. Finally, the leadership of the government, ministry, or implementing organization may change during the implementation process. That can affect the political will for implementing the policy reform.

A well-constructed monitoring and evaluation system can be used on an ongoing basis to direct discussion and examine delays and other challenges impeding policy implementation. Such a system also facilitates brainstorming on approaches to address these challenges. It stimulates feedback to help managers assess strategies for introducing and implementing reforms, and to assess the effect of those reforms on a country's critical sectors. When a monitoring and evaluation system is used in this way, it can foster ownership of the reform process by stakeholders, and increase their commitment to implementation.

Approaches to monitoring and evaluating policy implementation

Policy implementation milestones

The complexities discussed above reveal a number of common milestones in the implementation process. These milestone events underpin a monitoring and evaluation system that can closely track and assess where a policy is, in the implementation process. They can highlight challenges or opportunities which may allow managers to take actions to enhance the chance of effective implementation of a given reform. Milestones in the process of implementing a reform include the following:

- Achieving ownership and support of the policy reform within various levels of government and civil society through dialog, lobbying, and public discussion.

- Completing a detailed policy implementation plan.

- Achieving political compromises among stakeholder groups who stand to lose power or resources from implementing the new policy.

- Disbursing appropriate financial resources to implement the policy.

- Carrying out organizational changes, including the ones affecting power relations.

- Increasing technical and managerial capacity where needed.

These events rarely occur in sequential order. For example, dialog between stakeholders may have to be ongoing. Legislation may have to be redrafted if it is incomplete, ineffectual, or unacceptable to some key stakeholder groups. Advocacy groups may need to continue lobbying for a given reform throughout the process.

Therefore, monitoring and evaluation systems should be flexible enough to accommodate the iterative nature of policy reform. Because some of the milestone "events" listed above either occur repeatedly, or function more as processes than events, policy reform monitoring and evaluation systems should build in, where appropriate, periodic checks on the status of these events. For example, governments may want to design their monitoring and evaluation systems to note the first time a given milestone event occurs *and* to track whether and to what extent the "event"

continues, and to assess its contribution to the policy implementation. This type of system gives managers a more detailed and realistic picture of the policy reform implementation process.

Setting annual targets and assessing performance

Predicting year-to-year progress on the implementation of a given policy reform can be tricky. Unlike predicting progress in many other programme areas, one cannot look at a trend line of past performance or rely on scientifically grounded assumptions to predict the rate of change from year to year. When thinking about projecting movement toward policy implementation, a number of country-specific factors will affect the reform effort and the speed at which it is implemented.

Therefore, before setting policy reform annual performance targets, the following should be considered:

- The degree of political support for the policy within the current government, as well as by other groups and individuals affected by the reform.

- The resources available to lobby for, or to block, the reform.

- The amount of political power held by the probable "winners" and "losers" of the reform.

- The degree of political stability within the country.

- The availability of financial resources to implement the reform.

- The organizational capacity of those charged with implementing the reform.

This is not a comprehensive list, but one which represents considerations which should be taken into account when assessing the rate of the reform progress and the likelihood of its success. Judgments will be better informed through updated political assessment of the country and of the public and private institutions and organizations that will be involved in the reform process. Because political situations in any country are subject to rapid change, critical assumptions about the timetable proposed for completing the reform effort should be identified and stated.

Annual performance targets should be set by policy stakeholders from various constituent groups in the public and private sector. Given the political nature of reforms, it is vital that key stakeholders participate in this process, throughout implementation, in order to identify the stages through which a policy must proceed, and to predict the time required to reach each stage. The result will be a more realistic monitoring and evaluation system relevant to the specific reforms undertaken. Convening such a group to help create the monitoring and evaluation system and to set annual targets will also help when the time comes to assess progress.

Country-led monitoring and evaluation systems

As mentioned above, it is paramount that the system to monitor and evaluate implementation of policy reform is owned and led by the key stakeholders within the country, and not by external donors or agencies. Country-led monitoring and evaluation systems are systems in which the country (not the donors) leads and owns the monitoring and evaluation process, by determining:

- what elements of the policy or programme will be evaluated;

- what evaluation questions will be asked;

- what methods will be used;

- what analytical approach will be undertaken;

- how the findings will be communicated; and,

- ultimately how findings will be used.

Country-led monitoring and evaluation systems serve the information needs of the country and, therefore, are an agent of change and instrumental in supporting the implementation of policy reform. This is possible because, being country-led, they build on the culture and values, as well as the political processes, of the country. Therefore, international organizations and donors should always make sure their own monitoring and evaluation work strengthens rather than undermines national monitoring and evaluation systems and capacities.

It should be noted that country-led monitoring and evaluation systems are "country" led, i.e., not led exclusively by central governments. Local authorities and civil society should also be involved and contribute. For example, civil society organizations could play a key role in monitoring and evaluating the performance of policy reform implementation through different means which allow them to articulate their voice. In this context, professional evaluation organizations have a potentially significant role to play. In the last 15 years this potential has grown as the number of national and regional professional evaluation organizations has increased dramatically, passing from a dozen in 1997 to more than 100 in 2010. Most of the new organizations are located in middle income and developing countries.

Ownership and evaluation capacity development: the key ingredients to country-led monitoring and evaluation systems

National ownership is one of the best strategies to ensure policy relevance – and therefore use – of evidence. National capacity development is needed to enhance the technical rigour of evidence. The ownership principle endorsed in the *Paris Declaration*, and reiterated in the *Accra Agenda for Action*, states that countries should exercise effective leadership over their development policies and strategies. Donors are responsible for supporting and enabling country ownership by respecting the country's policies and systems, and helping the country to strengthen their capacity to implement them.

The implication for the monitoring and evaluation function is fundamental. The principle of ownership means that countries should own and lead their own national monitoring and evaluation systems. Donors and international organizations should support development of national monitoring and evaluation capacities to ensure their sustainability and that the information and data produced are in compliance with monitoring and evaluation standards. This support should take into consideration the value of diversity in evaluation approaches.

Towards a conceptual framework to strengthen country monitoring and evaluation capacities[2]

In addition to the arguments given above, there is strong demand by UN Members States for the development of country monitoring and evaluation capacities. The UN General Assembly stated that national governments have the primary responsibility for coordinating external assistance and evaluating its contribution to national priorities. It therefore mandated the United Nations system to promote national ownership and capacity development, as well as to make system-wide progress in harmonizing evaluation practices. The Executive Boards of UN agencies, including UNICEF, UNDP and UNFPA, emphasize that programme countries should have greater ownership and leadership in the evaluation of all forms of assistance. They underlined, in particular, the importance of increasing the participation of national counterparts and strengthening national capacities in evaluation. The 2009 UNICEF Executive Board encouraged UNICEF to use country monitoring and evaluation systems, where available, to continue to incorporate capacity-building mechanisms into programme design and implementation, and to ensure that evaluations are responsive to national demand.

Based on the above, it can be said that there is a wide consensus among countries, donors and international agencies, that systems to monitor and evaluate national development strategies, including policy reforms implementation, should be led and owned by countries. International agencies and donors should support the strengthening of country monitoring and evaluation capacities. Thus, the two crucial questions are: what do we mean by capacity development, and what should a strategy to develop country monitoring and evaluation capacities look like? The following sections are an attempt to address these questions.

---

2   While the conceptual framework presented below is relevant to monitoring and evaluation, specific details have been developed focusing on the evaluation function. This is particularly important when implementing policy reform in developing countries, where policy reforms are sometimes not truly country-owned, and country-led evaluation systems may offer the evidence needed to challenge the relevance of policy reforms within a specific country context.

A context-specific capacity development

Capacity might be defined (Segone, Patel, Rouge and Russon, 2006) as the ability to define and realize goals, where defining goals entails identifying and understanding problems, analyzing the situation, and formulating possible strategies and actions for response. *Capacity development* is about creating conditions which support the appropriate actors in assuming the appropriate roles in this process of identifying problems and defining and realizing goals.

The above view encourages the use of a *systems approach* to identify, define and achieve goals. This means that it is necessary not only to look at actors at different levels and across sectors but also, crucially, to look at the network of relationships or connections between them. Such a viewpoint illustrates the fact that weaknesses in capacity at any level or with any key actor, whether at the community level, nationally, or somewhere in between these levels, will affect the capacity of the whole system to deal with a problem in order to achieve a goal.

In addition, capacity must be understood in terms of a specific cultural, social and political context. Capacity must be understood as something that exists in degrees at all levels of society – community/national, individual, household, institutional and system. Capacity can exist without outside intervention, though it may be constrained. This implies that one must first understand capacities or elements of capacity that already exist before engaging in any effort to build on or strengthen them.

Intentionality of capacity development

Capacity development implies intentionality to strengthen capacities. For example, technical assistance may strengthen country capacities only if it's clearly oriented towards this aim, by engaging country stakeholders in the monitoring and evaluation process, creating space for learning by doing (i.e. in joint country-led evaluations), and facilitating access to knowledge. A capacity development-focused technical assistance is therefore demand-driven rather than supply-driven, and it focuses on the outcomes rather than the inputs of technical assistance.

Figure 1: The shift in paradigm to a capacity development approach
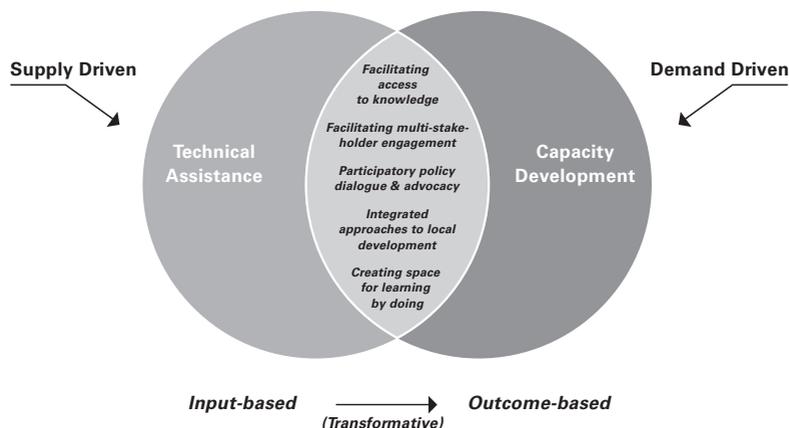
Source: UNDG, 2009. A collective approach to supporting capacity development

An integrated approach addressing supply as well as demand capacities

A distinction should be made between the capacity of policy-makers/policy advisors to use evidence and the capacity of monitoring and evaluation professionals to provide sound evidence. While it may be unrealistic for policy-makers/policy advisors to be competent doers of monitoring and evaluations, it is both reasonable and necessary for such people to be able to understand and use evidence produced by monitoring and evaluation systems in their policy practice. Integrating evidence into practice is a central feature of policy-making processes. An increasingly necessary skill for professional policy-makers is to know about the different kinds of evidence available; how to gain access to it; and, how to critically appraise it. Without such knowledge and understanding it is difficult to see how a strong demand for evidence can be established and, hence, how to enhance its practical application.

However, it is also important to take into consideration that the design and implementation of policy reform is a political process, informed by evidence. The use of evidence in policy reform and implementation depends on the combination of capacity to provide quality and trustworthy evidence on the one hand, and the willingness and capacity of policy-makers to use it on the other. The extent to which evidence is used by policy-makers depends, in turn, on the policy environment.

To strengthen an enabling policy environment, policy-makers may need incentives to use evidence. These include mechanisms to increase the 'pull' for evidence, such as requiring spending bids to be supported by an analysis of the existing evidence-base, and mechanisms to facilitate evidence-use, such as integrating analytical staff at all stages of the policy implementation.

Civil society organizations may also play a major role in advocating for the use of evidence in policy implementation. Think-tanks, with the support of mass media, may also make evidence available to citizens, and citizens may demand that policy-makers make more use of it.

Figure 2: Capacity development framework addressing the demand as well as the supply side
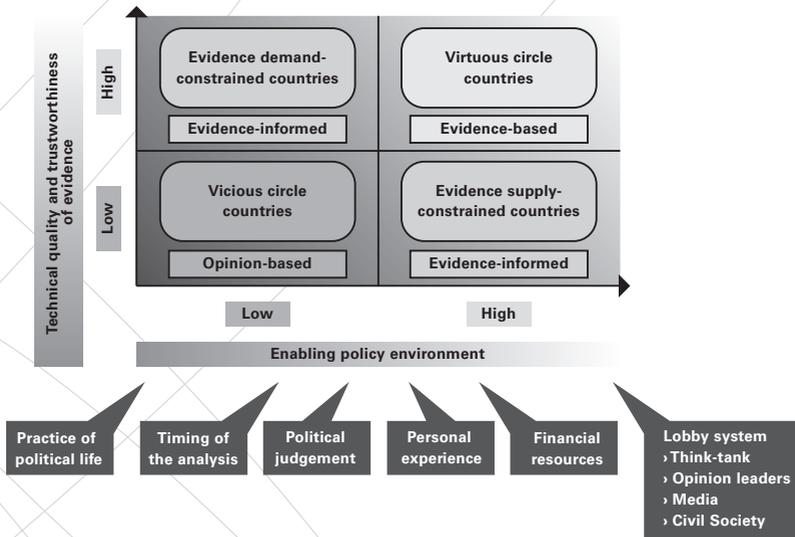
Source:Adapted from Segone, 2009, Enhancing evidence-based policy-making through country-led monitoring and evaluation systems

It is therefore clear that a strategy for developing country monitoring and evaluation capacities should be tailored to the situation and context in which a specific country finds itself, as illustrated in figure 2 above, and detailed below:

- *Vicious circle countries.* Evidence is technically weak and policy-makers have little capacity to make use of it. Policy decisions are mainly taken on political judgment and personal experience,

which may result in poor policy design and, consequently, poor results. In this case, it is necessary to adopt measures which will simultaneously increase both the demand and supply of evidence, as well as improve the dialogue between producers and users of evidence.

- *Evidence supply-constrained countries.* Although evidence is technically weak, it is increasingly demanded by policy-makers. However, evidence deficiency reduces the quality of decision-making and therefore the quality of services delivered. Policy-makers are likely to resent being held to account on the basis of inadequate evidence. Therefore, the priority should be to adopt measures to increase the quantity and quality of evidence. The challenge is to strike a balance between quickly generating improvements to evidence, while laying the foundations for better performance of the country monitoring and evaluation system in the long-run.

- *Evidence demand-constrained countries.* The quantity and quality of evidence is improving, but it is not demanded for decision-making because policy-makers lack the incentives and/or capacity to utilize it. Policy-makers are likely to be at the very least wary of (or may even actively dislike) having more and better evidence pushed at them when it may not support decisions they have already taken or wish to take. In this case, priority should be given to the adoption of measures to increase the demand for evidence, as well as to improve the dialogue between producers and users of data.

- *Virtuous circle countries.* Evidence is technically robust and is being used increasingly for decision-making. The production of good (or at least improved) evidence is matched by its widespread (or at least increased) use in decision-making. These two processes are mutually reinforcing, resulting in better policy design and implementation, and ultimately, better results.

The virtuous circle countries situation is the goal which should be set, since it provides a useful benchmark against which to compare the other three cases. Developing a culture of evidence-based policy-making is a slow process which may take years, but the potential rewards are worth the effort. Where this situation is approximated in practice, it is clear that good evidence is an integral part of good governance. Essential elements for sustaining a virtuous circle linking monitoring and evaluation professionals to policy-makers are: strengthening the democratic process by requiring

transparency and accountability in public sector decision-making, and establishing clear accounting standards and an effective regulatory framework for the private sector.

A systemic approach to country evaluation capacities development

In the past, evaluation capacity development focused on strengthening individual capacities. However, it is by now clear that capacity development should be based on a systemic approach that takes into account three major components: individual; institutional; and, external enabling environment.

Figure 3: A systemic and integrated approach to country evaluation capacities development

Demand

Supply

Individual Level

Institutional Level

Enabling Environment

• Tailored to the specific context of each country
• Country leadership and ownership
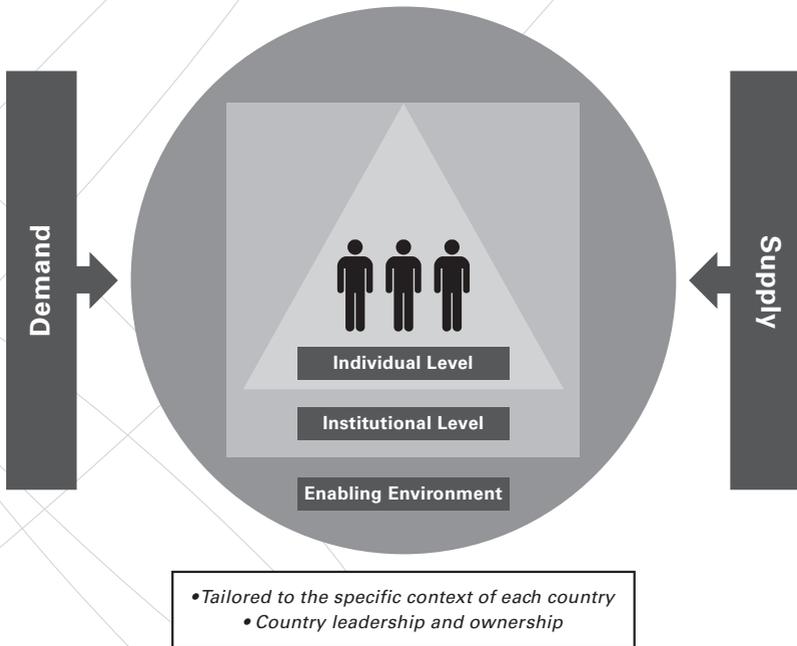
Individual Level

- Senior management capacity to:
    - strategically plan evaluations
    - manage evaluation for independence and credibility
    - use evaluation
- Identify and support leaders or natural champions
- Behavioural independence
- Professional competences

Institutional Level

- Strong evaluation culture
    - evaluative (critical) thinking
    - protective environment
    - managers value and use evaluation
- Evaluation policy
- Set up/strengthen an evaluation unit
- Quality Assurance systems
- Independence of funding for evaluations
- System to plan, undertake and report evaluation findings in an independent, credible and useful way
- Knowledge management systems

Enabling Environment

- Public administration committed to transparency and managing for results and accountability, through results-based public budgeting and evidence-based policy-making
- Legislation and/or policies to institutionalize monitoring and evaluation systems
- Duty bearers have capacity and willingness to be accountable for results
- Rights holders have capacity to monitor policy implementation
- National evaluation professional organization exists
- National evaluation standards and norms

A country strategy for national evaluation capacity development should strengthen the **enabling environment** by supporting *Public Administration commitment to accountability, transparency and managing for results*, including by strengthening results-based public budgeting and evidence-based policy-making.

Legislation and/or policies to institutionalize national evaluation systems should be designed, adopted and implemented. A two-tier strategy should be put in place to strengthen the *capacity of duty-bearers (policy-makers) to demand and use sound evidence* while developing *rights-holders' capacity to demand and to assess policy implementation*, by putting in place systems and mechanisms to engage citizen groups, and to capture and utilize their feedback.

National evaluation organizations should be supported to enable them to foster indigenous demand and supply of evaluation, including by setting national evaluation standards and norms.

Figure 4: Tips for stakeholders to help reinforce an enabling environment for evaluation

Governing bodies (Parliaments, Executive Boards, etc.)

- Be aware of the threats to independence, credibility and utility of evaluation and demand measures to safeguard these principles
- Adopt and oversee the implementation of legislation and/or policies which institutionalize the independence, credibility and utility of evaluation
- Exercise oversight over the quality of evaluation
- Demand evaluation and demonstration of the use of evaluation findings and recommendations

Head of State, Ministers, Policy-makers, Chief executives officers

- Demonstrate leadership in setting a culture of learning and accountability
- Seek and use evidence from evaluations to validate the attainment of goals and objectives and to improve performance whenever possible
- Understand evaluation as part of good governance that aims to use public resources effectively and efficiently to achieve the goals which governments or organizations aim to achieve

Evaluation function

- Demonstrate the value added of independent, credible and useful evaluations
- Raise the awareness of stakeholders of the role and importance of evaluation and of the evaluation principles
- Contribute to evaluative thinking through awareness building, dialogue, and training

Professional evaluation organizations

- Set standards as benchmarks which can be used to convince other stakeholders about the importance of the evaluation principles and measures to safeguard them

Source: Adapted from Heider, 2010, Conceptual framework for developing evaluation capacities

At the **institutional level**, a strategy for evaluation capacity development should strengthen the *evaluation culture* of the organization through institutional commitment to learning from evidence produced by evaluation systems. It should also support evidence-based decisions and demand for accountability, and establish a *protective environment* which removes repercussions on careers. Through a set of values and attitudes supporting *evaluative (critical) thinking* within an organization, individuals are more self-directed learners and use information to act; to take higher risks but, also to develop a greater sense of personal accountability and responsibility; and, to consult, coach, and support each other more. In this context, organizations with a culture of evaluation are able to develop innovative ideas and strategies; change more quickly according to variations in the external environment; and, increase efficiency and effectiveness by systematically using lessons learned to improve programmes and policies. There is less direction from top management and a much more positive attitude and self-accountability at all organizational levels.

An organization with a culture of evaluation has an effective, structured and accepted use of evaluation to support change and development. Managers *value and use evaluation* findings to test out innovation or assess progress towards expected results.

An *institutional framework to institutionalize the evaluation* function should be developed within the organization to ensure evaluation is part of the institution's governance and management functions, as well as to facilitate high-level management understanding of and support for the evaluation function. This should include assistance to develop an *evaluation policy*, which foresees the setting-up or strengthening of a *credible evaluation department*, including endorsement of *evaluation standards* and *quality assurance systems* (including peer review). A *diagnosis of* existing (or, if not existing, of the preconditions to develop it) *evaluation function and/or system*, which would include functional clarity, effective human and financial resources management and robust coordination mechanism,

should be carried out to enable the crafting of a context-specific evaluation policy and system. The diagnostic would identify *bottle-necks* in the policy or system which could be addressed through targeted technical support with the intention of strengthening country capacity.

Heider (2010) states that an institutional framework for evaluation should also secure the *independence of funding for evaluations*, at an adequate level, to ensure that necessary evaluations are carried out and that budget holders do not exercise influence or control over what is evaluated and how. Therefore, funding should be under the direct control of the head of the evaluation function, and should be adequate for a reasonable work programme. In addition, Heider writes that such a framework should also *set out a system to plan, undertake and report evaluation findings in an independent, credible and useful way*. To increase objectivity in the planning and conduct of evaluation, systems are needed which increase the rigor, transparency and predictability of evaluation processes and products. Such systems can include more or less detailed process descriptions or guidelines for the design of evaluations, the conduction of evaluations and reporting on findings.

A good practice is to strengthen *knowledge management systems* in support of the evaluation function. Knowledge has traditionally been fostered at the individual level, mostly through education. However, seen from the perspective of the three levels identified above, knowledge should also be created and shared within an organization through an effective knowledge management system, and supported through an enabling environment of effective educational systems and policies.

**Figure 5: Tips for stakeholders to help develop
an institutional framework for evaluation**

Governing bodies (Parliaments, Executive Boards, etc.)

- Be briefed about the evaluation system so as to understand whether the institutional framework includes adequate checks and balances and to become a discerning user of evaluation reports
- Introduce processes for the selection of the head of evaluation which ensures his/her independence
- Provide secure, separate and adequate funding for evaluation

Head of State, ministers, policy-makers, chief executives officers

- Be briefed about the evaluation system so as to understand whether the institutional framework includes adequate checks and balances and to become a discerning user of evaluation reports

Evaluation function

- Develop and document systems for the selection, design, conduct and reporting on evaluations
- Provide briefings on these standards to increase transparency and confidence in the process and products, and so to enhance credibility
- Develop mechanisms to ensure lessons from evaluation are systematically shared and integrated into debates and decision-making processes.

Professional evaluation networks, associations, etc.

- Set professional standards and good practice standards
- Develop and implement professionalization, accreditation and credentialing systems

Source: Adapted from Heider, 2010, Conceptual framework for developing evaluation capacities

At **individual level**, a capacity development strategy should strengthen *senior management capacity to strategically plan* evaluations (and to identify the key evaluation questions); to *manage* evaluation for independence and credibility; and to use evaluation.

MacKay (2007) underlines the importance of *identifying and supporting leaders or natural champions* who have the ability to influence, inspire and motivate others to design and implement effective evaluation systems. Leadership is not necessarily synonymous with a position of authority; it can also be informal and be exercised at many levels. Therefore, the evaluation capacity development strategy should, especially in the initial stages, identify and support as appropriate, national and local leaders in the public administration, in inter-governmental monitoring, and in evaluation groups and national evaluation organizations.

On the supply side, a capacity development strategy should enhance *behavioural independence* (independence of mind and integrity; knowledge and respect of evaluation standards; agreed evaluation processes and products) as well as *professional competences* through formal education; specialized training;

professional conferences and meetings; on the job training (such as joint country-led evaluations); and, communities of practice and networking. The ongoing work on "professionalization" of evaluation as described by Quesnel and Morra in this book will certainly contribute to strengthen the supply side. This article does not focus on the individual level as it is extensively covered in this book through the contributions of Preskill, Kuzmin, Quinn, Patton, Fetterman and O'Sullivan.

Lessons learned on strengthening country monitoring and evaluation capacities

Extensive experience on capacity development initiatives has highlighted three major lessons learned.

The first lesson is that *capacity development should be underpinned by the fundamental characteristic of national ownership.* Taking a capacity development strategy to scale requires linking to national and local plans, processes, budgets and systems. To be sustained, a comprehensive capacity development response must link to, and draw from, relevant national reforms.

The second lesson is that capacity development is about transformations and must *address how best to manage change within the existing policy environment.* The tendency to often look only inside an organization and to downplay the larger institutional context, in which the organization resides, has proven unsuccessful. To ensure continued political commitment and resource support, a capacity development strategy can and often must show both short and long-term gains. Experience shows that capacity traps are more often pertinent to the "soft side" of the policy environment, such as vested interests, ethic and attitudes, rather than to "hard" technical competencies.

The third lesson is therefore, that a good evaluation capacity development strategy should be *tailored to the specific context of each country and institution.*

Conclusion

In the last decade, several countries reformed their social policies. However, several times these policy reforms did not deliver the expected results to the target population, especially children and women. It seems the main challenge resides in implementing the policy reforms, rather than in designing and adopting them.

Country monitoring and evaluation systems can play a central role in producing the relevant evidence to facilitate effective decisions to keep policy implementation on track, and to yield insights on the effectiveness of the strategy used to implement policy reforms.

However, in certain cases countries do not have the needed capacity to produce relevant and sound evidence. National monitoring and evaluation capacity development is therefore a central strategy to enhance country capacity to assess the progress of policy reform implementation more effectively, through use of the evidence generated by country-led monitoring and evaluation systems. While there is a strong consensus that national monitoring and evaluation systems should be led and owned by countries, international agencies such as the UN system and donors have a central role to play in strengthening country capacities.

REFERENCES

Adrien, M. and Jobin, D. (2009). *Country-led impact evaluation. A survey of development practitioners*. In: *Country-led monitoring and evaluation systems*.

Adrien, M. and Jobin, D. (2008). *Country-Led Evaluation*. In: *Bridging the Gap: the role of monitoring and evaluation in evidence-based policy-making.*

BA Tall, K .O. (2009). *The role of national, regional and international evaluation organizations in strengthening country-led monitoring and evaluation systems*. In: *Country-led monitoring and evaluation systems*.

Bamnerger, M. and Rugh, J. (2009). *RealWorld Evaluation: conducting evaluations under budget, time, data and political constraints,* In: *Country-led monitoring and evaluation systems*.

Bamberger, M. (2006). *Evaluation Capacity Building,* In: *Creating and Developing Evaluation Organizations) - Lessons learned from Africa, Americas, Asia, Australasia and Europe.*

Bamberger, M. (2008). *Enhancing the utilization of evaluation for evidence-based policy-making*, In: *Bridging the Gap: the role of monitoring and evaluation in evidence-based policy-making.*

Crosby, B.L. (1996). *Policy Implementation: The Organizational Challenge.*
World Development 24 (9): 1403–15.

Feinstein, Osvaldo. (2009). *Country-led Evaluation: Learning from Experience.*
In: *Country-led monitoring and evaluation systems.*

Giovannini, E. (2009). *Bringing statistics to citizens: a "must" to build democracy in
the XXI century.* In: *Country-led monitoring and evaluation systems.*

Government of Bosnia and Herzegovina and UNICEF. (2007). *Joint Country-led Evaluation
of child-focused Policies within the Social Protection Sector in Bosnia and Herzegovina.
Sarajevo.*

Gustafson, D.; and M. Ingle. (1992). *Policy Characteristics Analysis: Technical Notes.*
Implementing Policy Change Project No. 3. Washington: U.S. Agency for International
Development.

Heider, C. (Forthcoming) *Conceptual framework for developing evaluation capacities.
Building on good practices in evaluation and capacity development.*

IDEAS. (2006). *Country-led Evaluations and Systems: Practical experiences of the Central
and Eastern European Region. Regional Workshop, June 19-20, 2006. Prague.*

Kusek Zall, J. and Rist, R. (2008). *Ten step to a results based monitoring and evaluation
systems*, In: *Bridging the gap. The role of monitoring and evaluation in evidence-based
policy-making.*

Kusek Zall, J and Rist, R. (2009). *Getting the logic right. How a strong theory of change
supports programmes which work!.* In: *Country-led monitoring and evaluation systems.*

Kushner, S. (2006). *Democracy and Evaluation.* In: *New Trends in Development
Evaluation.*

Lundgren, H. (2009). *Supporting partner country ownership and capacity in development
evaluation. The OECD DAC evaluation network.* In: *Country-led monitoring and evaluation
systems*

Mackay, K. (2008). *Helping countries build government monitoring and evaluation
systems. World Bank contribution to evidence-based policy-making.* In: *Bridging the gap.
The role of monitoring and evaluation in evidence-based policy-making.*

Mackay, K. (2009). *Building monitoring and evaluation systems to improve government
performance.* In: *Country-led monitoring and evaluation systems.*

National School of Government UK. Policy Hub, *How research and evaluation evidence
contributes to policy-making.* Available at: < http://www.nationalschool.gov.uk/policyhub/
~evid.asp >

Nutley, S. M. and Davies H. T. O. (2000). *Making a reality of evidence-based practice:
some lessons from the diffusion of innovations.* In: Public Money and Management
Oct/Dec 20(4), 2000, pp.35-42.

Nutley, S. M., Davies, H. T. O. and Tilley, N. (2000). *Getting research into practice.*
In: Public Money and Management Oct/Dec 20(4), 2000, pp.3-6.

Nutley, S., Davies, H. and Walter I. (2002). *Evidence Based Policy and Practice: Cross Sector Lessons From the UK,* UK.

Ocampo, A. and Segone, M. (2004). *Using internet technology to develop evaluation capacity and practices in Latin America and Caribbean*. The Evaluation Exchange. Harvard Family Research Project. Volume X, No. 3, Fall 2004. p 23.

ODI (Overseas Development Institute) Civil Society Partnerships Programme. *Evidence-based Policy-making: Lessons from the UK for Developing countries.*

Paris Declaration Secretariat. (2008). *Evaluation of the implementation of the Paris Declaration.* France.

Parker, D. (2008). *Monitoring and evaluation, and the knowledge function,* In: *Bridging the gap. The role of monitoring and evaluation in evidence-based policy-making*.

Perri. (2002). *Can policy-making be evidence-based*? In: MCC: Building Knowledge for Integrated Care *Feb 10(1),* UK, pp.3-8.

Picciotto, R. (2009). *Evaluating development. Is the country the right unit of account?* In: *Country-led monitoring and evaluation systems*

Pron, N., Oswalt, K., Segone, M. and Sakvarelidzeg. (2009*). Strengthening country data dissemination systems. Good practices in using DevInfo* In: *Country-led monitoring and evaluation systems*.

Quesnel, J.S. (2006). *The Importance of Evaluation Associations and Networks*. In: *New Trends in Development Evaluation*.

Quesnel, J.S. (2009). *The strategic intent. Understanding strategic intent is the key to successful country-led monitoring and evaluation systems*. In: *Country-led monitoring and evaluation systems*

Radaelli, C. (1995). *The role of knowledge in the policy process*. In: Journal of European Public Policy Jun 2(2), 1995, pp.159-83.

Segone, M. (2006). *Democratic approach to evaluation*. In: *New Trends in Development Evaluation*.

Segone, M. (2006). *Evaluation culture: a new approach to learning and change*. In: *New Trends in Development Evaluation*.

Segone, M. (2006.) *The evolution of the evaluation function*. In*: New Trends in Development Evaluation*.

Segone, M. (2008). *Evidence-based policy-making and the role of monitoring and evaluation within the new aid environment*. In. *Bridging the gap. The role of monitoring and evaluation in evidence-based policy-making*.

Segone, M. (2009). *Enhancing evidence-based policy-making through country-led monitoring and evaluation systems*. In*: Country-led monitoring and evaluation systems*.

Segone, M., et al. (2009). *Country-led monitoring and evaluation systems*. UNICEF, World Bank, IDEAS, MICS and DevInfo, Switzerland

Segone, M., et al. (2008). UNICEF, World Bank, IDEAS, MICS and DevInfo. *Bridging the gap. The role of monitoring and evaluation in evidence-based policy-making*. Switzerland

Segone, M. and Ocampo, A. (2006). IOCE *Creating and Developing Evaluation Organizations. Lessons learned from Africa, Americas, Asia, Australasia and Europe, Peru.*

Segone, M , Patel, M., Rouge, J. and Russon, C. (2006). *National and Regional Evaluation Organizations as a Sustainable Strategy for Evaluation Capacity Development*, In: IOCE *Creating and Developing Evaluation Organizations) - Lessons learned from Africa, Americas, Asia, Australasia and Europe, Peru.*

Segone, M. (1998), UNICEF Regional Office for Latin America and the Caribbean: *Democratic Evaluation*, Colombia.

Segone, M. (2006). UNICEF Regional Office for CEE/CIS and IPEN. *New Trends in Development Evaluation*, Switzerland.

Uhr and Mackay (1996). *Evaluating policy advice*, Australian National University

UNDG. (2008). *Evaluation of the implementation of the Paris Declaration.* USA

UNDG. (2009). *A collective approach to supporting capacity development*, USA

UNDP. (2009). *Capacity Development. A UNDP Primer.* New York.

UNDP. (2009). *Supporting Capacity Development. The UNDP approach.* New York.

USAID. (2000). *Recent Practices in Monitoring and Evaluation TIPS: Monitoring the Policy Reform Process*, USAID Center for Development Information and Evaluation, Number 14, 2000.

Vestman, O. , and Connor, R. (2008). *The relationship between evaluation and politics.* In: *Bridging the gap. The role of monitoring and evaluation in evidence-based policy-making.*

Vukovic, A. and McWhinney, D. (2008). *Joint Country-Led Evaluation of the policies related to child well-being within the social protection sector in Bosnia and Herzegovina.* In: *Bridging the Gap: the role of monitoring and evaluation in evidence-based policy-making.*

Walshe, K. and Rundall, T. G. (2001). *Evidence-base management: from theory to practice in health care,* In: The Milbank Quarterly Sep 79(3), 2001, pp.429-57.

Weiss, C. H. (1999). *The Interface between Evaluation and Public Policy.* In: Evaluation 1999; 5; 468.

WORLD BANK IEG (Independent Evaluation Group). (2006). Evaluation Working Papers ISSUE #5: Evaluation Capacity Development, *Institutionalization of Monitoring and Evaluation Systems to Improve Public Sector Management.*

WORLD BANK/OED (Operations Evaluation Department). (2004). *Evaluation Capacity Development: OED Self-evaluation.*

WORLD BANK/OED (Operations Evaluation Department). UNDP and IOB. (2003). DAC working Party on Aid Evaluation, 37th Meeting, 27-28 March 2003, *Country led evaluations: a discussion note.*

FUTURE TRENDS IN EVALUATION[1]

by Michael Quinn Patton,
Founder and Director, Utilization-focused evaluation and former
President of the American Evaluation Association

The profession of evaluation has changed immensely over the last quarter century. I date the beginning of the profession to 1975 when the first *Handbook of the Evaluation Research* was published. That same year the Evaluation Network and Evaluation Research Society were formed in the United States (They merged in 1984 to become the American Evaluation Association). At that time there were no evaluation journals, no regular training institutes, few textbooks, and no standards for evaluation. All that has changed dramatically in what has become the rich and diverse mosaic that is now the global evaluation profession. In this chapter I will outline the major trends I expect will shape evaluation over the next quarter century.

These are:

1. increased international and cross-cultural expansion of evaluation with greater globalization and diversity;

2. evaluation increasingly acknowledged and valued as a *trans-discipline* and profession;

3. increased political interest in accountability, performance indicators, and transparency;

4. growing emphasis on evaluation capacity-building and skill development;

5. continuing debate about what constitutes methodological rigour; and

6. deepening use of systems thinking and complexity science as frameworks for evaluation.

---

1   Future trends in evaluation was originally published in ″Оценка программ: методология и практика″. Reprinted with the permission of Process Consulting and the author.

International and cross-cultural expansion of evaluation: globalization and diversity

No trend has been more important to evaluation in the last decade than its expanding global reach. In 1995 evaluation professionals from 61 countries around the world came together at the first truly international evaluation conference in Vancouver, British Columbia. Ten years later a second international conference in Toronto attracted 2,330 evaluation professionals from around the world. The 1990's also gave rise to the European Evaluation Society, founded in 1994 in the Hague, and the African Evaluation Association, founded in 1999 in Nairobi having held its 5th continent-wide conference in Egypt in 2009. Now there are over 60 national evaluation associations around the world, including: Brazil; Colombia; Japan; Malaysia; Mongolia; New Zealand; Niger; Peru; Russia; South Africa; Sri Lanka; and, Zimbabwe to name but a few examples. In 2003 in Lima, Peru, the inaugural meeting of the new International Organization for Cooperation in Evaluation (IOCE) was held as an umbrella networking and support initiative for national and regional evaluation associations around the world. The International Development Evaluation Association (IDEAS) was formed in Beijing, in 2002, to support evaluators with special interests in developing countries. Its first biennial conference was held in New Delhi in 2005. The Network for Monitoring, Evaluation, and Systematization of Latin America and the Caribbean (ReLAC) was formed in 2005, in Peru.

The World Bank, through its International Programme for Development Evaluation Training (IPDET), offers month-long evaluation training for people throughout the developing world at Carleton University in Ottawa, Canada. International agencies have developed comprehensive guidelines for the conduct of evaluation. Various national associations have reviewed and adapted the Joint Committee Standards (1994) for evaluation to fit their own socio-political contexts, but still affirming that evaluations should be judged by their utility, feasibility, propriety, and accuracy.

Governments around the world are building new systems for monitoring and evaluation, aiming to adapt results-based management and performance measurement to support development. International agencies have also begun using evaluation to assess the full range of development efforts under way in developing countries. Most major international organizations have their own evaluation units with guidelines; protocols; conferences; training opportunities; websites; and resource specialists.

Such globally interconnected efforts made it possible for evaluation strategies and approaches to be shared worldwide. Thus, the globalization of evaluation supports our working together, to increase our international understanding about factors which support programme effectiveness and evaluation use. International perspectives also challenge Western definitions and cultural assumptions about how evaluations ought to be conducted and how quality ought to be judged. As evaluation standards are translated into different languages, national associations are adding their own cultural nuances and adapting practices to fit local political, social, organizational, economic, and cultural contexts. I expect this trend of cultural and political adaptation to continue to shape evaluation in the future.

Evaluation as a trans-discipline and profession

Philosopher and evaluation theorist Michael Scriven has characterized evaluation as a *trans-discipline*, because every discipline, profession, and field engages in some form of evaluation, the most prominent example being, perhaps, evaluations of students taking courses and completing disciplinary programmes of study, and refereed journals in which new research is evaluated by peers to determine if it is worthy of publication. Evaluation is a disciple that serves other disciplines even as it is a discipline in its own right, thus its emergent trans-disciplinary status. Statistics, logic, and evaluation are examples of trans-disciplines in that their methods, ways of thinking, and knowledge base are used in other areas of inquiry, e.g. education; health; social work; engineering; environmental studies; and so on. I expect evaluation to be increasingly recognized as a trans-disciplinary field, which will have implications for both how we conduct evaluations (using interdisciplinary teams), and how we conduct research on evaluation.

One important implication of this trend is recognizing that evaluation has its own knowledge base about the factors that affect programme effectiveness and how to conduct useful evaluations. Far too many evaluations are conducted by economists and other social scientists who have no evaluation knowledge or expertise. Many do not know that there are standards for high quality evaluation. They do not know the options that have emerged in evaluation methods and models. So one trend I hope to see is that as evaluation is recognized both as a profession and as a trans-disciplinary field of knowledge. Those who commission and fund evaluations will be sure to include trained evaluation professionals on the teams that conduct evaluations worldwide.

Increased political interest in accountability, performance indicators, and transparency

Performance measurement has infused politics globally. We see this in the increasing attention to performance targets, benchmarks, and milestones in the Millennium Development Goals and treaties, like the Kyoto Agreement on greenhouse gas emissions. Performance indicators have become so important and widely used that they have become a regular part of legislation, government management, and international monitoring agreements. Ongoing monitoring of indicators against desired target levels may be called performance measurement or performance monitoring. They are used for three main purposes:

(1) to assess the impact of government policies on services;

(2) to identify well-performing or under-performing institutions and public servants; and

(3) for public accountability.

Governments and international agencies are in the position of both monitoring public services and being monitored by performance indicators. This makes the political stakes quite high.

Performance monitoring done well is useful within a broader monitoring and evaluation framework. Done badly, it can be very costly and not merely ineffective but harmful and indeed destructive. Thus, in the future, work will have to be done to build more sophistication about useful performance monitoring systems, including their inevitable limitations. The potential positive contribution of performance monitoring is captured in the mantra "what gets measured gets done". Well-developed and appropriate indicators both focus attention on priority outcomes and provide accountability for achieving those outcomes. The shadow side of performance indicators is that measuring the wrong thing means the wrong thing gets done.

Education and training in the appropriate use and interpretation of performance indicators is needed. Special emphasis will need to be given to the importance of independent scrutiny and transparency as safeguards of public accountability, methodological rigour, and fairness to individuals and/or institutions being monitored.

Concerns about the misuse of performance indicators follow from Campbell's Law, formulated by Donald Campbell, one of evaluation's most distinguished pioneers: "The more any quantitative

47

social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."

Consider this example: Police officers in New Orleans manipulated crimes statistics to make it look like the crime rate was going down stimulated by the Department's policy of handing out awards to leaders of districts with the lowest crime statistics. Five police officers were fired over the scandal.

Governments and politicians are expected to set targets and report on progress as a basis for public accountability. The usefulness of performance indicators depends on their credibility, relevance, validity, transparency, and meaningfulness – and an appropriate and fair process for interpreting them. Performance indicators are one tool in a very large evaluation toolkit that includes a wide variety of methods, data collection techniques, measures, and models. Given the rapid and widespread proliferation of performance monitoring approaches, there is the danger that many will think that performance measurement is sufficient for, or equivalent to, evaluation. But performance measurement merely portrays trends and directions. Indicators tell us whether something is increasing, declining, or staying the same. Evaluation takes us deeper into asking why indicators are moving in the direction they are, how the movement of indicators are related to specific interventions, what is driving the movement of indicators, and what values should guide interpretation of indicators in making judgments. Utilization-focused performance measurement adds the importance of being clear about the primary intended users and intended uses of performance indicators. Given the increasing importance of performance measurement in the public sector around the world, evaluators, policy makers, and the general public need to understand both the strengths and limitations of key performance indicators.

Evaluation capacity-building and skill development

We have learned that organizations, programmes, and projects need resources, knowledge and organizational development in order to effectively manage and conduct evaluations. In the last decade especially, we have developed a deeper understanding of the importance of enhancing evaluation capacity in organizations by infusing evaluative thinking into organizational culture. Organizations increasingly have the capacity to generate, store, and retrieve massive amounts

of information and data. The problem is knowing how to use all of that information. The technological capacity for gathering and computerizing information now far exceeds the ability of most organizations to process and to make sense out of it all. Organizational leaders are constantly faced with deciding what's worth knowing, what to ignore, and how to translate findings into action and decisions. This means that increasingly, in the future, evaluators will be called on not only to conduct evaluations but also to build sustainable evaluation capacity in organizations. Evaluators will have to work to build and sustain interest in evaluation use. Identifying intended users is part selection and part nurture. Potential users with low opinion of, or little interest in, evaluation may have had bad prior experiences or have just not given much thought to the benefits of evaluation. Part of the evaluation task, then, will be to cultivate interest in evaluation and build commitment to its use. Even people initially inclined to value evaluation will still often need training and support to become effective information users.

Evaluators also need capacity development and skills beyond just methodological knowledge. We have learned that to facilitate effective use of evaluations, evaluators need skills in building relationships; facilitating groups; managing conflict; walking political tightropes; and, in effective interpersonal communications. Technical skills and social science knowledge are not sufficient to get evaluations used. People skills are critical. Ideals of rational decision making in modern organizations, notwithstanding personal and political dynamics, affect what really happens. Evaluators without the savvy and skills to deal with people and politics will find their work largely ignored or, worse still, used inappropriately.

Process use will be increasingly understood and valued

The chapter on utilization-focused evaluation in this book discusses *process use* and its importance in evaluation. I expect that importance to increase. *Process use* focuses on the learning that occurs and the capacity that is built through participation in an evaluation. Participatory and collaborative approaches to evaluation have an effect beyond the use of findings. Those who participate in evaluation learn to think more critically. They learn how to frame questions; interpret data; set priorities; clarify intervention models; and focus, on outcomes. They learn to use the logic and employ the reasoning of evaluation. Thus, process use builds capacity for sustainable evaluation engagement and use.

Thinking in terms of what's clear, specific, concrete, and observable does not come easily to people who thrive on, even depend on, vagueness, generalities and untested beliefs as the basis for action. They are in the majority. Practitioners of evaluation logic are a small minority. The good news is that thinking like an evaluator, once experienced, is often greatly valued. That is part of what creates demand for evaluation services.

Process use is distinct from use of the substantive findings in an evaluation report. It is equivalent to the difference between learning how to learn, versus learning substantive knowledge about something. Learning how to think evaluatively is learning how to learn and think critically, and those who become involved in an evaluation learn by doing. Facilitating evaluative thinking opens up new possibilities for impact which organizations and funders are coming to value, because the capacity to engage in this kind of thinking can have more enduring value than a delimited set of findings. This especially resonates for organizations interested in becoming what has popularly come to be called "learning organizations." Learning how to see the world as an evaluator sees it, often has a lasting impact on those who participate in an evaluation – an impact that can be greater and last longer than the findings from that same evaluation. Findings have a very short 'half life' – to use a physical science metaphor: they deteriorate very quickly as the world changes rapidly. Specific findings typically have a small window of relevance. In contrast, learning to think and act evaluatively can have an ongoing impact. The experience of being involved in an evaluation, then, for those stakeholders actually involved, can have a lasting impact on how they think, on their openness to reality-testing, and on how they view the things they do. Thus, I expect the future to bring increased attention to process use as part of capacity-building in evaluation.

Methodological debates about rigour

Evaluation's past has involved intense debates about what constitutes methodological rigour, including especially debates about the value and credibility of quantitative versus qualitative methods. The debates have taken different forms over time, including periods of intense rancor as well as times of rapprochement. The debate is once again in the ascendant, this time focused on whether randomized controlled experiments are the gold standard for impact evaluations.

The validity of experimental methods and quantitative measurement, appropriately used, has never been in doubt. By the 1990s, however, qualitative methods, appropriately used, had ascended to a level of comfortable respectability. Today, mixed methods have become especially valued. While a consensus has emerged in the profession that evaluators need to know and use a variety of methods, in order to be responsive to the nuances of particular evaluation questions and the idiosyncrasies of specific stakeholder needs, the question of what constitutes *the methodological gold standard* remains hotly contested. On the one hand, there is agreement that rigour should be judged by the appropriateness of methods for a specific evaluation purpose and question and that, where possible, using multiple methods – both quantitative and qualitative – can be valuable. However, there is also a widespread belief that one question is more important than others, the causal attribution question, and that one method, randomized control trials, is superior to all other methods in answering that question. This is where the gold standard issue comes into play.

The debate is not just among evaluation methodologists. Evaluation practitioners are deeply affected, as are users of evaluation: policymakers; programme staff; managers; and funders. All can become mired in the debate about whether statistical results from experiments ("hard" data), are more scientific and valid than quasi-experiments and qualitative case studies ("softer" data). Who wants to conduct (or fund) a second-rate evaluation, if there is an agreed-on gold standard? What really are the strengths and weaknesses of various methods, including experiments (which, it turns out, also have weaknesses)? What does it mean to match the method to the question?

If evaluators are to involve intended users in decisions about methods, evaluators and intended users need to understand the methods debate and evaluators need to be able to facilitate choices which are appropriate to a particular evaluation's purpose. This means educating primary stakeholders about the legitimate options available, the potential advantages of multiple methods, and the strengths and weaknesses of various approaches.

Both the American Evaluation Association and the European Evaluation Society have adopted policy statements which advocate methodological eclecticism, and the adaptation of evaluation methods to the nature of the evaluation question and information needs of primary intended users. These position statements assert that

methodological appropriateness should be the gold standard. The problem is that the alleged superiority of the quantitative/experimental approach has cut off serious consideration of alternative methods, and channels millions of dollars of evaluation funds into support for experimental designs. Those designs have strengths, but also they have significant weaknesses. The gold standard accolade means that funders and evaluators begin by asking: "How can we do an experimental design?" rather than asking: "Given the evaluation situation and the information needed, what is the appropriate evaluation design?" The prestige of the method determines the evaluation question and design rather than considerations of utility; feasibility; propriety; and, accuracy. Under the gold standard label, high quality impact evaluation is defined as testing hypotheses, formulated deductively, through random assignment of programme participants to treatment and controls, and measuring outcomes quantitatively. No other options are worthy of serious consideration – *by definition*.

Yet, alternatives exist. There are ways, other than experiments, of assessing programme processes, outcomes, and impacts. In the last quarter century, these alternatives have been used by evaluators and practitioners who found that the dominant paradigm failed to answer, or even ask, their questions. Debate about whether experimental designs constitute the methodological gold standard revolves, in part, around what level and kind of evidence is needed to determine that an intervention is effective. Let me illustrate with an example from my book, *Utilization-Focused Evaluation* (Patton, 2008, Chapter 12).

Consider the challenge of eradicating intestinal worms in children, a widespread problem in developing countries. Suppose we want to evaluate an intervention in which school-age children with diarrhea are given worm medicine to increase their school attendance and performance. In order to attribute the intervention to the desired outcome, advocates of randomized controlled trials would insist on an evaluation design in which students suffering from diarrhea are randomly divided into a treatment group (those who receive worm medicine), and a control group (those who do not receive the medicine). The school attendance and test performance of the two groups would then be compared. If, after a month on the medicine, those receiving the intervention show higher attendance and school performance at a statistically significant level compared to the control group (the counter-factual), then the increased outcomes can be attributed to the intervention (the worm medicine).

Advocates of qualitative inquiry question the value of the control group in this case. Suppose that students, parents, teachers, and local health professionals are interviewed about the reasons students miss school and perform poorly on tests. Independently, each of these groups asserts that diarrhea is a major cause of the poor school attendance and performance. Gathering data separately from different informant groups (students, parents, teachers, heath professionals) is called *triangulation*, a way of checking the consistency of findings from different data sources. Following the baseline interviews, students are given a regimen of worm medicine. Those taking the medicine show increased school attendance and performance, and in follow-up interviews, the students, parents, teachers, and health professionals independently affirm their belief that the changes can be attributed to taking the worm medicine and being relieved of the symptoms of diarrhea. Is this credible, convincing evidence?

Those who find such a design sufficient argue that the results are both reasonable and empirical, and that the high cost of adding a control group is not needed to establish causality. Nor, they would assert, is it ethical to withhold medicine from students with diarrhea when relieving their symptoms has merit in and of itself. The advocates of randomized controlled trials respond that without the control group, other unknown factors may have intervened to affect the outcomes and that *only the existence of a counter-factual* (control group), will establish with certainty the impact of the intervention.

As this example illustrates, those evaluators and methodologists on opposite sides of this debate have different worldviews about what constitutes sufficient evidence for attribution and action in the real world. This is not simply an academic debate. Millions of dollars of evaluation funds are at stake and the results of these evaluations around the world will affect billions of dollars of international development assistance.

In 2008 major funders interested in evaluation established the International Initiative for Impact Evaluation (3ie). The mission of 3ie is to "contribute to the fulfillment of aspirations for well-being by encouraging the production and use of evidence from rigorous impact evaluations for policy decisions that improve social and economic development programmes in low and middle-income countries." The future of evaluation will involve intense international debate about what constitutes "rigour."

Systems thinking and complexity science as frameworks for evaluation

The final trend I see is increased attention to and use of systems thinking in evaluation. Evaluation has become heavily dependent on linear logic models to conceptualize interventions. In the last decade, conducting an evaluation has often come to include conceptualizing and testing a programme's logic model or theory of change. As evaluators became involved in working with programme people to more clearly specify the programme's model (or theory), it became increasingly clear that evaluation was an *up-front activity* not just a back-end activity. That is, traditional planning models laid out some series of steps in which planning comes first, then implementation of the programme, and then evaluation, thus making evaluation a back-end, last-thing-done activity. But, to get a programme plan or design that could actually be evaluated has meant involving evaluators, and evaluative thinking, from the beginning. Evaluative thinking, then, becomes part of the programme design process including, especially, conceptualizing the programme's logic model or theory of change and asking the question: How will what the programme does lead to the desired results? Engaging in this work is an example of *process use* in which the evaluation has an impact on the programme quite apart from producing findings about programme effectiveness. The very process of conceptualizing the programme's theory of change can have an impact on how the programme is implemented, understood, talked about, and improved. The evaluative thinking process has these impacts, as noted earlier.

This has huge implications for evaluators. It means that evaluators have to be:

(1) astute at conceptualizing programme and policy theories of change, and

(2) skilled at working with programme people, policymakers and funders to facilitate their articulation of their implicit theories of change.

Given the importance of these tasks, it matters a great deal what theory of change frameworks the evaluator can offer. Systems thinking is one such framework, one which I expect to become increasingly valued and used in evaluation.

Linear logic models construct models and diagrams in which inputs lead to activities, activities lead to outputs, and outputs lead to outcomes. In contrast, looking at a programme from a systems perspective focuses on the interdependent configuration of factors that lead to outcomes rather than a simple cause-effect model. A systems framework is built on some fundamental relationships premises:

a. The whole is greater than the sum of the parts.

b. Parts are inter-dependent such that a change in one part has implications for all parts and their inter-relationships.

c. The focus of the model is on interconnected relationships.

d. Systems are made up of sub-systems and function within larger systems.

In 2006 the American Evaluation Association published its first-ever monograph, an expert anthology on *Systems concepts in evaluation* (Williams and Iraj Iman, 2006:6). That monograph provides a wide range of systems approaches and demonstrates the diversity of approaches which congregate under the systems umbrella. In commenting on this diversity the editors wrote:

"for those of you looking for coherence about what we consider to be relevant systems concepts for evaluation, our advice when reading this publication is to look for patterns rather than definitions. For us, three patterns stand out:

1. *Perspectives*. Using systems concepts assumes that people will benefit from looking at their world differently. For systems practitioners, this motivation is explicit, deliberate, and is fundamental to their approach. However, just looking at the "bigger picture," or exploring interconnections does not make an inquiry "systemic." What makes it systemic is how you look at the picture, big or small, and explore interconnections. A "system" is as much an "idea" about the real world as a physical description of it.

2. *Boundaries*. Boundaries drive how we "see" systems. Boundaries define who or what lies inside and what lies outside of a particular inquiry. Boundaries delineate or identify important differences (i.e. what is "in" and what is "out.") Boundaries determine who or what will benefit from a particular inquiry and who or what might suffer. Boundaries are fundamentally about values – they are judgements about worth. Defining boundaries is an essential part of systems work/inquiry/thinking.

3. *Entangled systems*. One can observe and perceive systems within systems, systems overlapping other systems, and systems tangled up in other systems. Thus it is unwise to focus on one view or definition of a system without examining its relationship with another system. Where does one system begin and the other end? Is there overlap? Who is best situated to experience or be affected by that overlap? What systems exist within systems and where do they lead? A systems thinker always looks inside, outside, beside, and between the readily identified systems boundary. He or she then critiques and, if necessary, changes that initial choice of boundary".

Evaluation as a profession and trans-discipline has just begun to consider and work with the implications of systems theory. I expect the future of evaluation to include increasing use of systems concepts and thinking in our work, and much greater attention to developmental evaluations specifically designed and implemented for complex adaptive systems (Patton, 2010).

Summary

The future of evaluation seems assured. The field has developed rapidly into a vital and vibrant profession. In this brief overview I have identified six trends that I think will continue to support the development of evaluation worldwide. I foresee:

1. increased international and cross-cultural expansion of evaluation with greater globalization and diversity;

2. evaluation increasingly acknowledged and valued as a *trans-discipline* and profession;

3. increased political interest in accountability, performance indicators, and transparency;

4. growing emphasis on evaluation capacity-building and skill development;

5. continuing debate about what constitutes methodological rigour; and

6. deepening use of systems thinking and complexity science as frameworks for evaluation.

References

Joint Committee on Standards for Educational Evaluation. (1994). *The Programme Evaluation Standards.* Thousand Oaks, CA: Sage.

Patton, Michael Quinn. (2008). *Utilization-Focused Evaluation,* 4th edition. Thousand Oaks, Ca: Sage Publications.

Patton, Michael Quinn. (2010). Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use. New York: Guilford Press.

Struening, Elmer L. and Marcia Guttentag (eds.). (1975). *Handbook of Evaluation Research.* Beverly Hills, CA: Sage.

Williams, Bob and Iraj Iman. (2006). *Systems Concepts in Evaluation: An Expert Anthology.* AEA Monograph. Point Reynes, CA: Edge

MEASURING THE PROGRESS
OF SOCIETIES

by Enrico Giovannini,
President of the Italian Statistical Institute, former OECD Chief
Statistician and Co-Chair of the Board of the Global Project
on "Measuring the progress of societies"

by Jon Hall,
Programme Officer for PARIS21, former Project Leader for the OECD
Secretariat of the Global Project on "Measuring the progress
of societies"

Introduction

The Global Project on "Measuring the progress of societies" is an international, inter-organisational initiative, hosted by the OECD. Its aim is to encourage the development of statistical measures which go 'beyond GDP' to provide a comprehensive picture of societal wellbeing. This paper outlines the 'measuring progress' approach and argues that better measures of progress will lead to more effective policy. Some of the potential implementation benefits of a country-led 'measuring progress' initiative include: fostering 'whole-of-government' cooperation; helping with reform challenges; developing an outcome-focused governance culture; and strengthening accountability mechanisms. However, the success of such projects relies to a large extent on conditions which remain a significant challenge in developing countries.

Around the world, there is growing acknowledgement of the need for new and broader measures of societal progress. That is to say, of the need to develop and use statistical indicators which go beyond a narrow focus on economic outputs to give a wider view of societal wellbeing and its change over time. In August 2009, the European Commission released a communication declaring: "There is a clear case for complementing GDP with statistics covering the other economic, social and environmental issues, on which people's well-being critically depends" (European Commission 2009). In 2008, when announcing the creation of the Commission on

the Measurement of Economic Performance and Social Progress,[1] French President Nicolas Sarkozy stated, "I have a deep conviction: we will not change our behaviour until we change our performance measures. And our behaviour absolutely must change." One of the lead experts on that Commission, Joseph Stiglitz, argued that: "Producing better, truer, ways of measuring economic, environmental and social performance is a critical step in making progress towards building a better world (Stiglitz 2009). In addition to these 'top-down' initiatives to develop new progress measures, a rise in the number of non-governmental initiatives around the world, developing new indicators for social change, suggests a powerful 'bottom-up' movement.

The Istanbul Declaration and the Global Project on "Measuring the progress of societies"

In 2007, the OECD held its 2nd *World Forum on Statistics, Knowledge and Policy* in Istanbul, Turkey. 1200 participants, from 130 countries, representing government, business, academia, civil society and the media, gathered to discuss diverse topics related to the conference theme of "Measuring and fostering the progress of societies". The culmination of the event was the release of the *Istanbul Declaration*, where six international organisations urged "statistical offices, public and private organisations, and academic experts to work alongside representatives of their communities to produce high-quality, facts-based information which can be used by all of society to form a shared view of societal well-being and its evolution over time."[2]

---

1   The Commission, headed by Nobel Laureates Joseph Stiglitz and Amartya Sen, and by the economist Jean-Paul Fitoussi was established with the mandate to: identify the limits of GDP as an indicator of economic performance and social progress; including the problems with its measurement; to consider what additional information might be required for the production of more relevant indicators of social progress; to assess the feasibility of alternative measurement tools; and to discuss how to present the statistical information in an appropriate way (Stiglitz, Sen, Fitoussi 2009). The Commission released its recommendations in September 2009. For more information see www.stiglitz-sen-fitoussi.fr

2   For the full text and list of signatories of the Istanbul Declaration, and for information about the Global Project on Measuring Progress in general, see www.oecd.org/progress.

In addition, the Declaration reflected an international consensus on the need to:

- undertake the measurement of societal progress, going beyond conventional economic measures such as GDP per capita;

- enhance a culture of evidence-based decision-making to increase the welfare of society;

- strengthen citizens' capacity to influence the goals of their societies; and,

- increase the accountability of public policies.

The Global Project on "Measuring the progress of societies" (henceforth, the 'Global Project') has emerged as a response to this call for action. The Global Project is a multi-organisational initiative, led by the OECD, the World Bank, UNDP, the European Commission and others, which brings together hundreds of actors from a multitude of disciplines, from all sectors of society and all regions of the world. Research institutes, non-governmental organisations and statistical offices, from both developing and developed countries are all associated with the Global Project. The Global Project seeks to become the world-wide reference point for those who wish to measure, define, and assess the progress of their societies by encouraging discussions on what progress means and by developing the measures needed to track progress.

Along these lines, the activities of the Global Project contribute to the following goals:

- Encouraging communities to consider for themselves what 'progress' means in the 21st Century.

- Identifying and sharing best practices on the measurement of societal progress and increasing the awareness to do so using sound and reliable methodologies.

- Stimulating national and international debates, based on solid statistical data and indicators, on both national and global issues of societal progress, their measurement and policy incidences.

- Working towards producing a broader, shared, public understanding of changing conditions, while highlighting areas of significant change or inadequate knowledge.

- Advocating appropriate investment in building statistical capacity, especially in developing countries, to improve the availability of data and indicators needed to guide development programmes and to report on societal progress, as well as on progress towards internationally agreed goals.

The 3rd *World Forum on Statistics, Knowledge and Policy* was held in October 2009 in Busan, Korea, with almost 2000 participants attending. The event reaffirmed the importance of the global project's role as an international focal point for efforts to measure progress.

"Measuring progress": what does it actually mean?

A natural starting point when measuring progress is to ask "What is progress?". This is no simple question to answer: indeed, it is a question that has exercised philosophers since the time of the ancient Greeks. The word "progress" refers to improvements, to move forward, to gain. Progress implies the notion of time and space, the certainty of knowing where we are and where we want to go, and having reference points and clear goals. The word "progress" can be used in many aspects of our life and culture. People speak about economic progress, social progress and scientific progress. But above all we talk about human progress. In the latter sense we define progress as an improvement in the overall well-being of humanity, or put in simple terms, "Is life getting better?"

Measuring the progress of a society is an ambitious and complex task. Just defining societal progress is problematic as "progress", and indeed "society", can mean different things to different people. The Global Project recognises this and embraces a wide diversity of different approaches within the measuring progress 'movement'. Indeed, it is important that each society defines for itself what progress means. Despite this heterogeneity, there are nevertheless some common aspects which emerge from this new wave of efforts to measure progress. These aspects concern both statistical content and process.

Content

The most obvious feature of current attempts to measure progress is that they generally widen the scope beyond economic indicators. For the greater part of the 20th century, progress was seen to be synonymous with economic growth. In recent decades, there have

been increasing challenges to this view. The work of the Brundtland Commission, and the subsequent 1992 Earth Summit in Rio, put the issue of 'sustainable development' firmly on the international agenda, placing economic growth in the context of its environmental and social impacts. The subject of quality of life has also risen in importance in policy circles, as there has been a recognition that increasing GDP does not always correlate with the number of people experiencing higher wellbeing or satisfaction. The financial and economic crisis of 2008/9 has reinforced the idea for many that a shift in policy (and therefore measurement) is needed.

Widening the scope for measuring progress beyond a narrow focus on economic growth makes sense, but it also introduces increased complexity into the task. Progress measurement which goes 'beyond GDP' to encompass wider economic, social and environmental issues is therefore multidimensional. The OECD has developed a *Framework of Societal Progress*[3], which aims to set out the key dimensions (in a non-prescriptive way) which could be considered when measuring progress. These are grouped under six categories: ecosystems; human wellbeing; economy; governance; and culture. 'Progress' can be seen through a number of different lenses which themselves cut across dimensions, such as quality of life, sustainability, poverty reduction, and gender equity to name but a few.

By rethinking progress, we are rethinking the goals of our societies. The shift towards a wider, multidimensional view of progress has also been accompanied by a shift towards a more qualitative (rather than quantitative) view of progress. That is to say, as societies question what is important to them, the important issue is increasingly not "Do we have more today than yesterday?" but "Is life better today than yesterday?". In order to answer that question, better measures are needed which can convey information about the immaterial aspects of our lives. A 2009 paper by the UK Sustainable Development Commission illustrates this point, using the concept of 'prosperity':

"....prosperity goes beyond material pleasures. It transcends material concerns. It resides in the quality of our lives and in the health and happiness of our families. It is present in the strength of our relationships and our trust in the community. It is evidenced by our satisfaction at work and our sense of shared

---

3   See www.oecd.org/progress

meaning and purpose. It hangs on our potential to participate fully in the life of society in our ability to flourish as human beings – within the ecological limits of a finite planet." (Jackson 2009, p. 5).

In order to measure such intangible aspects of progress /wellbeing/development as trust, social cohesion or empowerment, many projects are making innovative use of perception or subjective data, and methodological advances are rapidly being made in this area. The Stiglitz-Sen-Fitoussi Commission recognised in its final report that, adequately assessing quality of life will require both subjective and objective data (Stiglitz, Sen, Fitoussi 2009). Another of the key recommendations of the Stiglitz-Sen-Fitoussi report was the importance of looking beyond aggregate data to look at patterns of distribution alongside national averages. For example, income is an important indicator of wellbeing but national averages can hide large disparities amongst regions or between social groups.

Process

For many practitioners involved in progress measurement projects, the process of arriving at the final indicator or indicators is as important as the indicator itself. This is partly because the question of what to measure when considering societal progress is as political as it is technical. A list of societal progress indicators (or the choice of components of a composite indicator) is an implicit statement of the core values of a society, of what matters most to the people in it. *Who* is doing the choosing and *how* that discussion is taking place become as important as *what* is being chosen.

Increasingly, there is a consensus amongst practitioners that the process of selecting and developing indicators of societal progress needs to be based on discussion and interaction between representatives from different disciplines and different sectors of society. Including experts, practitioners and members of the public in a participatory process can increase the relevance and legitimacy of the indicators selected. The larger the geographic scale covered by the indicators, the more difficult it becomes to manage an inclusive process. However, it is possible. For example, for their publication *Measures of Australia's Progress*, the Australian Bureau of Statistics successfully consulted with a wide selection of interdisciplinary experts and members of the public in order to select the indicators.

Once the indicators are finalised and a report is published, the "measuring progress" process often does not end there. For many, the issue of societal progress is of urgent importance and the indicators need to be 'shouted from the rooftops'. In this respect, finding innovative and effective ways of communicating and building knowledge with the data can be an integral part of the process.

While there are as many different ways of organising the process of developing indicators as there are indicators, the Global Project has identified six key steps:

Step 1: Defining the issue of what matters most to a society. Any indicators project needs to be situated within a framework of what progress means in a specific place and time. A good framework will define the scope of the work. It will identify the specific aspects (or dimensions) of progress which one is seeking to measure and it will describe the ways in which they relate to one another.

Step 2: Identifying partners to carry out the effort and establishing a core group of stakeholders. No matter what institution you belong to, any organisation seeking to measure progress should engage a wide range of stakeholders. Of course, deciding who should be part of the "coalition" depends on the historical, political, institutional, and cultural elements which characterise the society whose progress is being measured. An institution that seeks to measure progress might be a governmental body; a research institute; an institution devoted to address the issue of accountability of public policies or to support their development; a civil society organisation; or, a statistical institute. In each case, institutions which are perceived as authoritative and open to collaboration with other institutions, are more likely to be successful. The higher the past record of launching and implementing successful projects, the easier it will be to engage others. Finally, it is possible that the original group of stakeholders may change over time. This is not necessarily a problem, unless the turnover of the participants is a sign of weakness of the leading institution.

Step 3: Producing an initial set of indicators. The process for identifying what to measure and how to measure societal progress, i.e. to define the key dimensions of what progress means (health, material well-being, etc.) and to select the most relevant indicators takes time. Patience and good management in organising the political and technical dialogue are key ingredients of this phase. The legitimacy of the final result will depend on the extent to which key stakeholders have been involved in the process. Of course,

the range of institutions that could be potentially involved is often large, and the more voices in the process, the more difficult it can be to find an agreement. This implies that an appropriate balance needs to be struck given the resources of the lead institution.

Step 4: Getting the information 'out there'. The fourth step is getting the information out to intended audiences. This is a step that most producers of indicators are familiar with and accept as a key objective. Getting the information 'out there' can be done in a variety of ways: releasing print publications and publishing the data on the web are the most common strategies. However, simply putting the information into the public domain is not the same as ensuring the information reaches the target audiences. It may be that the intended audience is 'everybody', but while it is admirable to be so ambitious, some thought should be given to the different needs of different groups within the target audience. For example, it is unlikely that policy-makers will have the same requirements and respond to the same kinds of information as school children. This does not mean that the same set of indicators cannot be used to reach different groups, but in order to design effective dissemination strategies, there needs to be a clear idea of the needs of the audience from the beginning.

Step 5: Communicating and building knowledge with the indicators. While it may seem obvious that dissemination is a key step for indicator producers, the problem comes when releasing the information is seen as the end of the process. As Albert Einstein said, "Information is not knowledge". The difference between data, information and knowledge is an issue of understanding and application. Information is data which has been given meaning (rather than rows of raw numbers in a spreadsheet). Knowledge comes about when information has been absorbed by users, allowing them to understand and relate the information to their context. Building knowledge is therefore the next important step in the journey from data to action, but it is not a straightforward task.

Step 6: Ensuring continuity and relevance. Indicators of societal progress, when published regularly over time, allow all members of society to judge where society is performing well and where action is needed to change course. Producing an influential indicators report cannot however, be a one-off exercise. To have any kind of meaningful impact, the exercise must be repeated regularly over time. It will probably also be necessary to make adjustments over time, to the indicators and to the communications approach, in order to ensure the continued relevance of the exercise.

Measuring progress to foster progress: from data to results

How can the statistical developments covered in the previous section help in overcoming the implementation challenge? Potentially, there are several ways.

First, better measures will assist with the design, implementation and monitoring of more effective policies and programmes. For example, a household may be classed as having adequate monetary income, and yet still not have access to education or healthcare. Multi-dimensional measures will take these deprivations into account, and so support better targeting of interventions.

Secondly, a country-led indicators approach, including agencies from across government, can foster 'whole-of-government' cooperation which will have knock-on benefits for the implementation stage. 'Progress' indicators make a strong statement about what goals are important for a country. If there is no support for the choice of indicators, this will undermine cooperation at the implementation stage. Including different government departments in the process of selecting indicators will help to ensure shared understandings and objectives at the policy implementation stage.

A third point, related to the one above, is that the 'progress indicators' approach can help to encourage a more 'outcome-focussed' governance culture. Indicators of progress tend to be about societal outcomes in the economic, social and environmental domains, rather than purely about government inputs and outputs. They can help to keep the long-term objectives in view (i.e. universal literacy) alongside data on funding and programme implementation.

The fourth point is that transparent information about country performance in important societal areas can help with reform challenges. If governments are to reform they need the support of citizens. Reliable, transparent and resonant measures of progress can assist that process. For example, in Mexico, indicators on educational performance are being used to illustrate the need for investment in the school system. Mexican students ranked very low in the 2006 PISA[4] assessment, and yet a 2007 survey of Mexican

---

4   Programme of International Student Assessment (see www.oecd.pisa.org). PISA surveys are carried out every three years in over 60 countries to assess the performance of 15-year-olds related to knowledge and skills essential for full participation in society (e.g. literacy, numeracy).

parents showed very positive assessments[5] of the quality of education in the country. Whilst it can be difficult justifying investment of public funds where no public demand exists, the Mexican government has incorporated the education performance indicators within their reform plan, to explain to the public the need for change.

Finally, where an active civil society exists, progress measures can act as accountability mechanisms. Although societal outcomes depend on much more than government actions, in areas where indicators show persistent lack of improvement, the indicators can be used by advocacy groups to put pressure on policy-makers to target problem areas.

Challenges and benefits for developing countries

It is sometimes argued that the development of a broad set of progress measures is a luxury that developing countries should not spend resources pursuing. But if one considers 'progress' as synonymous with 'development' then a set of measures might be rather useful for countries in the South. There is often little discussion about what development means, or should mean, for a developing country. Too often there seems to be an unspoken assumption that 'development' means becoming more like the West. Whilst most, if not all, developing countries do indeed seek more Western standards of living, levels of infant mortality and so on, it cannot be assumed that this is true for all dimensions of their progress and wellbeing. Aspects such as social capital and cohesion, or cultural factors, might be much more important to (and stronger in) some 'developing' countries than they are in the 'developed' West. Countries might well be reluctant to sacrifice these things in their quest for economic growth.

This is a discussion that each country should have for itself. But it is a discussion that will benefit from a set of progress measures. Only then will the citizens of a society understand how their community is changing as it develops, and whether the benefits of development outweigh the negative aspects which might accompany economic growth. Only then will they be able to change course if they do not like what they see.

---

5   77% of parents interviewed reported that the quality of educational services was good or very good (IFIE-ALDUCIN 2007).

As one might expect, there are challenges for a developing country in producing a set of progress measures. These challenges are similar to, but more pronounced than, those facing any nation attempting such an initiative. There is likely to be less suitable data in a developing country and also less resources to undertake an initiative, particularly when many countries are striving to meet the statistical burden already set by international processes such as the *Millennium Development Goals (MDGs)*, and the *Poverty Reduction Strategy Papers (PSRP)*.

An initiative to measure progress can however, also help build statistical capacity in a country and offer benefits beyond those which could be delivered by the existence of only the indicator set. The process itself can be very beneficial. Jean Le Nay, argues that, for most least-developed countries: "Compiling economic and social information remains a minor concern, even when it is included in the PRSP. A radical change in this situation, in this ranking of priorities, is needed. Thus, a mobilisation is required. However, it would be misguided to think that an information system could be built for the purpose of development steered from outside the country. Furthermore, such a mobilisation must not be aimed at statisticians themselves, although they remain the main contacts of development partners, but at the potential users of information: political decision-takers, the civil service of course, but also the private sector and civil society, since they can play their role only if they have access to an efficient information system ……. Supply must of course be improved to create and stimulate demand while helping build a wide consensus on the need to promote economic and social information to the level of a priority in any country's development strategy." (Le Nay 2009)

Le Nay goes on to note several steps to overcome this situation, which included promoting demand among civil society; the private sector; public finance and sector ministries; and, paying attention to governance issues. Le Nay argues that statistical supply needs to be adapted to better meet demand, which must include "widening the field of observation and analysis".

An initiative to develop progress measures will not solve all of these problems on its own. However, experience from more developed countries suggests that such an initiative can help a statistical office to better engage with the world outside statistics and so both to stimulate demand and to better understand the needs of users. Moreover, one can imagine that conversations about 'measures of

progress' are more likely to engage a more influential group of external stakeholders than conversations about 'statistical priorities'.

The benefits of a country-owned process to define and measure development, involving a diverse group of stakeholders, can also strengthen democratic ownership of a country's development plan, something which Andrew Ellis argues is often lacking:

"Although highly valued as one of the guiding principles of international development cooperation, 'national ownership' is often little more than ownership by the executive branch of government. Not only does this exclude popular input into each individual cooperation programme, it creates a perception that development is solely or primarily technical, and thus beyond the area of political debate. There is still a long way to go before 'national ownership' of development in practice means *democratic* ownership" (Ellis 2009).

References

Ellis, A. (2009). *Making Room for Politics: the Role of the Legislature in Oversight and Measurement – and Why the Technician's Approach isn't Good Enough*, Presented at 3rd OECD World Forum, October 2009: http://www.oecd.org/dataoecd/32/63/44110395.pdf

European Commission (2009). *GDP and Beyond : Measuring Progress in a Changing World*, Communication from the Commission to the Council and the European Parliament, Brussels, 20th August 2009, COM (2009) 433 final.

Jackson, T. (2009). *Prosperity without growth?: the transition to a sustainable economy*, Sustainable Development Commission, www.sd-commission.org.uk/.../prosperity_without_growth_report.pdf

Le Nay, J. (2009). *Strategy for statistical development: an introductory note*, presented to PARIS21 Consortium Meeting, Dakar November 2009 http://www.oecd.org/document/13/0,3343,en_21571361_41755755_41767373_1_1_1_1,00.html

Stiglitz, J. (2009). *Progress, what progress?*, *OECD Observer*, No. 272, March 2009.

Stiglitz, J., Sen. A., Fitoussi, J-P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*, www. stiglitz-sen-fitoussi.fr

AVOIDING FAILURE IN PUBLIC PROGRAMMES

by Billy C. Hamilton,
former Deputy Controller of the State of Texas,
and adjunct Professor of Public Affairs at the LBJ School of Public
Affairs, University of Texas, Austin, USA

by Jody Zall Kusek,
Advisor on Results and Capacity Development,
World Bank Institute

Introduction

In early 2004, an international organization invested millions of dollars in Latin America to fight a growing HIV/AIDS epidemic. Five projects were designed to be part of a regional programme aimed at reducing the transmission of the HIV virus in certain vulnerable populations. There was high level support for these projects both from the sponsor and the client. The programmes were designed using best practice that had been tested elsewhere in Latin America and focused on a series of interventions which were expected to be successful in mitigating the HIV problem in these countries.

Expectations were high, and yet, from the outset, the programme was seriously flawed in fundamental ways. The lines of authority over HIV/AIDS and all other health problems were confused, and no single organization had overall responsibility for monitoring the AIDS epidemic. It became apparent during implementation that the programme's goals could not be measured because there was no working system for monitoring and evaluation. Some of the programme goals were unrealistic, and soon, the government began to spend far too little of the available funds to be able to achieve any of the expected outcomes of the programme. Still, the programme continued to be implemented with little change to the design and no significant change to implementation. Even regular programme reviews failed to halt the programmes or make needed revisions. Eventually, the programmes failed.

This is not an isolated example. In programmes of this type, failure is common, and yet, despite a sometimes patchy track record, these programmes are not designed to fail. Failure is far from the

minds or desires of either the sponsors or the clients. But still failures persist. Why?

In some cases, the explanation lies with the complexity of the initiatives and the limitations of the governments involved. Today's developing governments are often forced to undertake complex programmes which touch on central aspects of society, including health, education, private sector, and the treasury. At the same time, the international community is placing greater demands on these governments to "do more with less," and show that results are being achieved with donor funds. The result of these clashing demands can be failure.

Where the international community is concerned, the problem is more one of resource limitations to meet burgeoning demands. It faces the increasing need for humanitarian efforts. Moreover, continued conflicts, earthquakes, floods and other "acts of God" emergency funding, and the need for rapid disbursement of funds is becoming the norm in providing international assistance. Such a situation creates an environment rife with the potential for failure, both because of poor decision making, and of the lack of effective tools for monitoring the results of decisions made in response to sudden and unexpected needs.

These situations are not unique to the world of international development. Recent history offers many examples of programme failures, both large and small. These failures typically do not result from a lack of good intentions or from the lack of genuine effort. The root cause of the failure lies in the complexity and scope of the programme, and how people cope with that complexity during programme implementation and beyond. Most government programmes, including development programmes, involve many processes, people and stakeholders. Demands are high, but resources are limited. Given the complexity of these programmes and the economic constraints which often accompany them, it is no surprise that failures happen; it is amazing that most programmes do, in fact, succeed.

When programmes fail, the failure can often be traced to a lack of attention to the possibility of failure and how it arises. In *The Logic of Failure*, Dietrich Dorner (1996), points out that "failure does not strike like a bolt out of the blue; it develops gradually according to its own logic." Failures can result from many factors, or combinations of factors. A very short list of potential problems might include: ill-defined development goals; poor planning; underestimated time requirements; inadequate resources; and ineffective

monitoring and evaluation. There are many more. In complex systems, failure typically occurs not because one thing goes wrong but because multiple, sometimes apparently innocuous, failures join to create an environment ripe for failure.

Too often, when the unexpected occurs, managers of development programmes are unprepared. They do not see the problem coming, and when it arrives, they over-react, under-react or choose the wrong strategy for dealing with it, often exacerbating the situation in the process. Once problems escalate to the point of failure, a fix may or may not be possible, but the cost will be significant in money, time and human resources.

Why is the failure to anticipate the unexpected so commonplace? The answer is complicated, and various experts have offered a range of explanations. These explanations are not limited to development programmes but are more generally related to the inherent difficulty people have in dealing with complex systems. They form fuzzy or ill-conceived goals. They do not set clear priorities. They do not pay enough attention to the process as it unfolds. They fail to recognize and correct errors as they occur. They do not learn from the past or from the experience of others.

This does not mean that failure in complex programmes is a foregone conclusion. People can learn how to deal with complex systems, how to recognize the precursors of failure and how to deal with them. They can, in short, learn to pay attention to the possibility of failure. University of Michigan professors Karl Weick and Kathleen Sutcliffe (2007), have called this attention to the possibility of the unexpected occurring "mindfulness." They contrast it with "mindlessness," that is, a situation in which "people follow recipes, impose old categories to classify what they see, act with some rigidity, operate on automatic pilot, and mislabel unfamiliar new contexts as familiar old ones." Mindlessness is not a lack of attention – or at least it is not that alone. It is the absence of a predisposition to the idea that things can go wrong – that failure is not just a possible outcome of a programme but can be the likely outcome if programme managers are not mindful of the risks inherent in the undertaking and are not ready to respond to them in a timely way.

Mindlessness in development management is not inevitable. To borrow Weick and Sutcliffe's word, development managers can become more mindful. The mindful manager is aware of the threat of programme failure. The mindful manager is cognizant of the fact

that failure is a potential result of any public undertaking. The mindful manager takes steps to avoid programme failure. Those steps begin with a comprehensive and effective monitoring and evaluation system, but they do not end there. The best information systems in the world will not protect a programme if the insights it brings are not analyzed and understood; if warning signs are not heeded as they arise; and, if the manager does not have a predilection for action to resolve small issues before they cascade into massive, unavoidable failure.

Our theory is that it is more than good luck or good design which determines the success of a development programme. It is a conscious decision to imagine worst case scenarios, to guard against their occurrence and to have a plan of action in place if they do occur, despite all prior planning. Any programme will face problems as it proceeds. Assumptions will not be borne out. Estimates will be incorrect. Mistakes will be made. The key is to correct these problems during implementation and to prevent them from compounding into failure. Ironically, then, the path to successful development begins with an understanding of why programmes fail.

What is programme failure?

Failure and success are often thought of as alternative conditions. This is misleading. They are very different phenomena. A common example of this distinction is a bridge collapsing. The failure of the bridge is an event; the fact that a bridge doesn't collapse is a state. Failure, in this sense, is the culmination of a series of events that lead up to the failure. There are many possible avenues of failure, but in most cases, one chain of events leads to failure. Thus, poor planning in the design of the bridge, substandard materials, and too much stress on the bridge at a particular moment can cause the collapse. Some, perhaps most, of those elements may have been present for a long period of time without a failure occurring. Some final element, or elements, is needed to trigger the collapse.

A second feature of failure in development programmes is not paying attention to the political economy. History is littered with examples of programmes which have been judged a success, although a closer examination of what actually happened would suggest that they had, in fact, failed. By the same token, success may be unfairly judged to be failure. The harsh reality of public management is that failure can be "spun" as success or at least perceived as something other than failure, and success also can be seen as failure.

Mark Bovens and Paul t'Hart (1996), addressed this paradox in their examination of what they call "policy fiascoes." They offer a broad definition of policy fiasco that seeks to encompass both the facts of, and the perceptions of a failure. A policy fiasco is, they write, "a negative event that is perceived by a socially and politically significant group of stakeholders in the community to be at least partially caused by avoidable and blameworthy failures of public policy-makers." As a simplified model, then, here is a policy fiasco in a nutshell:

- Something bad happens in a programme.

- Someone of sufficient political influence notices.

- The failure is publicized.

- Blame is assigned.

It is, in effect, a variation on the old joke about the six stages of a project: enthusiasm; disillusionment; panic; search for the guilt; punishment of the innocent; and, reward of the non-participant.

In his book *Why Government Programs Fail*, James S. Larson (1980), defined "failure" broadly as it applies to public programmes, encompassing both the programmatic and the political nature of problems. "Failure," he wrote, "is any significant short-coming in a government programme that brings about subsequent changes in the law or in its implementation." Again, there is the triggering event (the short-coming), and the political result, in Larson's definition, taking the form of a policy response to the perceived failure.

In the most general sense, then, programme failures can occur in three circumstances. First, something that is expected to happen does not happen. Second, something that was not expected to happen does happen. Third, something that was not predictable happens. This is a useful simple division, but it is not sufficient because it omits the element of control i.e. what can a manager do, or what could he or she have done, to avoid the problem in the first place. In this regard, Bovens and t'Hart offer a typology of failures in programmes that takes into account the idea of controllability of failure. Their divisions are:

- Foreseeable but controllable.

- Foreseeable but uncontrollable.

- Unforeseeable but controllable.

- Unforeseeable but uncontrollable.

This final category suggests the distinction that they make between mis-management and misfortune. That is, the recognition that some results are simply the result of, for lack of a better term, bad luck.

Public managers are usually well aware that the success or failure of programmes can turn on events outside of their control. They generally don't make final budget decisions, and they can be taken by surprise by the same economic and natural events that surprise us all. "Man plans, and God laughs," as the old saying goes.

Our interest, though, is in the very large percentage of any programme that is not subject to uncontrollable events, beyond the control of programme managers. In this regard, attention to the possibility of failure is the best guarantee of success. Understandably, public managers may feel uncomfortable about such an inherently negative approach to managing public programmes. Programmes are, after all, typically designed and intended to produce a public good or to solve public problems. The point is not to be pessimistic but to be realistic in managing public programmes. By anticipating and solving problems, it is possible to avert their compounding and the resulting failure of a programme. The recognition and consciousness of possible failure is what mindful management is all about.

Where things go wrong

Potentially, in any public programme, many things can go wrong. Inevitably, some things *will* go wrong. However, the programme need not fail if the problems are recognized and dealt with before they begin to coalesce and cascade. In this regard, the first key to successful management is to recognize the precursors of failure through the information provided by the monitoring and evaluation system. The figure below shows one way of thinking about the precursors of failure. The sources of programme problems can be divided for convenience into four categories: organizational weaknesses; programmatic weaknesses; management weaknesses; and, externalities. As the chart indicates, the various categories can interact and combine to form a single pattern of failure. One factor typically is not enough to insure failure; it is the accretion of problems, often across all of the categories, that produces the sequence that leads to failure.

Organizational weaknesses. Public programmes are typically implemented within the context of larger ministerial oversight, and the actions of those ministries can have a dramatic impact on the probability of programme success. Again, few ministries implement

programmes with the idea of failure, but at times, the collective actions of the organization can contribute to failure. Often, organizational weaknesses plant the first seeds of failure.

An organization may have vague goals or lack a codified strategic direction, thus making it difficult to design the programme in a way that promotes success. Organizations often fall prey to over-optimism, promising more than can reasonably be delivered. This problem is often compounded by budget constraints or multiple, conflicting programmatic requirements. Later there may be a shift in the direction or scope of the programme due to slow fund disbursement by donors or agreement that goals are not likely to be met as planned. Also, as often happens, poor communication within the programme team can lead to mis-understandings that promote failure.

Management weaknesses. Much of the possibility of success or failure in a programme lies with the programme's management and the programme team, itself. To be successful, managers must plan adequately for the programme and anticipate things which can go wrong. Research shows that the root cause of problems in many programmes lies in the poor or inadequate decisions of the programme team and its management. These can take the form of: failure to involve stakeholders in the project; poor communication between team staff and staff within the larger organization; ineffective cost control; and, the lack of an efficient monitoring and evaluation system to measure the progress and success of the programme as it unfolds. The mindful manager is aware of these pitfalls and actively works to avoid them.

Figure 1: Four interacting sources of programme failure
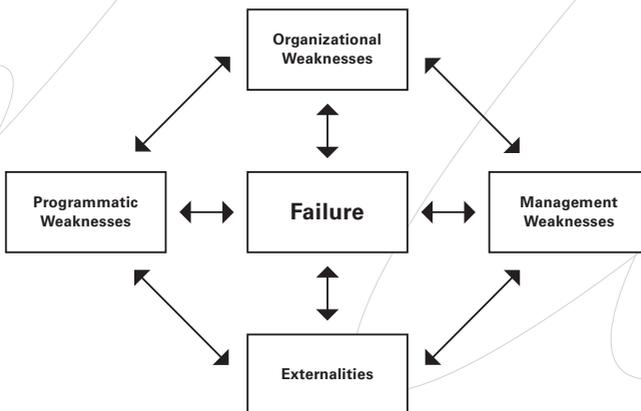
Programmatic weaknesses. Closely related to managerial weaknesses are programmatic weaknesses. Public programmes are often complex and often have narrow, politically defined timelines which can pose real challenges for the success of the programme. The potential for failure often begins with the design of the programme, particularly where the application of technology is involved. Poor design specifications in complex programmes is frequently a source of problems.

There is often a gap in the understanding of the programme, responsibilities and expectations, between the government public managers and the international funding partners, who have programme oversight. Managers, staff, donors, and other stakeholders must be willing to think through programmes from the beginning to the end to be certain that contingencies are correctly anticipated and that realistic timelines are set. The progress of the programme must be closely monitored on a routine basis and corrective action must be taken as needed to insure the programme stays, as nearly as possible, on time and on budget.

Externalities. The fourth category of potential precursors to failure is the one that is the most difficult for even the best managers to influence, but it is one that must be recognized. These are external factors which may influence the success of the programme. In public programmes, this can be something as simple as the proverbial "act of God," a situation that can not be foreseen but which has a critical adverse effect on the programme.

Depending on the programme, these externalities might include such factors as a change in the government, the weather, or changes in economic conditions. More probable, though, in public programmes is the impact of external political factors. Some government programmes are controversial. There may be international scrutiny or public opposition, such as stigma noted in some countries to actively help HIV/AIDS victims get the programme support they need. This opposition may arise unpredictably and have unpredictable effects on the success of the programme. While these situations may be temporally uncontrollable and unpredictable, they are, typically, not unknowable. The mindful is aware of the potential for external disruptions, considers ways for dealing with them and has contingencies for when and if they arise.

Figure 2 expands on these four categories, showing specific examples of each weakness. There are many more examples which can be taken from the experience of most programme managers.

In search of mindfulness

Most of the precursors of development failure are recognizable to anyone who has managed or been involved in the implementation of a public reform programme in a developing country. Most participants could probably add to the list. Since these factors are often well known and recognized, it raises the question of why they are not intercepted more often in practice. Why are well known risks to development programmes allowed to multiply with few real strategies put in place to manage these risks? This potential for incipient failure can be part of the antecedent strategy of how programmes are managed to achieve desired results.

The faults which allow the elements leading to programme failure can take many forms. In some cases, managers and their organizations develop expectations for how a programme should proceed, and these expectations can produce a blind spot where a potential point of failure can fester and grow. In other cases, the manager may become distracted by one apparent problem and focus on the wrong signal, allowing other problems to develop. In other cases, poor or inadequate communications or information systems may present problems. Other cases may be due to inherent problems being under-estimated or under-valued even when they are recognized.

Based on our studies of successful programmes, there are certain traits that seem to be present in most or all. Among these, the most important are the following.

1. **Clarity on what success looks like**. Successful programmes start with clearly defined goals. The programme management team converts those goals into detailed plans, monitors progress in detail and makes adjustment during implementation – and quickly – if the programme does not proceed as originally expected.

2. **A shared concern within the programme team for identifying possible failures**. The final responsibility for the success or failure of a programme may lie with programme management, but the full responsibility cannot reside there alone. All members of the programme team must be alert to problems as they arise

and must be encouraged to communicate concerns and to suggest solutions. Management, in turn, must be sensitive to input from the frontlines where the real work is done and be proactive in dealing with problem situations which are brought to their attention.

Figure 2: A programme failure taxonomy

A Programm Failure Taxonomy	
Organizational Weaknesses • Lack of executive support • I11-defined goals • Conflicting priorities (too many programs) • Rigid perceptions and beliefs about the organization • Changes in program scope • Poor communications	**Management Weaknesses** • Irrational promises • Inadequate stakeholder involvement • Poor planning for contingencies • Poor program management • Poor quality control • Confused lines of authority • Poor decision making structure • Poor communications • Poor cost control • Inadequate monitoring and evaluation • Information difficulties
Programmatic Weaknesses • Over-optimism • Incomplete specification of requirements • Inadequate training • Human error – bad judgement • Groupthink • Inadequate or failed technology • Unrealistic deadlines • Multi-year project horizon • Poor contracting practices • Poor communications • Inadequate resources • Insufficient measurable outcomes • Scope creep	**Externalities** • Changing political priorities • Changing demands • Externally imposed restraints • "Acts of God"

3. **Realistic and flexible norms and rules about possible failure points.** This is a function of thorough planning. It is, of course, impossible to foresee every contingency that will arise in implementation and management of a programme, but thorough planning and flexibility in adjusting plans is essential.

4. **Data for decision making.** An understanding of the need for continual examination of the programme through monitoring, analysis and feedback systems. Many organizations have a difficult time dealing with changes, particularly if they require changes in well-established practices and norms. Successful programmes stay alert to changing situations and how various mistakes arise and are flexible enough to change the organization's processes so that they do not recur.

5. **Open and continuous communication.** Mis-communications, or the failure to communicate at all, lie at the heart of many programme failures. Successful programme managers and teams recognize the limits of their own perception of the programme and encourage communications within the programme as a way of surfacing problems which might not otherwise be available from their own vantage point. Also important is communication with those outside the programme – the potential users, the stakeholders, the overall organizational management – to make sure that the progress of the project is understood and that concerns are considered as they arise.

6. **Clear responsibilities, accountabilities and decentralized decision-making, where appropriate.** All team members should have a clear understanding of their roles and duties. There is clear awareness of what exactly is expected from them. Decentralized decision making permits prompt and flexible field-level responses to surprises.

7. **A refusal to over-simplify problems.** It is certainly true that most programme situations require a certain amount of simplification to avoid becoming lost in the details. However the task of implementing or managing a programme is fraught with potential complexities and the more information that is available, the more likely incipient problems will be caught as they arise.

8. **An ability to "imagine" possible failures.** To be able to avoid possible failures, managers and programme teams must be able to conceive of the worst possible occurrence. There are a number of techniques for creating an organization that has an

active imagination when it comes to anticipating and dealing with potential problem. Managers can use good meeting management techniques to elicit various viewpoints from within the organization. They can plan "what if" games to anticipate potential problems before they arise.

9. **A refusal to assume that everything will work right the first time and that every contingency is known and prepared for.** This is the chief operating reality of a successful programme. It promotes a dynamic tension within the organization which translates into alertness for problems as they arise – not after they have escalated.

10. **Continual dissemination of lessons learned from his or her experience and the experiences of others.** Every programme is an opportunity to learn lessons which will improve future programmes. The reality of public programmes is that negative feedback must be provided internally, or it will most certainly be provided externally.

An Assessment checklist for the mindful manager

Problems are inevitable in any development programme; failure is not. Our goal has been to show, in a broad sense, how programme failures germinate and take root and where to look for the sources of potential failure in any programme. With that in mind, the following is an assessment checklist that can be applied to any development programme to guard against programme failures. Obviously, a simple checklist is no substitute for alert and insightful management or strong monitoring and evaluation systems. It does not eliminate the need to be willing to make mid-course adjustments when the need arises, but it can help spot the potential for failure in a programme, and that is the first step towards avoiding failure.

Figure 3: Programme failure assessment checklist

Part 1: Organizational weakness

1. Is there a mission statement and/or strategic plan that codifies the goals and national or sub-national development objectives from the government?

2. Are programmes prioritized and aligned to the annual national budget?

3. Is THIS programme among the prioritized and funded programmes of the organization?

4. Is there a champion for this programme, and if so, does he or she have decision making authority over resource allocation?

5. Are there formal mechanisms to disseminate information across the organization? Are there informal mechanisms? Do they work?

Part 2: Management weakness (of the programme)

6. Are there clear lines of authority for decision-making?

7. Have lines of authority been clarified to the client?

8. Is there a codified programme management plan, including oversight criteria and regular assessment of progress and quality?

9. Do the programme manager and client counterparts have the requisites skills in project/programme management?

10. Are key programme indicators of success monitored at a regular basis?

11. Does the programme have an effective (that is, well-managed) cost control system?

12. Are there effective lines of communication between the programme manager, members of the team and other key actors within the larger organization?

Part 3: Programmatic weaknesses

13. Are the development or programme goals rational?

14. Has the manager signed off on all elements of the design?

15. Is the manager involved in day to day programme implementation?

16. Is monitoring and evaluation data used to improve the implementation of the programme or used to alert the organization that goals are or are not being met?

17. Are appropriate resources available to implement the programme on time and of the desired quality?

18. Are there regular programme reviews, quarterly, mid-term etc to assess progress against planned goals?

Part 4: Assessing the role of externalities

19. Is this programme likely to be affected by a change in government?

20. Has there been or is there likely to be opposition by counter-reformers with strong political or popular support?

21. Have risks from "Acts of God", to successful implementation, been considered in management planning?

Conclusion

We live in a world of increasing complexity and extraordinary international development challenges. The demands to develop sound programmes which deliver the results promised have never been higher. At the same time, the potential for failure has never been higher. In this environment, things-going-wrong is an all too common modern management experience. Development programmes inherently involve trade-offs between programme goals and other pressures, such as budget considerations. Pressed for time, the manager can make a hasty decision that remedies the problem but creates a myriad of new problems. The list of potential problems is seemingly endless. Although programmes are launched with the best intentions, too often managers court failure in predictable patterns, from simple confusion and mis-perception to short attention spans, the failure to collect and understand performance data, and unwillingness to change tactics.

However, all is not lost. Despite the potential for failure which is part of any large development programme, managers can, through the information provided by monitoring and evaluation systems, learn to recognize defective organizational and management behaviors and to correct them. Our goal in this chapter has been to encourage managers to recognize that an understanding of the sources, and the potential, for programme failure are, in fact, part of a strategy for promoting programme success. By understanding how failure arises within a programme, it is possible to develop some basic understanding of how the possibility of failure can be reduced or eliminated.

Begin with our simple assessment of the major sources of programme failure, but do not stop there. Our goal is to instill in programme managers the certain knowledge that failure is a potential in any public programme, and to encourage the manager to develop a mindfulness for this potential. Problems will always be part of any programme, but they can be identified and controlled before they escalate to the point where they produce a full-scale programme breakdown. The best managers recognize the stress points in their programmes. They collect and analyze data with an eye toward developing problems. They admit problems as they arise, and they are flexible in their response. When mindfulness replaces mindlessness in our programme management, then success will replace failure. It is that simple – and that difficult.

References

Boven, Mark and t'Hart, Paul. (1996). *Understanding policy fiascoes*, New Brunswick, New Jersey: Transaction Publishers. p. 15.

Doerner, Dietrich. (1996). *The Logic of Failure: recognizing and avoiding errors in complex situations*, New York: Metropolitan Books, 1996, p. 10.

Larson, James S. (1980). *Why Government Programmes Fail. Improving policy implementation*, New York: Praeger.

Weick, Karl E. and Sutcliffe, Kathleen M. (2007). *Managing the Unexpected*, 2nd Edition, San Francisco: John Wiley & Sons, Inc. p. 88.

EVALUATING POLICY: A PERSPECTIVE FROM MULTILATERAL EXPERIENCE

by Caroline Heider,
Director, Office of Evaluation,
World Food Programme and Vice-chair
of the United Nations Evaluation Group

Introduction

This paper suggests three key evaluations questions to be addressed in policy evaluations: understanding the quality of a policy, the results it has achieved, and how these results can be explained. These questions generate evaluation evidence that is concrete and usable by policy-makers and implementers. It discusses the implications of policy evaluations for the evaluation principles of independence, credibility and utility, as well as the challenges embedded in addressing the three key questions.

The paper is based on the practical experience of evaluating policies in the context of several multilateral organizations. It is an approach that has been shaped over several years in the course of a number of policy evaluations. The paper is structured to introduce a definition of policy so as to ensure a common understanding of readers, a discussion of evaluation principles and how they are affected in the case of policy evaluations, and a detailed explanation of the three key evaluation questions proposed here. In so doing, the paper covers why the proposed evaluation questions are important, the associated stakeholder issues and how stakeholders will use the evaluation evidence, as well as methodology issues. Examples are provided to illustrate the points.

Policy – What is it?

This paper uses the definition of policy commonly found in dictionaries: "a definite course of action to attain specific goals".[1] Policies aim to influence behaviour: they either aim to make sure that people maintain certain behaviour or change it in specific ways. For instance, a policy that subsidizes students from low-income households aims to encourage young people from this income-group to enrol and pursue studies and thereby gain greater equality in building human capital. Policy, therefore, is something that entails the stated intent of attaining a desired result.

Policy can be determined by an authoritative figure or institutions, or can be defined through participatory processes.[2] This latter approach of *evolving policy* recognizes the complex interplay between "three levels of social reality": society, institutions (or also rules, roles and relations as well as policy), and the individual.[3] These three levels influence each other: society is made up of individuals, their norms are expressed in institutions, but both society and norms determine behaviour of individuals. Changes in society (the collective of individuals) are driven by individuals and institutions which at the same time influence each other. Policy is in part the expressed norms of society and in part an effort to change behaviour through measures which create incentives or disincentives for changes in behaviour. The complexity that results from this complex reality has implications for the more linear and simplified models such as logical frameworks and, as some would argue, also for the extent to which policies are, or can be, goal oriented.

---

1   Policy is defined as: a definite course of action adopted for the sake of expediency, facility, etc.; a course of action adopted and pursued by a government, ruler, political party, etc.; action or procedure conforming to or considered with reference to prudence or expediency (see http://dictionary.reference.com/browse/policy). Or, as defined in Webster: a definite course or method of action selected from among alternatives and in light of given conditions to guide and determine present and future decisions or a high-level overall plan embracing the general goals and acceptable procedures especially of a governmental body (http://www.merriam-webster.com/dictionary/policy).

2   For instance, Carlsson speaks of "all policy areas in contemporary society can be characterized as complex: a great number of actors are involved who together pursue a multitude of goals." Hanberger suggests that "[c]onsidering different stakeholders' arguments is more urgent today because the state and its representatives no longer have unquestioned authority and legitimacy." And van der Meer suggests that "policy-making and implementation are increasingly shaped by the interplay between the efforts (concerted or not) of different actors, from both the public and private spheres".

3   Sanderson refers to the work of Reed and Harvey (1992) and Bhaskar (1998) from which the "transformational model of social action" is derived.

In either case, some common aspects to policies are of particular interest to policy evaluation:

- **Stakeholders.** There is a multitude of stakeholders: policy-makers, policy implementers, those affected by policy prescriptions and those monitoring policy outcomes. Each of them will have different interests. Relationships are complex, be it because of power structures, different contexts and backgrounds of people, and/or because of relationships that mutually influence each other.

- **Policy Objectives.** Policies, defined as "definite course of action to attain specific goals", implicitly or explicitly pursue objectives. These are sometimes stated clearly, in other cases they are left vague in particular when the agreement of stakeholders is difficult to attain.[4] In such cases, it is easier in the short term to approve a policy that has ambiguous objectives rather than negotiate an agreement. In other cases, policy-makers are not sure of the outcomes a policy will attain and therefore do not want to specify expected outcomes.

- **Assumptions.** Even when objectives are stated clearly, underlying assumptions about how the expected changes will follow from policies are not spelled out in detail. They are dependent on cultural context, value systems, diverse perspectives and responses to policies. Therefore, this dimension is particularly complex in an international and multilateral context, not necessarily within the organizations themselves, but working across countries with people from different backgrounds.

Evaluation principles: implications of policy evaluations

This paper adopts three key evaluation principles – independence, credibility and utility – as fundamental to the quality of evaluations (see Box 1). These are discussed in greater detail in the *Conceptual Framework for Developing Evaluation Capacities*.[5]

---

4   Van der Meer and Edelenbos speak of "considerable ambiguity in and around policy processes: lack of shared goals, lack of information (or abundance of information but lack of meaning) on what the problems are, what causes them, what will follow from possible measures".

5   Conceptual Framework for Developing Evaluation Capacities, Building on Good Practice in Evaluation and Capacity Development, Caroline Heider, forthcoming.

Box 1: Evaluation principles

- *Independence* forms the bedrock of good evaluation practice. It is fundamental to attain credibility and utility of evaluation. It should lead to impartiality reflected in, for instance, the choice of subjects of evaluation or the selected evaluation method. It is achieved through means at structural, institutional and individual levels.

- *Credibility* of evaluation is enhanced with greater independence but needs to be accompanied by competence of the evaluators, transparency of the evaluation process, and impartiality of evaluators and process.

- *Utility* of evaluation is not guaranteed through its independence and credibility, but requires that commissioners and evaluators undertake the evaluation with the intention to use its result, undertake it at a time when the results can meaningfully inform decision-making processes, and that evaluations are accessible. This principle exists to ensure evaluations are undertaken to influence change so that governments and organizations can achieve their objectives and achieve them better.

Together these principles are markers of high quality evaluation and ensure "good practice" in evaluation. Independence, credibility and utility are three equal sides of the triangle at the centre of which the quality of evaluation rests. As the cited article explains, these principles do not exist in perfect harmony. Some stakeholders believe that only by controlling evaluation results will they be useful to them, seemingly unaware that such partial evaluations will have only partial utility and can adversely affect credibility. If evaluations cannot safeguard against influence, they will have limited credibility with the cross-section of stakeholders.[6] Equally important, evaluations that do not maintain impartiality are of little, if any use in solving problems. They contribute little to learning: when problems remain unknown or covered up, solutions cannot be found to them and stakeholder will continue to be hampered in achieving their objectives.

The principles apply in general, yet, for policy evaluations additional considerations are warranted.

The multitude of stakeholders has an effect on all three principles. The challenge of safeguarding independence can become more pronounced when policy-makers want to pursue certain policies

---

6   As Boyle put it "[e]ducated consumers can help create an evaluation "ethos" where evaluation is valued as an integral part of the government decision-making process.

and fear evaluation evidence would advise otherwise. In such cases, stakeholders with strong policy interests might want to influence the way in which an evaluation is conducted, or its findings. They can be among the proponents of the policy or opposing it. They can be policy-makers or those affected by it. These politically motivated interests are understandable, but evaluators need to balance the call for this form of *utility* with the purpose of evaluation, namely to provide accountability and learning, and its principle of impartiality.

Some authors who speak about the complexity of policy-making and policy implementation processes also suggest that evaluation should be part of policy-making[7] rather than be independent. There are merits in such close inter-relationship between policy-making and ongoing or self- evaluation: learning can be embedded throughout the processes, especially when new policy measures are being tested and uncertainty exists about outcomes. However, this paper argues that there is also merit in independent evaluation to support policy-making and implementation "at arm's length" to gain an independent view (in additional to self-evaluations) and to reduce potential political pressures to report policy success regardless of supporting evidence. This paper focuses on independent evaluations without disregarding the value and importance of in-built self-evaluation into policy-making processes.

To ensure independence it is important to understand stakeholders and power structures to determine how they influence policy-making and, potentially, policy evaluation:

- The **power structures** between stakeholders, especially policy-makers in a position of power and those affect by policy who may be disenfranchised, can create strong polarities. In particular, it is important for the evaluators to understand and listen to those who are affected by policy, but cannot influence it.

- The **interests and convictions of stakeholders** in policies can be strong because of the political stake that can be gained or lost since policies are associated with investments of public resources for policy implementation. Evaluators have to safeguard against being absorbed into these policy interests.

---

7 Frans-Bauke van der Meer and Jurian Edelenbos, *Evaluation in Multi-Actor Policy Processes, Accountability, Learning and Co-operation*, Evaluation, Sage Publications 2006.

- The *assumptions* stakeholders make are often "hidden" in that they are internalized and assumed to be shared. Often this is not the case nor realized until well into policy implementation. Evaluators have to be aware of their own hidden assumptions to counter potential biases.

Box 2: Understanding stakeholders

To understand stakeholders, a number of tools exist:

- **Stakeholder matrixes** to summarize the roles of different stakeholders in policy-making and implementation, their interest in the evaluation.

- **The "power/interest grid"** that helps divide stakeholders into groups with different levels of interest and power and suggests, based thereon, the relationship that the evaluation (manager and/or team) should have to each of the stakeholder groups.

- **Venn diagrams** to illustrate the importance and proximity of each stakeholder to the policy or the locale where policy results should be observed. Venn diagrams provide a powerful illustration of the relative influence of each stakeholder group and the relationship between them, and differences in perceptions.

- **Accountability maps** describe who is accountable to whom, which is particularly complex in policy evaluations when policies are the result of negotiation processes that leave accountabilities undefined, or where accountabilities are dispersed.

- **Force-field analyses** provide a tool that help understand who may be supporting or hindering policy adoption or implementation and their reasons for doing so. The same tool can be applied to the evaluation itself to determine who stands to gain or lose as a result of the evaluation.

For evaluators it is important to use these tools to gain an understanding of who the stakeholders are, what they stand to gain or lose and how they will relate to the evaluation. Maybe not all tools need to be used in all cases, but a thorough stakeholder analysis needs to be done at the time of designing the evaluation and to manage stakeholder relationships throughout the process to safeguard independence and impartiality of the evaluation.

Ensuring credibility of policy evaluations is also challenging, not least because of the importance of impartiality as already discussed. The competence of evaluators, including their ability to understand and manage complex multi-stakeholder processes, and the transparency of evaluation processes is even more important in policy evaluations with different power relationships among stakeholders and increasing demands on accountability of policy-makers. Maintaining impartiality while demonstrating an understanding of the issues is essen-

tial for gaining credibility.[8] In addition, evaluation methods have to be designed to cope with complexity since a number of factors could influence each other and therefore can render simple linear logic models less meaningful especially given possible ambiguities in policy objectives and underlying assumptions.

Policy evaluations have led to questions about the utility of evaluations and their effectiveness. Considerable resources are spent on evaluation, the argument goes, but the use of their results is not evident in new policy design or in policy implementation. Too many policies are adopted without adequate evidence, and too many evaluations are not influencing policy-making processes. This situation can be explained in part by the way in which policies are made and in part by the way in which evaluations are conducted. Policy formulation involves complex processes which are driven by many factors[9] and not just by rational choices and evidence or insights from evaluation. Unfortunately, it is often not understood that the absence of impartial evaluations and subsequent actions to address problems can result in major losses: of confidence in policy-makers, of outcomes as a result of inappropriately used resources, and of opportunities which could otherwise have achieved more.

Figure 1: Evaluation principles and the implications of policy evaluations

---
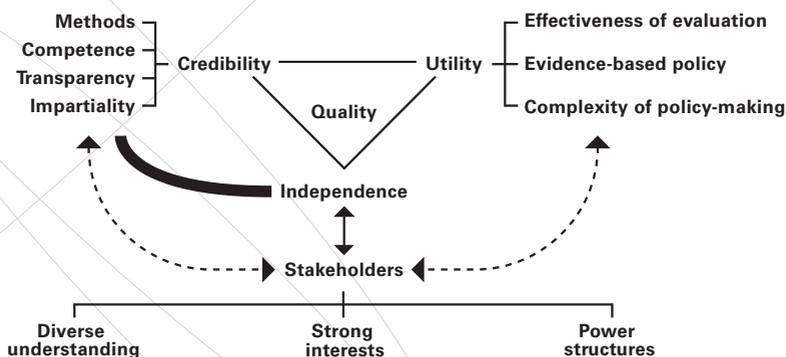
8   Buchanan-Smith, in the ODI working paper on the Sphere project, suggests an evaluation had been found credible, as it presented a clear and level-headed analysis of the situation.

9   Davies suggests that there are a number of factors that influence policy decisions, including experience, expertise, judgement, resources, values, habits, pressure groups, pragmatics and contingencies.

The above paragraphs help the understanding of where challenges arise in policy evaluations for the principles of independence, credibility and utility. These challenges need to be taken into account when developing evaluation capacities[10] and when designing and conducting policy evaluations. The following section explains how these challenges are addressed in the three key evaluation questions.

Key evaluation questions

The three evaluation questions proposed here take into account the challenges of policy evaluations to the above principles. They are based on practical experience and demonstrated their usefulness in evaluating policies. They integrate issues of accountability: was policy implemented and what results did it produce; with learning: what were the factors that affected the observed changes and what can we learn from them for policy-making and implementation. In addition, the first question focuses on the quality of policy itself, rather than accepting it as a given.

Box 3: Defining quality criteria

It is important for policy evaluations to define the criteria to assess the quality of policy and discuss and agree them with stakeholders.

Question 1: How good is the policy?

The importance of this question arises from the need to keep the evaluation independent from policy-making and to take a look at the quality of a policy itself. Taking policy "for granted" rather than including it in the scope of an evaluation would pre-suppose the policy as right. Such an assumption would introduce a bias into the evaluation, in that the evaluators would act on the assumption that the policy is correct, rather than generating an understanding of its virtue from the perspective of a number of stakeholders, and from a technical point of view. For instance, the policy-making process could have been driven by a few powerful stakeholders or

---

10   See the article on *Conceptual Framework for Developing Evaluation Capacities* for details on evaluation capacity development in general. Capacities for undertaking policy evaluations will need to be set up in ways that they are safeguarded against the interests of policy-makers, which can be difficult in national contexts, and develop the technical capability to deal with the complexities of policy evaluations discussed in this paper.

have involved many stakeholders through participatory processes, which would have affected the quality of policy in different ways. Or, the policy design might have been evidence-based, in which case assessing its quality against current knowledge can validate whether past evidence and understanding were accurate. In either case, it is important to understand how good policy prescriptions were in order to determine whether policy implementation, or a lack thereof, originated in its quality and was a positive thing or not.

The challenge lies in determining what is "good" policy and who decides this. Not many standards exist to define what is "good policy". Should it include a logic model, a result matrix, indicators or a costing of policy implementation? Or should it remain less prescriptive and give room to interpretation and adjustments? The Policy Hub[11] includes in its resource materials for improved policy-making papers that cover the following aspects: forward looking; outward looking; innovative and flexible; joined up; inclusive; evidence-based; evaluated and reviewed; and, include lessons. This list is indicative of what the UK National School of Government considers to be part of good policy. Picciotto[12] suggests that an important dimension of policy is its coherence, nationally and internationally. The evaluation of the World Food Programme's (WFP) capacity development policy[13] used three dimensions to determine its quality: the underlying logic, the coherence with internal and external policy frameworks, and its practicability. The last of these three dimensions was based on joint work of the European Union and the Organisation for Economic Co-operation and Development (OECD) to provide support for the improvement of governance and management[14] but adapted to the needs of the WFP evaluation. The criteria to determine practicability included: flexibility without being arbitrary; free of internal contradictions; unambiguous; clear and understandable; defined competences and responsibilities; estimated costs; and, efficiency of institutional set-up.

---

11   UK National School of Government, Policy Hub http://www.nationalschool.gov.uk/policyhub/.

12   The Evaluation of Policy Coherence for Development, Robert Picciotto, Evaluation, 2005, Sage Publications.

13   World Food Programme, Evaluation of WFP's Capacity Development Policy and Operations, 2008 (http://documents.wfp.org/stellent/groups/public/documents/reports/wfp216450.pdf).

14   Improving Policy Instruments Through Impact Assessment, SIGMA Paper: No. 31; 2001; http://appli1.oecd.org/olis/2001doc.nsf/43bb6130e5e86e5fc12569fa005d004c/c1256985004c66e3c1256a4f004b5bd4/$FILE/JT00107877.PDF

Technical standards can serve to benchmark the content of policy. "Good policy" needs to be technically sound. Technical Standards can be used for accountability, if policy-makers had committed to them previously or, if this is not the case, they can still be used for learning about the match of the current policy against latest thinking and technical standards. The evaluation of WFP's capacity development policy used international good practice for capacity development, which had been published by the OECD,[15] as a benchmark. The evaluation of rural finance of the International Fund for Agricultural Development (IFAD) looked to industry standards as a reference point.[16] The policies were, in both cases, compared to good practice to determine coherence and application of the latest thinking in the profession.

In some settings, policy might not be finite but evolving over time, be it because it is less clearly defined (in which case the evaluation faces challenges in determining the actual objectives) or be it because of changing parameters to which policy-makers respond, in which case the evaluation needs to determine whether changes were meaningful and timely. For instance, the United Nations Industrial Development Organization's evaluations of its programmes to promote human resource development, or economic and technical cooperation among developing countries,[17] involved analyses of policy directions given by the governing bodies of the organization. At the time, the governing council and industrial development board met frequently and adopted decisions and resolutions at each session. The evaluations analyzed how these policy directions evolved over time, and used the analysis to determine whether programmes were in line.

Stakeholder issues arise from the complexity of the policy-making process and the diversity of interests and understanding of the policy content. This complexity exists for all policy evaluations and is heightened in multilateral contexts. As indicated above, different stakeholders may make different tacit assumptions and

---

15 OECD/DAC, The Challenge of Capacity Development, Working Towards Good Practice, 2006.

16 The evaluation built on the Donor Guidelines on Good Practices in Microfinance of the Consultative Group Against Poverty (December 2004), see IFAD, Evaluation of IFAD's Rural Finance Policy, 2007 http://www.ifad.org/evaluation/public_html/eksyst/doc/corporate/rural.pdf

17 United Nations Industrial Development Organization, evaluations of the Economic and Technical Cooperation Programme (1992) and of the Human Resource Development Programme (1993).

have objectives which are only partly expressed in the policy. An evaluation has to draw out these differences and consider how to incorporate them in the evaluation design, since different stakeholder groups may consider the quality of the policy and its outcomes very differently. This can be done by analyzing written policy, and its implementation instructions or guidelines which provide interpretation of policy prescriptions, and by discussing with stakeholders their understanding of the policy, its underlying logic, its assumptions and intended outcomes. This part of the evaluation process in itself can be helpful to stakeholders in clarifying their own expectations and to understand differences between different stakeholder groups.

The users of the evaluation will gain a better understanding of diverse underlying assumptions and how they affect policy implementation and results, and thus can use evaluation results to improve on policy formulation. Equally, the comparison with benchmarks (good practice or industry standards) will help evaluation users to determine how close they want be to such benchmarks. However, the usefulness of evaluation does not mean its results will automatically be used: difference in policy objectives between different stakeholders might continue to exist, as different points of view and convictions about the course of action cannot be reconciled.

Question 2: What were the results of the policy?

Asking this question is essential to understand what difference the policy made. It is asked partly for accountability reasons, but also to learn whether policy prescriptions are effective or need to be changed to achieve results. Results can be observed at various levels. A policy might foresee a change in the way an organization delivers its services, thus the first level of results is at the organizational level. Policy could also suggest a change in partnerships, the results of which would be measurable in the changes in relationships. Ultimately, results should be measured at the level of people affected by the policy, as the policy impact shows-up in changes in behaviour.[18]

---

[18]   It can be challenging to undertake a policy evaluation that covers the three questions and includes an impact evaluation of people affected by policy, especially in multilateral organizations where policies have global implications. To ensure sufficient depth in impact evaluation while addressing policy questions, WFP is undertaking five impact evaluations of its school feeding programme and an evaluation of its school feeding policy. These six evaluations are being carried out over a three year period, 2009-2011.

Answering this question is methodologically challenging. Societal change – a change in behaviour following policy prescriptions is hardly ever linear. Changes are induced by many factors and an interplay between them, or situations may remain static for a number of reasons. These factors, including stakeholders, are often in complex interdependent relationships which influence each other. Therefore, simplified linear models (logical frameworks or results chains) can be useful to express the expected chain of events and underlying logic of policy-makers and implementers. However, these models might oversimplify reality and may not take into account the many varying factors which influence behaviour. itself, rather than accepting it as a given.

Box 4: Tracking intended and unintended results

Whichever evaluation method is adopted it is important to track intended and unintended effects of policy, positive or negative.

Evaluation methods need to cater for this complexity and for how to determine the contributions a policy has made to observed change. There are various ways in which evaluations can deal with this challenge. One is to follow the logical framework and collect data against the stated objectives and so establish the plausibility of these changes occurring as a result of the policy. However, ambiguous objectives pose a further challenge. If not stated clearly objectives need to be re-defined to understand the changes which the policy (policy-makers) intended to bring about. The dialogue with stakeholders, already done in response to question 1, will help the evaluation to clarify objectives and to determine how to assess progress against them. Alternatively, it is possible to work from good practice as the underlying model as done in the evaluation of WFP's capacity development policy. It used the good practice model to organize its inquiry into results and their presentation. Using this model was helpful in that it showed areas of strengths and gaps. Another approach is to observe changes in a setting influenced by the policy and then to collect information about the various factors which played a role to explain these changes (see question 3 for the discussion of factors explaining results). In either of these cases, it will be difficult if not impossible to "attribute" the observed changes to specific policy measures. At best a contribution analysis is possible, which requires the evaluation analysis taking into account a range of relevant related factors – policy measures of other agents, investments, unforeseen initiatives, etc.

Measuring behavioural change is challenging. Descriptions of the situation at the outset of the policy intervention (the behavioural pattern) might not exist to help understand what changes have occurred. Or behaviours might not lend themselves to direct measurement with the help of indicators, but require finding and testing proxies. Furthermore, it is important to gauge the sustainability of results: will the changes in behaviour last even if the policy measures (incentives and disincentives) were to be discontinued?

Box 5: Maintaining impartiality

The independence of evaluation can also be affected by potential biases of evaluators. Just like stakeholders in the policy-making, implementation and monitoring processes, evaluators will have convictions, make assumptions and be influenced by the values of their background. The commissioner's interest might determine the evaluators' understanding of the theory of change; something that is even more likely when the commissioner and evaluators come from the same background and have a "blind spot" for assumptions about behavioural change and reactions to policy prescriptions. In other cases, evaluation questions might be limited to compliance with policy and its implementation, without evaluating the validity of the policy itself. Therefore, evaluation managers and evaluation teams need to be aware of their own internal biases and about implicit biases to determine ways to manage these in the evaluation methodology and processes.

Stakeholders are instrumental to understanding the changes that occurred and which of them could have resulted from the policy. They may have definitions and criteria of success which are different from the logic model expressed in the policy or defined in the evaluation approach. Through participatory methods these diverse definitions of success can be brought into the evaluation (at the outset when designing the evaluation if there is sufficient time and the necessary resources, or in reporting back on the findings if the necessary steps are built into the evaluation process), which in turn can help redefine policy if necessary or warranted. However, stakeholders will also have strong vested interests, be it as policy-makers, implementers or those affected by policy, whereas the evaluators should maintain a balanced and impartial position.

The users of the evaluation will gain a better understanding of the success and failure of policy measures, and of unexpected effects of the policy. These insights will help policy-makers consider whether the resources for policy implementation are well spent. It can help policy-makers and implementers to correct policy, especially if the policy is generating undesirable unintended results.

However, this does not mean evaluation evidence will automatically result in improved policy-making and implementation. If the results shown in the evaluation do not support the convictions of policy-makers, they may question the evaluation and seek other ways to prove the policy is successful. There are also many examples where the political cost of changing policy is so high that policy-makers shy away from such decisions. For instance, there may be a need to revise policies on cost-recovery (or subsidies) for public services to enable maintenance and increase efficiency of service provisions.[19] But such changes would result in price increases for consumers, which often is politically hard to pursue. Then there can also be situations where policies are necessary, even if they are not as effective as hoped.

Question 3: What factors explain results and performance?

This question is essential to generate an understanding of why results occurred, or failed to do so. It is the question that generates the most learning about the factors which contribute to success and shortcomings, and about how policy has worked.

To answer this question, the evaluation has to understand the drivers of policy implementation. Partly, this has to do with stakeholder attitudes and commitment to policy, the relationship between stakeholders and the knowledge they have in common. It relates to the underlying assumptions about what motivates people and their behaviour, whether in support of policy implementation or actually changing in response to policy prescriptions.

Findings can relate to factors which are under the control of policy-makers and implementers, but also those outside their control, although the inter-twined relationships mean that a clear separation is not always possible. An evaluation will be perceived as "fair" when it distinguishes these factors, and it will be useful when it can explain some of the factors beyond the control of some stakeholders, since then it may be able to generate new approaches and solutions. The interplay between these factors will often be complex.[20]

---

19 There are many examples of this nature. One of them is the evaluation of the Asian Development Bank's capacity development assistance to water utilities, 2003.

20 DeGroff and Cargo suggest three main groups of factors that play a role in policy implementation that have immediate consequences for evaluation. These are: (1) networked governance, which emphasizes the interrelationship between agents of change; (2) socio-political context and democracy, which relate to power structures; and (3) new public management, which introduces results-orientation into governance.

For instance, a number of agencies will be responsible for policy-making and involved in policy implementation. Responsibilities might not be clearly defined and relationships will vary over time. Accountabilities will not be unique or even clear, and neither will the power structures. An evaluation will have to understand the network of change agents involved in the policy and the interrelationship among the network members. The reach of the network will depend on the policy in question and the scope of the evaluation; in a multilateral context, the complexity of networks multiplies, as the same policy is implemented in a number of diverse countries.

An evaluation can easily become overwhelmed with the number of stakeholders and the complexity of their interrelationships. In the case of three recent policy evaluations of WFP, the scope of the evaluations was narrowed down to focus on the programme's capacity to implement the policies under evaluation. This choice was made as policy-makers wanted to understand the comparative advantage of WFP. Other choices can be made when an evaluation is designed. For instance, these drivers can fall into a number of groups, which can be gleaned in part from the indication of what "good policy" should look like. Some examples are given below, but other unexpected explanations may arise and should be captured during evaluation:

- *An analysis of the underlying assumptions of different stakeholder groups* might reveal why results appear different from expectations; something that can be even more pronounced in an international context when policy-makers, implementers and those affected by policy come from different cultural backgrounds. In these cases, an evaluation can add value to the knowledge of various stakeholders.

- *Policy prescriptions may be highly relevant to the problem*, which increases the likelihood that, if implemented, policies will produce expected results. However, as indicated above, policy objectives might be ambiguous and may not be shared by all stakeholders, in which case the policy might be broadly relevant, but not significant enough to achieve desired change.

- *A policy might be formally adopted, but resources are not available to implement it*. The actual policy commitment, thus, has to be gauged from the human and financial resources and systems put in place for policy implementation. The efficiency of use of resources can also explain results, or the lack thereof.

- *Partnerships and stakeholder relationships are essential*, as illustrated above, and will help explain successes and shortcomings, be it in ensuring policy relevance, attaining results and doing so efficiently. An evaluation can provide insights into the management of these partnerships.

Stakeholders are key informants who can relate "what happened" during policy implementation and what other events occurred simultaneously. They may, however, also fear that the evaluation is "about them" and their behaviour and be guarded for that reason. Or they may not know about or understand interrelated factors, and thus be able to give only a partial explanation. For this reason, it is essential to consult with a wide cross-section of stakeholders and triangulate findings.

Users of the evaluation findings will gain a better understanding of the reasons why, or why not policy was implemented, how changes occurred and why others did not. These insights, from explanations of the responses to policy prescriptions, can help adjust policy formulation and its implementation.

Conclusion

Policy evaluations pose challenges which need careful management. They are not insurmountable, but they do require considerable consideration in evaluation design and implementation. This paper has highlighted the implications of policy evaluations for the three fundamental evaluation principles: independence, credibility and utility. It suggests ways in which they can be managed through stakeholder analyses and management of relationships and evaluation methods which can address questions of complexity and uncertainty. The three main evaluation questions about the quality of policy, its results, and the factors which explain performance and results, help ensure the independence and utility of policy evaluations, while evaluation methods and processes need to be designed to gain credibility.

References

Asian Development Bank. (2003) *Evaluation of Capacity Development Assistance to Water Utilities*

Asian Development Bank. (2009) *Understanding Complexity*.

Barnes, Marian; Matka, Elizabeth; Sullivan, Helen. (2003) *Evidence, Understanding and Complexity, Evaluation in Non-linear Systems*, Evaluation, Sage Publications.

Boyle, Richard and Lemaire, Donald. (Eds). (1999) *Building Effective Evaluation Capacity, Lessons from Practice*.

Buchanan-Smith, Margie. (2003). *How the Sphere Project Came into Being: A case study of Policy-Making in the Humanitarian Aid Sector and the Relative Influence of Research*, Overseas Development Institute.

Carlsson, Lars. (2000). *Non-Hierarchical Evaluation of Policy, Evaluation,* Sage Publications.

Chelimsky, Eleanor. (2008). *Clash of Cultures*, American Journal of Evaluation, Vol. 29, 2008.

Davies, Philip (2004). *Is Evidence-Based Government Policy Possible?* Jerry Lee Lecture.

deGroff, A and Cargo, M. (2009). *Policy Implementation: Implications for Evaluation,* in Knowledge utilization, diffusion, implementation, transfer and translation: Implications for evaluation, New Directions for Evaluation.

Hanberger, Anders. (2001). *What is the Policy Problem?,* Evaluation, Sage Publications.

Howlett, Michal and Ramesh, M. (1996). *Studying Public Policy, Policy Cycles and Policy Subsystems,* Oxford University Press, 1996

International Fund for Agriculture Development. *Evaluation of the Rural Finance Policy*

Lang, Jochen (2001). *Improving Structural Policy Under Conditions of Hybrid Governance*, Evaluation, Sage Publications.

Meer, Frans-Bauke van der and Edelenbos, Jurian (2006) *Evaluation in Multi-Actor Policy Processes, Accountability, Learning and Co-operation,* Evaluation, Sage Publications, 2006 ODI

Picciotto, Robert. (2005). *The Evaluation of Policy Coherence for Development*, Evaluation, Sage Publications.

Policy Hub. *Better Policy-making,* http://www.nationalschool.gov.uk/policyhub/

Ramalingam, Ben, Jones, Harry, Exploring the science of complexity: Ideas and implications for development and humanitarian efforts, ODI, 2008

Sanderson, Ian (2000). *Evaluation in Complex Policy Systems*, Evaluation, Sage Publications, 2000

Sato, Yuriko. (2005). *A Case of Policy Evaluation Utilizing a Logical Framework*, Evaluation, Sage Publications.

Trosa, Sylvie. (2003). *L'evaluation des Politique Publiques, Institut de l'entreprise.*

United Kingdom Cabinet Office. (2003). *The Magenta Book*, July 2003

United Nations Industrial Development Organization. (1992). *Evaluation of Economic and Technical Cooperation*

United Nations Industrial Development Organization. (1993). *Evaluation of Human Resource Development.*

Van der Knaap, Peter. (2000). *Performance Management and Policy Evaluation in the Netherlands*, Evaluation, Sage Publications.

Vlassopoulou, Chloe. (2005). *Une evaluation constructiviste des politiques publiques?,* Revue Espaces Temps.

World Food Programme. (2008). *Evaluation of WFP's Capacity Development Policy and Operations.* 2

World Food Programme. (2008). *Evaluation of WFP's Gender Policy.*

World Food Programme. (2008). *Evaluation of WFP's HIV and AIDS Interventions.*

THE ROLE OF POLICY ANALYSIS IN OVER-COMING THE IMPLEMENTATION CHALLENGE

by Craig Russon,
Senior Evaluation Officer,
International Labour Organisation (ILO)

Introduction

Within the context of this publication theme, the author has decided to focus on policy analysis systems as a special case of country-led monitoring and evaluation systems. Such systems can help over-come the implementation challenge in two ways. First, these policy analysis systems can conduct analyses **for** policies, which will address the implementation challenge. The results of this type of analysis can serve as inputs to policy proposals. The second way they contribute is by conducting analyses **of** policies, which will attempt to determine the impact of the policy proposals.

Some evaluators think that the best way to measure impact is through randomized controlled trials. Others think that the methods should be determined by the questions which drive the evaluation (Leeuw & Vaessen, 2009). It is this author's view that no greater impact can be attained than through policy-making at the national level. This is because policy determines: what should be done; what resources should be expended; what means should be employed; and, who is responsible (Trochim, 2009). However, the author does not believe that policy-making is amenable to randomized controlled trials.

In this chapter, the definition of policy analysis will be examined. Then, three general approaches to policy analysis will be presented. The implications of these approaches for the way in which policy analysis systems are constructed are also discussed. The process of conducting policy analysis will be presented. Lastly, a case example from the International Labour Organization will be given.

Definition of policy analysis

Policy analysis suffers from the challenge of multiple definitions, in the same way as does the field of evaluation. Patton and Sawicki (1993) have defined policy analysis as a process through which one

identifies and evaluates alternative policies or programmes which are intended to lessen or resolve social, economic or physical problems. Such a definition leads this author to think of policy analysis systems as a special case of monitoring and evaluation systems.

However, there are many other definitions. Dunn (2004), for example, defines policy analysis as a process of multidisciplinary inquiry designed to create, critically assess, and communicate information that is useful for understanding and improving policies. As we shall see below, the potential difficulty of multiple definitions is overcome through the careful operationalization of terms. The approaches and processes used to conduct policy analysis have been very well defined.

Approaches to policy analysis

As with evaluation, there are various approaches to policy analysis. Three general approaches are the analycentric, the policy process and, the meta-policy approach (Wikipedia, 2009).

The analycentric approach focuses on individual problems and their solutions. Its scope is at the micro-scale and its problem interpretation is usually of a technical nature. The primary aim is to identify the most effective and efficient solution in technical and/or economic terms (e.g., the most efficient allocation of resources).

The policy process approach puts its focus onto political processes and involved stakeholders. Its scope is at the meso-scale and its problem interpretation is usually of a political nature. It aims at determining what processes and means are used and tries to explain the role and influence of stakeholders within the policy process. By changing the relative power and influence of certain groups (e.g. enhancing public participation and consultation), solutions to problems may be identified.

The meta-policy approach is a systems and context approach. Its scope is at the macro-scale and its problem interpretation is usually of a structural nature. It aims to explaining the contextual factors of the policy process, i.e., what are the political, economic and socio-cultural factors influencing it. As problems may result because of structural factors (e.g. a certain economic system or political institution), solutions may entail changing the structure itself.

Organizational design and culture of policy systems

As we saw above, each of the three general approaches to policy analysis has unique aims. A policy analysis system that uses one of these approaches should put in place an organizational design[1] that would facilitate the accomplishment of the corresponding aim. Four dimensions of organizational design which should be taken into account are: division of labour; authority; departmentalization; and span of control (Russon & Russon, 2005).

Figure 1: Four dimensions of organizational design

	Analycentric	*Policy Process*	*Meta-Policy*
Division of Labour	Specialized staff trained in economics	Specialized staff trained in political science	Specialized staff with extensive experience in the system
Authority	No special status	Special status vis-a-vis stakeholders	Special status vis-a-vis top policymakers
Departmen-talization	Independence guaranteed by locating system outside the responsible ministry	Access to stakeholders facilitated by locating system inside the responsible ministry	System located above various ministries
Span of Control	Reports up chain of command	Reports to stakeholder groups	Reports to top policymakers

In addition to organizational design, an organizational culture[2] should be created that will enable the policy analysis system to accomplish its aims. Organizational culture has been thought of as the collec-

---

1 & 2   When conducting training on evaluation capacity development, the author often uses computers as a metaphor. The organizational design of an evaluation system is like the computer hardware, and the organizational culture is like the operating system. Unless these two elements are in place and functioning well, no evaluation work can be accomplished. The same capacity development principle would be true for policy analysis systems.

tive personality of the system. Three dimensions of organizational culture which should be taken into account are: artifacts; values; and, assumptions (Russon & Russon, 2005).

Figure 2: Three dimensions of organizational culture

	Analycentric	*Policy Process*	*Meta-Policy*
Artifacts	Economic data and annual budgets	Agendas and minutes of stakeholder meetings	Ministry records and reports
Values	Policy is intended to achieve maximum social gain	Policy is intended to establish and enforce compromise between various, conflicting stakeholder groups in society.	Policies are consistent across and harmonious among ministries
Assumptions	Problems are viewed in technical terms	Problems are resolved through dialog and consensus among stakeholders.	Problems are resolved through structural rather than individual change

Process of conducting policy analysis

A policy analysis system put in place with careful attention to the system's organizational design and organizational culture will, in large measure, determine the processes which can be carried out. It would be beyond the scope of this chapter to attempt to explain the processes which correspond to all three general approaches to policy analysis. Therefore, the author has chosen to focus on the following processes which correspond to the analycentric approach (Saint Germain, 2009):

• Verify, define and detail the problem.

• Establish evaluation criteria.

• Identify alternative policies.

• Evaluate alternative policies.

- Conclusions and recommendations.
- Monitor and evaluate the implemented policy.

Verify, define, and detail the problem

The first process of a policy analysis system, using an analycentric approach, would be to verify, define and detail the problem. The same problem can often be framed using an economic, a sociological or a political science conceptual framework. The way in which a problem is framed determines the manner in which it can be solved. It has been said that policy analysts more often fail because they solve the wrong problem, than by getting the wrong solution to the right problem (Dunn, 2004).

Establish evaluation criteria

Every time a policy problem is identified, some statement of goals is adopted. The goals are what the policy alternative should accomplish. Goals are often broken down into objectives. Good objectives have three characteristics (Raab, Swanson, Wentling, & Clark, 1987):

(1) they contain a statement about expected performance;

(2) they describe the conditions under which the performance will take place; and,

(3) they contain criteria for evaluating performance.

Policy analysis using an analycentric approach most often includes technical or economic criteria. Technical criteria ask whether the technology exists, or is readily available, to implement a proposed alternative. Efficiency and effectiveness are examples of economic criteria (Saint Germain, 2009).

Identify policies options

There are two phases to the identification of alternative policies. The first is the divergent phase in which many options from a large number of diverse sources are considered. Sources include: past experience; brainstorming; research; experiments; writing scenarios; and, concept mapping. The second is a convergent phase in which the options re screened in order to arrive at a manageable number. These are then submitted to careful evaluation during the following process.

107

Evaluate policies options

It becomes necessary to evaluate how each possible option meets the criteria previously established. At the heart of evaluating policy options establishing a counterfactual or in other words to determine what would happen if one policy alternative were selected over another one. According to Dunn (2004), this can be accomplished by extrapolation of current and historical trends into the future (projection) forecast based on explicit theoretical assumptions (prediction), or forecast based on informed or expert judgment (conjecture).

Conclusions and recommendations

The process of drawing conclusions and making recommendations enables policy analysts to produce information about the likelihood that future courses of action will result in consequences that are valuable to some individual, group, or society as a whole. This process involves the transformation of information about expected policy outcomes into information about best policy options. For this reason, the process involves taking into consideration ethical and moral issues (Dunn, 2004).

Monitoring and evaluation of the implemented policy

The consequences of policy alternatives are never fully known in advance, and for this reason, it is essential to monitor the results of implementation. Usually this is accomplished by adopting indicators to track progress toward desired outcomes. The first time that the indicators are measured, baselines are established. The baseline is the policy targets plus the amount of improvement that is expected from the policy action. Targets are divided into time-bound increments, called milestones (Kusek & Rist, 2004).

Monitoring can tell policy analysts if there is progress towards the targets. However, it cannot determine to what aspect of the policy action the progress is attributable. For this, evaluation must be conducted.

Policy analysis example from the International Labour Organisation

In the preceding section, the author considered the definition of policy analysis, general approaches, and the process of conducting policy analysis using the analycentric approach. This section

now presents a current case example of how one UN agency, the International Labour Organisation (ILO), assists its member States to conduct policy analysis.

The primary goal of the ILO is to contribute, with member States, to achieve full and productive employment and decent work for all, including for women and young people. The global financial and economic crisis has made it extremely challenging to achieve the decent work goal.

As economies across the globe have plunged into deep recession, a global employment crisis of unprecedented scale has emerged. Open unemployment is expected to increase by some 40 to 60 million, and the ranks of the working poor are expected to grow by some 122 to 233 million, severely undermining the prospects of attaining the Millennium Development Goals. In response to this crisis, a *Global Jobs Pact* was adopted at the 98th session of the International Labour Conference, in June 2009.

The ILO is committed to supporting its member States and Social Partners in assessing the impact of the crisis, and in determining the policy responses options. Recognizing the need to respond urgently and coherently at country level, the ILO produced a *Country Level Rapid Impact Assessment of Crisis on Employment* guide (ILO, 2009). The guide aims to provide practical advice on how to rapidly build up an adequate country-specific knowledge base with a view to informing decision-making under these difficult circumstances.

The guide suggests a four-step approach to conducting a country level rapid impact assessment of the crisis on employment:

1) an assessment of the pre-crisis situation including existing vulnerabilities;

2) an assessment of the impact of the crisis at the macro level, in key sectors, as well as on the labour market;

3) an assessment of the mitigation capabilities of both households and government; and,

4) an assessment of mitigation measures undertaken so far.

Ugandan Rapid Impact Assessment

In 2009, the Ugandan delegates to the International Labour Conference requested ILO assistance to assess the impact of the global economic crisis on the labour market. An assessment mis-

sion to Uganda was conducted in response to this request (Sender & von Uexkull, 2009). The mission team used the methods outlined in the *Country Level Rapid Impact Assessment of Crisis on Employment* guide (ILO, 2009).

The most important finding was that there is an ongoing steep decline in real wages, caused by food price inflation combined with stagnation in nominal wages. The report argued that the global economic crisis contributed to this situation in two ways: first, rapid outflows of portfolio investment led to a steep depreciation of the Ugandan shilling in late 2008, which continued throughout the first half of 2009. As a result, food exports to neighbouring countries remained very profitable and continued to grow despite increasing domestic food prices (Sender & von Uexkull, 2009).

Second, in the context of the crisis, most employers refused to pay an inflation adjustment to workers even in sectors which have clearly been affected. Both effects were particularly strong for low-wage casual workers who spend a higher share of their income on food and are much less likely to receive wage inflation adjustments, given their low bargaining power (Sender & von Uexkull, 2009).

The crisis also led to a substantial reshuffling of Uganda's export portfolio. Total exports actually increased, but this was driven entirely by informal cross-border trade. Some of Uganda's traditional export crops, most notably coffee but also tobacco and cocoa, suffered a decline in export value caused by lower world market demand. In other cases, such as fish and flowers, declines in exports seem to be attributable to supply side constraints rather than world market conditions. Some export commodities such as tea and maize even benefited from higher market prices (Sender & von Uexkull, 2009).

The boom in informal exports seemed to be mainly driven by ad-hoc trading by individuals in consumer goods following the depreciation in the Ugandan shilling. Whilst it demonstrated the large demand potential for exports to regional markets, and especially into Southern Sudan, it was not associated with increases in output and employment in Uganda. Its sustainability was therefore questionable (Sender & von Uexkull, 2009).

The report also found a substantial decline of earnings in tourism which led to immediate layoffs in the sector. Import values declined during the crisis. That decline was partially a price effect due to

lower world market prices for oil. This also led to a substantial improvement in the terms of trade, following the start of the crisis (Sender & von Uexkull, 2009).

The main message of the report was that workers, and especially low-wage casual workers, have born a disproportionate share of the impact of the economic crisis in Uganda. It highlighted a number of problems which had contributed to the extreme vulnerability of large parts of the workforce. These were: a massive discrepancy between supply and demand for unskilled labour, caused by a rapidly growing workforce and under-investment in key employment-intensive sectors; weak labour market institutions and labour organizations; insufficient education of labour force entrants; and, a social protection system that covers only a small minority of the population (Sender & von Uexkull, 2009).

Policy-makers were encouraged to focus investment efforts on creating decent jobs and to strengthen programmes to reduce population growth and to promote education, especially primary education for girls. In addition, strengthening of labour institutions and organization, and improvement of the social protection system, were seen as key policy priorities. The policy recommendations contained in the report argued in favour of a minimum wage to counter the ongoing decline in real wages (Sender & von Uexkull, 2009).

A pre-condition for the success of efforts to create decent jobs (especially for those rural labour market participants and new entrants who are most at risk), is the availability of improved labour market statistics. The report proposed new survey and research initiatives and highlighted the need to address the quality of national data on the impact of the crisis on vulnerable workers (Sender & von Uexkull, 2009).

The preliminary findings of the mission team (Sender & von Uexkull, 2009) were validated at a roundtable discussion with representatives from the Ministry of Gender, Labour and Social Development as well the Social Partners. The ILO has disseminated this report to a high-level meeting of government, social partners, and other stakeholders in Kampala.

In addition, the mission team (Sender & von Uexkull, 2009) provided detailed suggestions for amendments to the draft of Uganda's National Employment Policy. These proposed amendments mean that the final version can now incorporate an up-to-date and

rigorous analysis of new and ongoing challenges to providing decent work for all Ugandans in the context of the new issues emerging from the global economic crisis.

Conclusion

Policy analysis systems offer a unique opportunity to evaluate the implementation of development initiatives, from policy to results. Approaches to policy analysis are somewhat different from those used in the field of evaluation. However, the principles of capacity development, and the processes used to carry out work, are similar for both fields. An example was provided of the ILO's work to ameliorate the influence of the current financial and economic crisis on employment in Uganda.

References

Dunn, W. (2004). *Public policy analysis: An introduction*. Third edition. Upper Saddle River, NJ: Person Prentice Hall.

Kusek, J. & Rist, R. (2004). *Ten steps to a results-based monitoring and evaluation system*. Washington, D.C.: The World Bank.

ILO. (2009). *Country level rapid impact assessment of crisis on employment*. International Labour Office: Geneva: ILO.

Leeuw, F. & Vaessen, J. (2009). *Impact evaluations and development*: NONIE guidance on impact evaluation. Washington, D.C.: NONIE.

Patton, C. & Sawicki, D. (1993). *Basic Methods of Policy Analysis and Planning*. Englewood

Cliffs, NJ: Prentice Hall. Second edition.

Raab, R., Swanson, B., Wentling, T., & Clark, C. (1987). *A trainer's guide to evaluation*. Rome: Food and Agriculture Organization.

Russon, C. & Russon, K. (2005). *Evaluation Capacity Development Group Toolkit*. Mattawan, MI: ECDG.

Saint-Germain, M. (2009). *Introduction to Policy Analysis.* http://www.csulb.edu/~msaintg/ppa670/670intro.htm

Saint-Germain, M. (2009). *Policy Analysis Criteria*. http://www.csulb.edu/%7Emsaintg/ppa670/p&sch5.htm#p&sch5

Sender, J. & von Uexkull, E. (2009). *A rapid impact assessment of the global economic crisis on Uganda*. International Labour Office: Geneva: ILO.

Trochim, W. (2009). *Evaluation policy and evaluation practice.* In W. Trochim, M. Mark, & L. Cooksy (Eds.), Evaluation policy and evaluation practice. New Directions for Evaluation, 123, 13-32.

Wikipedia. (2009). *Policy Analysis*. http://en.wikipedia.org/wiki/Policy_analysis

Part 2
Developing
country monitoring
and evaluation capacities

Strengthening individual monitoring and evaluation capacities

A PERSPECTIVE FROM THE UNITED NATIONS ON NATIONAL OWNERSHIP AND CAPACITY IN EVALUATION

by Saraswathi Menon,
Director, Evaluation Office, UNDP and Chair,
United Nations Evaluation Group

Evaluation as a discipline and as a means of dialogue within countries and between governments in the course of development cooperation has advanced considerably in the past few decades. This paper argues that there has been greater investment in the development of analytical methods and approaches from the perspective of development cooperation and its requirements than in recognizing fully the potential of evaluation as part of governance within countries. This recognition must be rooted on a fuller understanding of national ownership and capacity in evaluation than is normally the case. This recognition will transform the understanding of evaluation itself.

Much has been achieved in capacity in evaluation in countries throughout the world. As a professional discipline standing on the shoulders of several social sciences from whom it gains methodologies and analytical approaches, evaluation may be said in a certain sense to have come of age. Evaluation's visibility as a profession may be measured crudely by the number of universities offering courses in evaluation, dedicated evaluation units in many governments and organizations, regional and country level associations of evaluation professionals, conferences on evaluation where new methodologies are discussed, consulting companies that specialize in evaluation, and not least the plethora of evaluation reports published annually that seek to make a constructive contribution to change.

There is richness and synergy in this diversity that is remarkable and difficult to capture. But when looked at through the lens of national ownership and national capacity these formations are only means to an end. And although as means they are critical and important, losing focus on the ends that we want to achieve, can warp the effectiveness of our engagement with strengthening these means, and dilute the achievement of the ultimate end.

What is the ultimate end of evaluation? Simply put I would say that it is to improve public action to contribute to people's wellbeing. Public action is seen in its broadest sense as the agency of people combined with the actions of the state and its partners. Learning and

accountability, which are so often seen as the underpinning of evaluation, are solely the channels by which the process and results of evaluation temper and galvanize public action.

Despite the huge advances in evaluation, for too long we have been reductionist and ahistorical in our understanding of national ownership and national evaluation capacity. I will illustrate this point of view primarily from the field of development evaluation.

The way in which the United Nations (UN) deals with development, unlike security, exemplifies the essence of what the UN represents – each country, whether rich or poor, large or small, has an equal stake. They have the same voice in determining how normative frameworks are established, such as the Universal Declaration of Human Rights, the Convention on All Forms of Discrimination Against Women (CEDAW) or the global environment conventions such as those on biodiversity and climate change; global public goods are managed (health, space, postal communications); priorities in supporting developing countries are established, with UN General Assembly resolution after resolution underlining the importance of economic growth, sustainable development and poverty reduction; and progress is monitored by the global community, through the Millennium Development Goals (MDGs) and internationally standardized statistics on national income accounts, population, literacy and so on.

The right of equal voice is coupled with the obligation to further these commitments. While nation states subscribe to these commitments and common ways of working, actual achievement is mediated by national policies and practice. The space for making decisions that affect people's lives, even in a globalised world, rests largely with the nation state. And it is at the national level that people can determine politics. Civil society, which has since Rio, played an important role in shaping many of the global commitments, plays a key role within societies through political parties, the media, unions, business associations and other groupings to influence public policies and action. An understanding of national ownership and capacity in development must take state, citizen and civil society responsibility and action into account.

And yet in the UN, as in the bilateral aid community, development is too often coterminous with development cooperation. National ownership is seen from the outside looking in. Reference is made to building national ownership as though it is to be created from outside or has to be restored having been hijacked earlier. In strengthening capacity a huge shift has been made within the UN from talking about capacity building as though it is to be fostered from outside to speaking about capacity development that recognizes intrinsic

existing capacity that gains from contact with partners. But this has not yet translated fully into evaluation capacity development.

In what way is our approach to national ownership and national evaluation capacity reductionist? Just as development is closely entwined with development cooperation, development evaluation has largely become the handmaiden of development cooperation interventions. The extremely useful criteria developed by Development Assistance Committee (DAC) of the bilateral donors in the context of evaluating projects – relevance, effectiveness, efficiency, sustainability and impact – have today become core criteria for all types of evaluation, and are embedded in professional evaluation capacity development efforts. These criteria are a natural fit for the discrete universe of projects through which development cooperation, and sometimes narrow spheres of government action, is delivered. Similarly, the path breaking work on impact evaluation begins, and too often ends, from the perspective of an intervention and its results.

But the evaluation questions around these criteria have to be fundamentally rethought if they are to yield meaningful information when looking at more complex phenomenon including outcomes, strategies, policies and public goods. They yield only partial information and not necessarily the most crucial for influencing public action that span the state, civil society and the citizen. They do not capture the complexities and interaction between policies. If evaluation is to reorient public action, evaluation criteria and evaluation questions must answer what is important for the citizen and policy maker and not only for the funder and manager of the initiative, whether national or international.

This shift in thinking can be seen in the real world. The South African Public Service Commission, which has the responsibility for evaluation of the public sector in South Africa, has an ethical, value-laden, development-oriented approach to evaluation. Responsiveness and appropriateness are important concepts that are integrated into their approach because the touchstone is the perspective of the citizen and not the intervention. Chinese evaluation professionals have developed and applied evaluation criteria that translate directly from their national aspirations. They include equity and innovation as criteria, thus placing a premium on equality as a development outcome and on a 'learning through experimentation" or "crossing the river by feeling the stones" approach to development change.

Both these experiences underline the importance of relating capacity to national vision and national aspirations. Investment in technical rigour is an intrinsic and necessary part of evaluation capacity. But it is not sufficient. If national ownership is to imbue evaluation practice, national vision and aspirations must be captured through evaluation

criteria, questions, methods and approach. National evaluation capacity needs to be able to generate appropriate criteria and methods and apply them to answer meaningful questions. From this perspective, development evaluation needs to be subsumed within national ownership and leadership rather than the latter being understood as something to be fostered through the practice of evaluation practitioners working in development cooperation.

In what way is our approach to national ownership and national evaluation capacity ahistorical? All evaluation systems and approaches are born within a national social and political context. The creation of the Government Accountability Office (GAO), in the United States or of the Swedish Agency for Development Evaluation (SADEV), more recently in Sweden answered specific political requirements for transparency and accountability in the public sector. These institutions may use standard evaluation techniques and methods and produce high quality reports but their continued legitimacy and relevance rests largely on their essential raison d'etre which lies within the arena of governance.

Similarly, when looking at the emergence of evaluation in newly independent India, the establishment of the Programme Evaluation Organisation in 1952 just one year after the first Five Year Plan, reveals how integrally bound up evaluation was with the national project of development. India's Plan represented the role of the state in catalyzing a mixed economy through investment in infrastructure, higher education and other priority areas. Evaluation was seen as a key instrument in assessing whether this public role was delivering results. Evaluation, therefore, derived its legitimacy and relevance from a specific policy vision. Many innovations in evaluation methodology that were introduced in those early years happened because of the close links between evaluation and the national statistical system, which was also seen as an essential element in understanding the trajectory of development in an independent country. The South African emphasis on responsiveness to citizens similarly accompanied the national purpose of overcoming deep historically created inequality and divisions through freedom and equal rights for all. History and governance determine the parameters and relevance of evaluation and provide their authorizing mandate.

Evaluation capacity development as it has occurred in many countries for the last decades has been influenced by the political context of aid. The emphasis has been on setting out results frameworks acceptable not only to national authorities but also to their international partners through defined frameworks, such as the Poverty Reduction Strategy Papers (PRSP), which can be jointly tracked and measured and become the basis for funding and prioritization deci-

sions on both sides. It is then an easy step to creating and embedding evaluation frameworks and capacity within these partnership frameworks. The impetus and legitimacy for evaluation in such cases emerge from cooperation rather than a national vision or mandate. And it, therefore, has weak links to public accountability of public action that concerns the citizens of the country.

But when looking at actual experience such situations are quickly shaped and transformed by the evolution of national politics and the national context. The interesting experience of the evaluation commissioned by the Government of Uganda of their Poverty Eradication Action Plan, the national PRSP, demonstrates the way in which tools and instruments introduced by international partners can be adapted and serve a national purpose. Many countries that were first introduced to monitoring and evaluation systems in their aid projects have introduced their own national systems to address specific requirement of monitoring actions and evaluating public policy.

These and other experiences reveal the importance of the historical authorizing or legitimizing context for evaluation. Evaluation is an intrinsic part of governance. And the fullest expression and space for democratic governance is within the nation state where citizens have a voice, can organize, can act and demand action. If evaluation is to be effective and nationally owned it must be part of this democratic space within countries. This requires adaptability while retaining rigour and credibility.

Evaluation when embedded in governance becomes a public good. The quality of governance is, therefore, a determining factor of the effectiveness of evaluation in influencing public action. But even where there is room for improvement, external assistance needs to be sensitive and respectful of national processes.

The toolbox should be inclusive. In many countries there are social audits and citizens' report cards, all of which are means for assessing results in terms that concerned citizens can grasp. The shadow reporting process for CEDAW and other UN conventions provides a unique way of bringing the official government report on progress in a country together with an alternative viewpoint to the attention of the international commissions dealing with these conventions. These different perspectives strengthen the evidence base for the experts to form a collective assessment and to be able to better advise all countries on how to improve performance against commitments that they have made. Evaluation cannot stay apart from these approaches.

There are a number of common principles of evaluation that are enshrined in documents such as the UN Evaluation Group's (UNEG)

Norms and Standards for Evaluation, DAC's Standards for Evaluation, the Evaluation Cooperation Group's Good Practice Standards and so on. The prisms of history, legitimacy and ownership need to be used to understand and apply them. Let us take independence as a case in point. Independence of evaluation has been examined like the proverbial elephant in a variety of ways – structural, behavioural, reporting lines, resource decisions, staffing decisions, link to decision making, access to information, conduct of evaluation, content of evaluation, disclosure and so on.

The fundamental question, however, is why do we need independence? If evaluation is to influence public action then both the authorities and the people, whose agency is an intrinsic part of public action, must see the evaluation process and result as impartial. Any perceived conflict of interest that can bias the information base or tamper with the evaluative judgment undermines the credibility and hence the usefulness of an evaluation. But it is the authorizing and legitimizing basis of evaluation in any country or organization that determines the nature and expression of this independence.

In the UNEG Norms for Evaluation, one of the means by which independence is protected is by requiring that the head of a UN evaluation unit should report to the head of the organization or the governing body. The notion of public accountability, interestingly enough, is not pursued. The need to build credibility while engaging with those who will effect transformation – decision-makers and staff of the organization is not addressed. The solution to independence is structural not political or governance related. In my view, each principle such as independence has to be explored to the fullest, within an organizational or more importantly, within a national context. National ownership and the capacity to influence public action can only be promoted if these principles are not applied in a technocratic way but rather embedded in governance.

I have argued that the approach to national ownership and national capacity in development evaluation suffers from a broader malaise. The tendency is to stay at one remove and look in as though national ownership has to be fostered and capacity developed. This is indeed what is happening.

In the United Nations Development Programme (UNDP), the Evaluation Office conducts independent end-of-programme cycle evaluations of global, regional and country programmes for accountability purposes. When we assess UNDP's contribution to development results in countries we seek to engage the partner country government in determining what will be evaluated, assuring the quality of the evaluation, and in disseminating and using the evaluation.

This may be done through a jointly managed or conducted evaluation, through the establishment of a broad national reference group, or by commissioning a national evaluation authority to conduct the evaluation. But at the end of the day even if the programmes are part of national programmes we are assessing UNDP's contribution and addressing UNDP's accountability.

Interestingly enough evaluation has not even caught up with new approaches in programming. For many decades, the programmes that UNDP funded and evaluated in developing countries were executed by specialized UN agencies, such as FAO and ILO. Over time the organization shifted to national execution where a national entity executed the programme replacing a specialized UN technical agency. Today there is an attempt to move to national implementation to more deeply integrate UNDP funded programmes into national programmes. These are dynamic attempts to address ownership in a more meaningful way.

In evaluation, however, we are still seeking national ownership and endorsement of an instrument that is required solely for the accountability of the multilateral organization, without fully addressing national and public accountability. At best we talk of alignment to national priorities and processes, and then continue an evaluation practice that addresses primarily the international partner's needs. We need to develop instruments that will cover organizational needs within national and public accountability.

The joint Paris Declaration Evaluation was a huge breakthrough in joint evaluation. The approach used was to develop a common framework terms of reference globally which was then adapted by participating developing countries for their country evaluation. The coherence was across the several country evaluations as the purpose was the preparation of a synthesis report. Only in rare cases was the country evaluation an integral part of the national evaluation project or linked to national governance processes.

This is not national ownership or national capacity in evaluation. National ownership would require that development partners join in the evaluation project of the country; that this project is intrinsically linked to national vision and national accountability to citizens; that it be an integral part of strong democratic governance processes; that the evaluation method and process is complex and sensitive enough to capture a range of influences on public policy; that the results of the process are meaningful and can effect change in public action.

What then needs to change? The UN can make a huge contribution to shifting the paradigm on national ownership and capacity in evaluation.

The first prong is normative. UNEG has Norms, Standards, Ethical Guidelines and a Code of Conduct for Evaluators. These are strong documents reflecting UN values and sentiments. When translated into policy documents for individual agencies, these norms are further developed. For instance, in the UNDP evaluation policy human rights and human development are basic principles. Agencies with the mandate and knowledge have moved the frontiers further, with UNIFEM and the Office of the High Commissioner of Human Rights working on methodologies to integrate gender equality and human rights into evaluation. If evaluation is to serve national vision, aspirations and public action, the importance of linking evaluation to the growing body of rights and obligations that have been framed within the UN cannot be overemphasized.

The second prong is to understand the role of history and political context of evaluation in countries and providing a platform for learning and experience sharing across countries that takes these contexts into account. Countries and organisations benefit from the knowledge of what has worked elsewhere in order to go beyond technocratic solutions to build evaluation into governance. Rather than co-opting national evaluation systems to enhance development cooperation evaluation needs, the approach would be to learn from strong national systems, such as GAO in the United States and the Public Service Commission in South Africa. National ownership and capacity in evaluation require going beyond government led evaluation, although it is an important part of public trust and transparency. Experience sharing across parliaments, academia, civil society, the media, evaluation associations and governments would benefit society as a whole. And the creation of an effective national evaluation system must include and be responsive to the full range of actors.

The third prong is to change the way we work. The many agencies of the UN support national systems for monitoring, data gathering, analysis, and evaluation in their mandated areas. We are blinkered sectorally and do not often bring the full strength of the UN to serve public systems. Our approaches to evaluation are blinkered organizationally and do not encourage our joining the national evaluation project. Integrating the best of our normative, technical and governance knowledge into national evaluation will ultimately be the most effective way to promote a people-centred evaluation that catalyzes public action for human development.

GENDER EQUALITY AND HUMAN RIGHTS RESPONSIVE EVALUATION: RETHINKING APPROACHES

by Belen Sanz Luque,
Head of Evaluation, UNIFEM, and Co-Chair,
UNEG Taskforce on Gender Equality and Human Rights

Introduction

2010 constitutes the tenth anniversary of the *Millennium Declaration*, a historical agreement among the United Nations Members States to end poverty and promote sustainable development with eight specific goals, indicators and targets to be met by 2015.[1]

The recognition of the right to development as a human right on its own indicates that "every human person and all peoples are entitled to participate in, contribute to, and enjoy economic, social, cultural and political development, in which all human rights and fundamental freedoms can be fully realized" [2]. Therefore, the achievement of the *Millennium Development Goals* (MDGs) cannot be fulfilled without fully ensuring the active and equal access and participation of all individuals, men and women, in development processes and the fair and equal distribution of its benefits. The centrality of this to the achievement of the MDGs is reflected in the fact that gender equality and women's empowerment constitutes one of the eight goals, and is shown to be a critical factor for the achievement of the other seven goals.

---

[1]   http://www.un.org/millenniumgoals/

[2]   Declaration on the Right to Development. A/RES/41/128, Art. 1.

While acknowledging the progress made thus far in a number of the indicators established for the MDGs related to gender equality and the empowerment of women (such as girls access to education, women's incorporation into the labour force, or women's share of national parliament seats), the targets set for each goal for 2005 have not been achieved and there are important challenges still to be overcome 10 years after the agreement was made.[3]

The existing data indicate that more effort is needed from governments and society as a whole. It also suggests that, while certain areas related to women's access to basic services have seen some progress (although still far from the targets), indicators related to shifts in power relations, such as equal decision making and political participation, are still lagging behind. This further calls for the need to enhance the indicators and data gathering to measure progress on gender equality and women's empowerment.[4] Most importantly for the topic of this book, it implies that we need to enhance our analytical frameworks for the evaluation of policies and programmes aiming at social justice and gender equality; rethinking how we measure and judge contribution to social change in order to overcome the challenge of implementation at the country level.

This paper is a humble attempt to identify and stress the role of evaluation, in the current development context, to achieve gender equality and women's empowerment. It calls for the need to rethink and enhance the approaches and methods we use so that evaluations contribute to social transformation processes.

---

3 Regional disparities are noticeable in access to education, and even within countries, nation-wide averages overshadows significant disparities between rural and urban areas. In addition, women's share of waged non-agricultural employment has increased only slightly in the last decade. Women account for almost 40 per cent of the total employment in this sector, but in many regions like South Asia, North Africa and West Asia, employment opportunities for women remain extremely limited; while more women have secured paid jobs outside agriculture, they have generally failed to access decent work. Close to two-thirds of all employed women are working in vulnerable jobs, either as contributing family workers or as own-account workers (MDGs Report, United Nations, 2009). Finally, there is a slow improvement in women's share of national parliamentary seats, averaging globally 18 per cent as of January 2009, representing a rise from 11.6% in 1995, but still far from the 30% target envisioned in the Beijing Platform for Action (PFA) (Progress of the World's Women. New York, UNIFEM, 2008/09).

4 There is growing recognition about the limitations of the targets and indicators of MDG3. Proposals to take on an "MDG3 plus" perspective – that take into account other critical factors that perpetuate gender inequality, such as violence against women, early marriage, unequal access to land and property rights, and the unequal share of unpaid care work that women carry out – have been proposed in analytical and position papers produced by many bilateral and multilateral organizations.

I will highlight the key policy/normative frameworks within the United Nations which establish the relevance of human rights and gender equality, and I will explore emerging evaluation paradigms which mirror some of those frameworks within the evaluation practice. Furthermore, I will conclude with exposing the key parameters of what constitutes the emerging approach of gender equality and human rights responsive evaluation, for which guidance is currently being developed by a number of United Nations (UN) organizations through the UN Evaluation Group.[5]

Gender equality, human rights and links to evaluation in the development field

The Universal Declaration of Human Rights, adopted in 1948, recognizes universally agreed human rights as the rights inherent to all human beings regardless of their residence, sex, nationality, religion or any other status, without discrimination. It has been followed by two legally binding Covenants agreed by member states of the United Nations. Of these Covenants recognizing human rights as the foundation of freedom, justice and peace, one is on economic, social and cultural rights, the other is on civil and political rights. There are also additional, separate conventions on the rights of specific groups subject to discrimination.

Emanating from the Universal Declaration and the two Covenants, the United Nations has placed specific attention on vulnerable groups who historically have not been fully afforded their human rights based on factors of sex, race, disability, age, political standing, or status in the workforce. As a result, there are now a number of Conventions and Declarations ratified and in force which specifically target those groups most subject to discrimination and exclusion. This has resulted in a comprehensive normative and legal framework for human rights that includes nine core conventions, clearly identifying the content of those rights which have achieved global endorsement.

---

5   The contents of this paper draw substantially from the work done by the United Nations Evaluation Group Task Force on Gender Equality and Human Rights as well as the experience of the United Nations Development Fund for Women in evaluation. The opinions and judgments in this paper are those of the author.

Table 1. Core human rights treaties and declarations.

Human Right Treaty / Declaration	*Date Adopted / entry into force*	*State Parties*
Universal Declaration of Human Rights (UDHR)	1948	161
International Covenant on Civil and Political Rights (ICCPR)	1966/1976	161
International Covenant on Economic, Social and Cultural Rights (ICESCR)	1966/1976	157
Convention on the Elimination of All Forms of Racial Discrimination (ICERD)	1965/1969	173
Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW)	1979/1981	90
Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (CAT)	1984/1987	145
Convention on the Rights of the Child (CRC)	1989/1990	193
International Convention on the Protection of the Rights of All Migrant Workers and Members of their Families (ICRMW)	1990/2003	37
Convention on the Rights of Persons with Disabilities	2006/2008	20
International Convention for the Protection of All Persons from Enforced Disappearance	2006/not yet	4
Declaration on the Right to Development	1986	
Declaration on Human Rights Defenders	1998	
Declaration on the Rights of Indigenous	2007	

Source: OHCHR, United Nations

Gender equality and women's human rights commitments are, as with the MDGs, included in all of the core conventions, but are more fully encapsulated in the *UN Convention on the Elimination of All Forms of Discrimination against Women* (CEDAW) and the *Beijing Platform for Action* (PFA). The binding UN Security Council resolutions 1325, 1820, 1888 and 1889 also recognize the central role of women in peace building processes. Together with the commitments in the *Millennium Declaration* and the *2005 World Summit*, the *Accra Action Agenda* and the *Doha Declaration* in 2008 as well as a series of International Labour Organization Conventions

on working women's rights – they offer a road map for advancing gender equality in countries worldwide.

With regard to the analytical frameworks which have mirrored the progress made by these important declarations, it is well documented[6] that the role of women in development was first explicitly acknowledged as a critique of the 1960s and 1970s economic growth model, with its focus on growth rather than on re-distribution. The 'women in development' approach emerged as a distinct development approach that responded to the failure of development policies to account for women's needs and priorities and the role of women in development.[7] In addition, there was a shifting focus in anti-poverty strategies during the 1980s and 1990s which introduced the variables of inequity and exclusion into the measurement of poverty. Another shift in the development approaches of the 1990s and 2000s focused on human development, participatory development, ownership and empowerment[8]. Likewise, the focus on women evolved from welfare to equality, anti-poverty and efficiency to empowerment.

The gender analytical framework that has emerged from these different paradigms since the 1980s also acknowledged and assessed, in their first generation, the difference in access and control over resources on the basis of gender relations between women and men, and later in their second generation, recognized the importance of identifying the practical needs and especially the strategic interests of women in the development processes.[9]

Importantly for evaluation, the gender analytical framework challenged the assumption that communities and households are harmonious social units. Research showed that men and women have differential access to household resources and diverse needs and interests over them. It also identified the implications of women's multiple social and economic roles. As accounted by Bamberger & Podems (2002), this framework serves to identify and measure changes and differences in women's and men's time use; to determine access and

---

6   See Bamberger & Podems (2002) account and extensive bibliography cited (pp. 94-96). Key sources are Boserup (1970); Moser (1993); Kabeer (1992); Mohanty (1997); Molyenex (1985).

7   Bamberger; Podems (2002), "Feminist Evaluation Explorations and Experiences" in New Directions for Evaluation. Jossey-Bass: San Francisco.

8   Ibid., p. 83.

9   Moser (1993), p.39.

control over resources before and after an intervention; and, to assess the effects of differential factors (demographic, socio-cultural or political) on who has access to and control over development resources.

While being very relevant for gender impact evaluation, feminist researchers' critiques highlighted the fact that the gender analysis framework helps to account for, but not to change, women's subordinated role through development interventions[10]. The evolution of the gender analysis framework after the 1990s was prompted by Moser (1993) and others by identifying not only the practical needs, but also the strategic interests of women with regard to their subordinated position to men, which is intrinsically related to gender divisions in labour, power and control which lead to unequal gender relations. The 'Gender and development' (GAD) approach emerged from this thinking with gender relations analyzed in terms of power differentials between women and men. A number of development agencies have used and further refined this GAD approach.[11]

Feminist research has further challenged the 'neutrality' of the analytical and research process. It focuses on power, privilege and gendered social forces, attends to the political realities of participant's lives, focuses on the research context and the research itself, and uses "centralized gender and power dynamics as a lens for understanding oppressed groups" (Beardsley; Hughes Miller, 2002). It has also placed great focus on the interactions between gender, ethnicity and class.

For the purpose of this paper it is important to acknowledge that gender analysis and feminist approaches have made a relevant contribution to research and evaluation by making the issue of power a key analytical dimension, placing the focus on giving voice to those who are disempowered or voiceless, or who are discriminated against. Feminist research "has at heart a commitment to change for women" and focuses on how the research is conducted and with what purpose. Feminist research places attention to gender as a variable of analysis: it analyzes gender in context; it pays attention to the researcher's subjectivity; the negotiation of gender in the research process; and, it explicitly commits to the empowerment of women as part of its agenda.

---

10   Ibid. See Moser, 1993; Mohanty, 1997; Reid, 1995.

11   See among others the framework developed by the World Bank in the context of the HIPC initiative. www.worldbank.org/poverty/strategies/chapters/gender/gen0427.pdf

Human Rights Based Approach (HRBA) is a model strongly grounded on the recognition of power imbalances as the basis of discrimination and serves as the UN's standard approach to human rights in programming, as stated in the *UN Common Understanding on the HRBA*[12]. A rights-based approach to development promotes justice, equality and freedom and tackles the power issues that lie at the root of poverty and exploitation. To achieve this, a rights-based approach makes use of the standards, principles and methods of human rights, social activism and development. It involves the use of a conceptual framework to understand the causes of non-fulfillment of human rights by analyzing inequalities, discriminatory practices and unjust power relations which are often at the heart of development problems. It is based on international human rights standards and principles, and develops the capacities of rights-holders to claim their rights, and duty-bearers to fulfill their obligations as outlined in the core conventions discussed earlier.

The human rights analysis looks at the inter-related human rights standards and groups suffering from a denial of rights; analyses the root or underlying causes for the situation that the intervention proposed to address (e.g. what is the cause of exclusion, discrimination or inequality); identifies the individual and/or institutional duty-bearers and their corresponding obligations under international human rights law; and, analyses the capacity needed (skills, resources and authority, etc.) by both rights-holders to claim rights, and duty-bearers to meet obligations that would lead to actual changes in the realization of rights.

The UN Common Understanding on HRBA, agreed by the UN system in 2003, defined three main areas of understanding and six core principles of human rights. The three areas can be identified as different levels of results guiding UN programming as in table 2.

---

12   The HRBA is the result of the Secretary-General's set of reforms for the UN since 1997 and 2002 . The UN system agreed on a 'UN Common Understanding of HRBA' (2003) and later on Action II, a programme to strengthen the UN's support to promote and protect human rights, so that they can support the efforts of Member States in establishing and strengthening national human rights promotion and protection systems consistent with international human rights norms and principles, including through the promotion of a human rights-based approach (HRBA). The Second Interagency Workshop on Implementing a Human Rights-based Approach in the Context of UN Reform http://www.undg.org/index.cfm?P=16 contains the basis of the HRBA.

Table 2: Three areas of understanding and their level of result.

Area of understanding	*Level of result*
1. All programmes of development cooperation, policies and technical assistance should further the realization of human rights as laid down in the UDHR and other international human rights instruments	Long term impact of development interventions
2. Human rights standards contained in, and principles derived from, the Universal Declaration of Human Rights and other international human rights instruments guide all development cooperation and programming in all sectors and in all phases of the programming process	Outcome intended by development interventions
3. Development cooperation contributes to the development of the capacities of duty-bearers to meet their obligations and/or of rights-holders to claim their rights	Outcome / Output to be achieved and delivered by development interventions

Source: adapted from UN Common Understanding

In addition, from the framework of six core 'principles of the HRBA,'[13] to guide programming, the three of particularly relevance to evaluation, are: equality and non-discrimination; participation and inclusion; and accountability and rule of law[14].

Finally, it is possible to highlight 'key components' of the rights-based approach with important implications for evaluation[15], these are:

- **Ensuring the development process is normatively based on international human rights standards and principles.** Being based on international human rights standards and principles, a

---

13   The remaining three principals include: universality and inalienability, indivisibility and inter-dependence and inter-relatedness.

14   The remaining three are universality and inalienability; indivisibility; and, inter-dependence and inter-relatedness.

15   Typology presented in the UN Common Learning Package on the Human Rights Based Approach http://www.undg.org/index.cfm?P=531 and used for the "draft UNEG guidance on integrating Human Rights and Gender Equality in evaluation".

rights-based approach focuses explicitly on discrimination and marginalization in the development process, and uncovers the underlying and root causes of major development challenges and unfulfilled rights. A rights-based approach should confront patterns of inequality and discrimination, and formulate responses which change the structural causes of exclusion and marginalization, and the denial of human rights. For evaluation, this means using the observations and recommendations of treaty bodies as references, and assessing the quality of human rights analysis conducted to design the intervention and the extent to which an intervention has challenged/changed inequalities and led to improved realization of human rights.

- **Determining the claims of rights-holders and the obligations of duty-bearers.** The HRBA is a capacity-building and social inclusion approach that establishes a relationship between duty-bearers and rights-holders. For evaluation, this means: assessing the extent to which the capacities of rights-holders and duty-bearers have been adequately identified and supported by an intervention; determining how far increased capacity has led to duty-bearers meeting their obligations, and to greater exercising of rights by rights-holders; and, how the relationship between rights-holders and duty-bearers has changed at the level of actual results.

- **Focusing on groups subject to discrimination.** There is a strong emphasis in the HRBA on identifying and supporting the capacity of women and any group subject to discrimination. A human rights analysis therefore needs to determine which groups are being discriminated against, and the means by which this discrimination can be overcome. Evaluations will need to pay close attention to whether these groups are receiving the support they need.

- **Giving equal weight to the outcomes and processes.** An HRBA gives the same importance to process as it does to results, with the focus on the process to achieve those rights being as important as achieving the enjoyment of these rights. Participation is a human right enshrined in many conventions and declarations and thus should be promoted and analyzed in evaluation.

Having briefly explored the two analytical frameworks which emerge from gender and human rights, the next section presents how evaluation practice has evolved and how it reflects these two analytical frameworks.

Evaluation context and shifting paradigms

How has evaluation thinking evolved and mirrored the frameworks delineated above? Evaluation in the development context has gone through relevant changes which are reflected both in the theories behind them and in the methods and practice deployed by experts and development organizations.

First, in the context of international development, there has been an evolution of trends in evaluation thinking and practice from measurement through accountability to organizational learning[16]. The first generation, from the 1950s to the 1970s placed a particular focus on appraisal rather than on evaluation. The emphasis was on the use of logic models translated later on into logical frameworks, and on the measurement of outputs, where evaluation was an external activity at the end of the project cycle, as a tool for measurement and comparison. In the second generation during the 1980s, an expansion of evaluation occurred and different development organizations started to institutionalize the practice as a tool for transparency and accountability. There was an interest in measuring longer term results, although investment in evaluation within agencies still remained limited. During the third generation, in the 1990s, evaluation increasingly became a function within international organizations and started to be considered as a strategic tool for knowledge generation and organizational learning. In this way, evaluation became 'utilization-focused' and several emerging approaches appeared in the field which paid attention to fostering empowerment, participation and inclusiveness. In the 2000s, with the signing of the *Paris Declaration on Aid Effectiveness* (2005), a number of 'rules of operation' are agreed between the donor and partner country community affecting also the development evaluation practice. A greater emphasis is put on "mutual" accountability; ownership; alignment; harmonization; and, management for results, therefore opening a space for evaluation to focus from projects to policies; to foster democratic approaches to evaluation; promote joint evaluations; and, support and enhance the use of country owned monitoring and evaluation systems.

---

16   See the synthesis made by Segone (2003); other relevant sources cited Cracknell (1988); Rebien (1997).

Different authors have explored the roots of evaluation theories and trends.[17] Patton (2008) distinguishes two core paradigms which have encapsulated much of the debate between research paradigms: quantitative/experimental paradigm (with a focus on precision and on things that can be counted, including assigning numerical values to experience); and, qualitative/naturalistic paradigm (with an interest in capturing how participants experience a programme). These approaches competing in the earlier 1970s and more reconciled in the present, still generate debate in the evaluation community about which one constitutes the "gold standard" for evaluation (Donaldson et.al., 2009), i.e. which generates more rigour and valid, credible data.

Furthermore, and useful for this article, another categorization is made along five sets of criteria to define evaluation quality and different evaluation perspectives or approaches. These criteria are: 'traditional scientific research' oriented, 'social construction and constructivist'; 'artistic and evocative'; 'pragmatic and utilitarian', and 'critical change'.[18] The *Fourth Generation of Evaluation* (Guba and Lincoln, 1989), belonging to 'social construction and constructivist', highlights the inevitability of subjectivity in the evaluation process and commits to capture and respect multiple perspectives, favoring qualitative enquiry and the use of mixed methods. The author situates feminist evaluation both within this paradigm and most directly within the 'critical change' criteria, where evaluators see themselves as change agents with the explicit intention to contribute to social justice.

The increased use and consolidation of these approaches, as in a number of publications: *Fourth Generation of Evaluation* (Guba & Lincoln); *Developmental Evaluation* (Patton); *Participatory Evaluation* (Whitmore, Cousins); *Empowerment Evaluation* (Fetterman); *Transformative Evaluation* (Mertens); demonstrates that there is a need to respond to emerging social challenges with evaluation models which can capture different levels of complexity.

---

17   See for example Alkin M. (2004). Evaluation Roots: Tracing Theorists' Views and Influences. Thousand Oaks. Ca.; or Stufflebeam and Shinkfield (2007).

18   See Patton (2002) pp. 101-103.

In this context the work of Mertens (2009) is a clear example of this evolution, providing a new evaluation paradigm that emerges from the human rights agenda and social justice, and positions the evaluator as an agent of change committed to social transformation.[19] The transformative paradigm "provides a framework for addressing inequality and injustice in society using culturally competent, mixed methods strategies. The recognition that realities are constructed and shaped by social, political, cultural, economic, and racial/ethnic values indicates that power and privilege are important determinants of which reality will be privileged in a research context".

Table 3. Characteristics of the transformative paradigm

- Emphasizes human rights and social justice.
- Places central importance on the lives and experiences of communities which are pushed to society's margins.
- Analyses asymmetric power relationships.
- Links results of social enquiry to action.
- Advocates culturally competent relations between the evaluator and community members.
- Uses transformative theory to develop the programme theory and the inquiry approach.
- Employs culturally appropriate mixed methods tied to social action.
- Applies feminist theory, critical race theory, post-colonial and indigenous theories.

Source: Mertens 2009

Finally, the focus of current debates and topics of evaluation conferences confirm the premise of shifting paradigms. This is demonstrated for example by the discussion on what counts as credible evidence in evaluation practice (Donaldson ibid.).

This new evaluation paradigm is also emerging in the field of gender equality and women's empowerment, as exemplified among others by a recent paper, by the Association for Women's Rights in Development,[20] on challenges in monitoring and evaluation to capture change in women's realities. After reviewing a number of existing models, the paper identifies: the limitations of current monitoring and evaluation frameworks to help us understand how change

---

19   Mertens (2009). Transformative Research and Evaluation. Guilford Press: New York.

20   Srilatha B., Pittman A. (2009).

happens or how gender relations have been altered; the linearity of the most commonly used monitoring and evaluation frameworks which do not acknowledge the complexity of social change; the lack of monitoring and evaluation frameworks to allow for tracking negative change, reversals and backlash on women's rights and the need to capture less tangible but equally important results of programmes targeting women's rights, such as the effects of research, capacity building, advocacy, etc.

Lastly, recent evaluation association conferences are showing a concern about the role and interaction of evaluation with broader social discussions by placing greater attention on issues related to democracy, governance and the role of values and context in evaluation.[21]

Gender and human rights, as core elements of social justice, are at the heart of this discussion. Both in domestic policies and in the context of international development, the failure to account for changes in power relations and how social programmes affect distinct groups in the population (especially those who are marginalized and/or excluded), implies a failure in the understanding of how social change occurs and how structural unbalanced power structures sustain injustice and unequal opportunities, and so limit access to, and enjoyment of, development outcomes.

Among other networks, the United Nations Evaluation Group (UNEG) has identified the need to address these dimensions in evaluation and is currently working on the development of practical evaluation guidance on integrating human rights and gender equality in evaluation. The work has focused on mapping existing evaluation models in the development context and on developing guidance for attending to the gaps in order to address these issues. It is expected that the practical guidance and tools from this group will soon be developed and made available. The following section draws substantially on the progress from this work in the recent years, and reflects on the challenges found along the way.

---

21   Some examples from last conferences illustrate this. African Evaluation Association (AFREA), in 2007: 'Evaluating Development, Developing Evaluation: A Pathway to Africa's future, 'Evaluation Matters, Africa matters', or in 2004: 'Joining Forces for Democracy, Governance and Development'; from the American Evaluation Association in 2009: 'Evaluation in Context'; or the Canadian Evaluation Society in 2009 'Value in Evaluation'.

Key principles and parameters for gender equality and human rights responsive evaluation

In this paper, evaluation is defined according to the *United Nations Evaluation Group Norms*.[22] Key aspects of this definition highlight the systematic nature of evaluation, which assesses an activity, project; programme; strategy; policy; topic; theme; sector; operational area; institutional performance; etc,[23] and its focus on both expected and achieved accomplishments; examining the results chain; processes; contextual factors; and, causality, in order to understand achievements or the lack thereof. It further emphasizes its intention to judge and determine the relevance, impact, effectiveness, efficiency and sustainability of the interventions, and the contributions of the organizations of the UN system.

The consideration of gender and human rights in evaluation practice implies the integration of these dimensions in all steps of the evaluation processes. It involves, on the one hand, the analysis of the effects of interventions in enhancing, or negatively affecting, gender equality and the empowerment of women (and all individuals subject to discrimination) to enjoy their rights. On the other hand, it involves the incorporation of these dimensions in the actual evaluation processes, from evaluation design to methods, data gathering and interpretation, and the use of evaluation results. In order to do this (as explained in section one), it draws importantly on the gender analysis framework and feminist and transformative research paradigms, which have key implications for gender and human rights responsive evaluation[24] including:

- **Gender inequalities**: explicit attention to gender inequalities and to the extent possible, analyze the intersectionality between gender inequalities with other forms of social injustice (racism, social class, disabilities, etc.).

- **The subject is treated with respect and as equal**, and there is power sharing between the evaluator and those in the programme being evaluated.

- **Ethical considerations** throughout the evaluation process are given a high priority.

---

22   UNEG (2005), http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=21.

23   For the purpose of this paper, we will name this 'units' of analysis "interventions".

24   Patton M.; Bamberger M.; in Seigart and Brisolara ed. (2002).

- **The political nature of evaluation** is acknowledged and thus power relations are explicitly addressed throughout the evaluation.

- **Experience and 'women's way of knowing'**, intuition and emotion is valued and reflected in the evaluation process: in the questions asked; in the methods deployed; in the use of findings; and the evaluator is self reflective about her/his own role and values.

- **The evaluation design and methods**: design includes, but goes beyond simple disaggregation of data to ensure critical gender analysis (such as time use, control of resources, participation in decision making, gender-based violence) and use appropriate methods: mixed methods (quantitative and qualitative); participatory processes which support consciousness raising; capacity-building; reflexivity; gather and value the knowledge and experience of all participants, especially disadvantaged women; and, data gathering techniques that are gender sensitive and promote inclusion and empowerment.

- **Use of the knowledge** that emerges from the evaluation process, to bring about change and social justice benefitting and empowering women in their particular contexts.

In addition, we can identify the following dimensions for evaluation emerging from the HRBA:

- Puts people first, especially disadvantaged groups, and promotes human centered development.

- Identifies rights-holders, such as women, and duty-bearers, and thereby enhances accountability for the realization of rights.

- Has a normative stance on the side of the most discriminated and excluded, thus compelling a focus on women and any group subject to discrimination.

- Underlines a holistic understanding of rights.

- Promotes empowerment of participants, especially the most discriminated groups, through participatory and inclusive processes.

- Directs its efforts at the roots of structural injustices and inequalities rather than effects.

- Focuses on women and other groups subject to discrimination.

- Focuses on the process (how?) and outcomes (what?) to enjoy human rights.

- Focuses on people's knowledge, attitudes and practices.

- Demands tools and approaches appropriate for collecting the required data from the identified sample groups and in suitable formats for them.

- Ownership and accessibility of evaluation results to all groups.

Building both on the gender analytical framework and the human rights-based approach, the principles for evaluation are:

- **Equality and non-discrimination**: All individuals are equal as human beings and by virtue of the inherent dignity of each person, and all human beings are entitled to their human rights without discrimination of any kind.

- **Participation and inclusion**: Every person and all peoples are entitled to active, free and meaningful participation in, contribution to, and enjoyment of civil, economic, social, cultural and political development in which human rights and fundamental freedoms can be realized. In particular, participation and inclusiveness in the process of decision-making by all affected, particularly women and the poor is critical to development outcomes.

- **Empowerment**, that gives women and men the power, capacities, capabilities and access to resources to enable them to change their lives.

- **Accountability and rule of law**: State parties and other duty-bearers are answerable for the observance of human rights, and rights-holders – women and men – are enabled to claim their rights and ensure that States fulfill their obligations.

The key evaluation parameters considered under this framework are:

- Recognize power relations;

- Identify the structural causes of inequality and discrimination;

- Determine the impacts of programmes on impoverished people and groups facing discrimination;

- Gather data disaggregated by sex, ethnic groups and other categories, and undertake gender analysis of such data;

- Involve participatory and reflective processes;

- Acknowledge women's human rights and gender equality and determine the claims of rights- holders and the obligations of duty- bearers; and

- Aim for the progressive realization of human rights and gender equality.

An evaluation that is responsive to gender equality and human rights would ask the following questions and assess if the intervention:

- Was based on a gender analysis and human rights based analysis; identified core rights to be furthered according to international conventions; considering how structures contribute to inequalities for women, especially those belonging to groups subject to discrimination; and, to challenge these structures by building the capacities of women to claim their rights and duty-bearers to fulfill them.

- Considered which human rights issues were to be furthered; appropriately defining who are the duty-bearers responsible to respect, protect and fulfill obligations and the rights-holders to claim and enjoy their rights and their respective capacity gaps; encouraged participation and inclusiveness, particularly of women and groups subject to discrimination.

- Collected and analyzed sex, age, ethnic, and regional disaggregated data.

- Contributed to progress/results on women's empowerment, gender equality and women's human rights and other social transformations.

In order to assess these dimensions, the evaluation design should deploy approaches which are cognizant of power relations and which promote mixed methods.

Having highlighted the key parameters of gender equality and human rights responsive evaluation and how this has been informed by both the gender analytical framework and the HRBA, the following considerations are required for the different evaluation phases in order to put these parameters into practice in evaluation processes.

Preparation and evaluation design: gender equality and human rights issues are considered in the design and planning of the evaluation. Thinking ahead on information needs means that timely information will be available and there will be real involvement of users:

- Consider and estimate the resource needs to incorporate gender equality and human rights perspectives in evaluations (time, money, expertise);

- Consider gender and human rights in any evaluability assessment, i.e., adequacy of the intervention's design to identify needs and interests of women and groups subject to discrimination; existence and availability of data, and cost/time to gather data disaggregated by sex and other variables; conduciveness of the context and use of the evaluation;

- Stakeholder analysis that identifies and assesses women/other groups subject to discrimination; duty-bearers and rights-holders, and their level of involvement in the evaluation process;

- Choosing the right criteria, approach and methods – inclusive and participatory approaches, mixed methods; culturally responsive and sensitive; cognizant of power relations.

Conduct of the evaluation: calls for the full engagement and participation of rights-holders and duty-bearers throughout the evaluation process:

- Design and conduct of the evaluation must be transparent, inclusive and accountable.

- Both rights-holders and duty-bearers should be engaged to the extent possible (e.g. as informants, as part of reference groups, etc.).

- Participatory processes should be used.

- The power relationships within the evaluation process should consider:

  – how the process does not reinforce discrimination or discriminate/marginalize against women and other groups subject to discrimination;

  – how the process can contribute to the empowerment of women and other groups subject to discrimination, promote gender equality and realization of Human Rights.

Methods and data collection: deploys mixed methods and ensures ethical standards:

- Methods should be responsive and sensitive to the situation of women and of any group subject to discrimination in their particular context:

  – Data gathering techniques should take into account the times, spaces and circumstances which will allow the meaningful participation of women and any group subject to discrimination;

  – Women and any group subject to discrimination to feel comfortable in the cultural context to speak their views and concerns;

  – Compliance with ethical guidelines specific to the population being informants.

- Mixed methods should be encouraged and used. Along with quantitative methods deployed to gather the broad spectrum of the affected population to be targeted by interventions, and qualitative techniques to be used in order to capture the voices and perceptions of women and other groups subject to discrimination.

- Indicators should be sensitive to differences among groups and include measures of programme effects on groups facing discrimination.

Follow-up and use: ensures findings can be used by users, finds ways to disseminate results

- Evaluation findings to be shared using dissemination strategies which are accessible and barrier-free to women and any group subject to discrimination, ensuring users are identified within righs-holders and duty-bearers.

- Targeting specific user's groups who may benefit from knowledge emerging from evaluations: civil society women's organizations and networks; human rights advocates; knowledge networks and academia.

- User-friendly language.

- Stakeholder workshops which include women and other groups subject to discrimination.

- Management response should be issued to ensure follow-up on key issues to advance gender equality and human rights.

In summary, evaluations which are responsive to gender equality and human rights dimensions are evaluations which: recognize power relations; identify the structural causes of inequality and discrimination; determine the impacts of programmes on impoverished people and groups facing discrimination; that involve participatory, inclusive and reflective processes; acknowledge women's human rights and gender equality; and, determine the claims of rights-holders and obligations of duty-bearers. Ultimately, they are evaluations which aim for the progressive realization of human rights and gender equality.

These evaluations recognize the gendered nature of development and the importance of gender equality to economic and social development. They recognize that poor and marginalized groups of people are gendered, and women and men are differentiated by race, ethnicity, age, disability, class and caste; that power relations exist within the home/family; and, that these relationships intersect with power relationships in the society, policy, and the economy.

Challenges ahead and concluding remarks

This paper has briefly outlined some key analytical frameworks which shape the formulation of an evaluation approach that is responsive to both gender dimensions and human right. It has highlighted the parameters which should be included in evaluations under this paradigm that aims for social transformation through the evaluation process – where evaluation practice itself becomes an agent of change.

As this constitutes an evolving approach that results from the work of people and organizations within the UN and beyond in the evaluation and development communities, there are a number of challenges that need to be considered.

Firstly, gender equality and human rights approaches are not the same. They both focus on fundamental rights, recognize power imbalances at the root of inequality and discrimination, but gender approaches are derived from the women's movement against discrimination, whilst human rights are derived from the sphere of political science and law. Whilst the pursuit of gender equality is integral to human rights, it is not subsidiary to it. Gender equality is also a factor and precondition for sustainable development. Therefore, evaluation that is responsive to both frameworks, as presented in this paper, can draw from their common principles and use their analytical frameworks according to the nature of the intervention evaluated.

An additional challenge has to do with evaluation in the current international development context, which has to respond competing demands, from accountability and focus on measurability, to learning and better understanding of social transformation. The approach presented here should respond to both demands by deploying evaluation processes and methods which can answer existing questions to advance development, whilst ensuring methodological rigour.

Lastly, the evaluation of gender and human rights issues require adequate resources and expertise. These aspects are not always guaranteed in development programmes struggling for funding but are very important if we want to advance this agenda.

Relevant implications from this gender and human rights paradigm need to be considered, both in terms of what we evaluate in public policies and, most importantly for the evaluation field, of how we evaluate. The paradigm responds to the current development context and forces the evaluator to be self-reflective and understand his/her own self as a factor in influencing power relations and to acknowledge the responsibility that this entails when conducting any evaluation.

Finally, the gender and human rights approach broadens the scope of the evaluation field to deploy creative and suitable approaches and methods to give voice to those populations most discriminated against, and marginalized, whilst ensuring that these approaches and methods are based on sound advancements in the evaluation and research fields. The gender and human rights responsive approach is worth investing in if we are to contribute to facing the challenges that mark our current era, such as advancing towards the *Millennium Development Goals* which the international community agreed to achieve by 2015.

References

Bamberger, M., Podems D. (2002). *Feminist Evaluation Explorations and Experiences.* In New Directions for Evaluation. San Francisco: Jossey-Bass. 2002.

Bamberger, M. (2001). *Engendering Development Through Gender Equality in Rights, Resources and Voice.* Washington: World Bank: 2001.

Boserup, E. (1970). *Women's role in Economic Development.* London.

Donaldson, S., Christie A., Mark M. (2009). *What counts as Credible Evidence in Applied Research and Evaluation Practice?* California: Sage.

Fetterman, D., Wandersman, A. (2005). *Empowerment Evaluation Principles in Practice.* New York: Guilford Press.

Mertens, D. (2009). *Transformative Research and Evaluation.* New York: Guilford Press.

Moser, C. (1993). *Gender Planning and Development.* London: Routledge.

Office of the High Commissioner for Human Rights. United Nations. (2008). *The United nations Human Rights System. How To Make It Work For You.* Geneva.

Patton, M. (2008). *Utilization Focused Evaluation.* 4th Edition. California: Sage.

Segone, M. (2004). *New Trends in Development Evaluation*. UNICEF CEE/CIS Office & IPEN.

Srilatha, Batliwala, Pittman, A. (2009). *Capturing change in women's realities: The Challenges of Monitoring and Evaluating Our Work*. AWID. http://www.awid.org/ Issues-and-Analysis/Library/The-pitfalls-of-monitoring-and-evaluation-Do-current-frameworks-really-serve-us

Stufflebeam, D., Shinkfield A. (2007). *Evaluation Theory, Models & Applications*. San Francisco: Jossey Bass.

Theis, J. (2004). *Promoting Rights Based Approaches: Experiences and Ideas from Asia and Pacific.* http://seap.savethechildren.se/upload/scs/SEAP/publication/ publication%20pdf/child%20rights%20programmeming/Theis_Promoting_Rights_ Based_Approaches.pdf

United Nations Evaluation Group. (2009). *Draft Guidance on Integrating Human Rights and Gender Equality in Evaluations in the UN System.* Draft under revision.

Waldorf, Lee, *CEDAW and the Human Rights Based Approach to Programmeming*. New York: UNIFEM.

DEVINFO AND WEB 2.0. STRENGTHENING DUTY-BEARERS DEMAND FOR HIGH-QUALITY DATA

by Nicolas Pron,
DevInfo Global Administrator, UNICEF Headquarters

by Kris Oswalt,
Executive Director, DevInfo Support Group

by Douglas Hsu,
Project Officer, DevInfo Support Group

The Millennium Development Goals (MDGs), adopted by all United Nations Member States in 2000, have become a universal development framework for developing and transition countries. There is a growing consensus that these goals cannot be achieved without the full participation of all stakeholders, particularly the poor, in the decisions and policies which affect their lives. Along with this, there is also a growing consensus that a human rights-based approach to development programming can effectively empower all people to make the best development-related decisions for their own lives.

In tackling international development issues through the prism of a human rights-based approach, a few key principles emerge as fundamental. These include the principles of universality; equality; non-discrimination; participation; and, inclusion. From a human rights perspective, it is imperative that these principles be applied equally across the globe so that every man, woman and child can participate in, contribute to and enjoy, full civil, economic, social, cultural and political development – the precursor to realizing the promises of the Millennium Declaration.

A human rights perspective, therefore, requires increased attention on the part of duty bearers (governments and other entities responsible for guaranteeing rights), to be constantly assessing, monitoring, evaluating and analyzing, in order to identify and address disparities. Identifying disparities allows duty-bearers to enact policies which enable vulnerable sub-populations to realize their potential and overcome discrimination and other forms of exclusion, which deprive them of development opportunities. A human rights perspective also requires that relevant and appropriate indicators be selected, which can identify the unintended impact of policies and practices, as well as expose overlooked subpopulations.

As such, duty-bearers at the international, national and sub-national levels, are more than ever before, seeking access to relevant, reliable and timely data – the indispensable input for evidence-based decision-making. Ready access to relevant, reliable and timely data to duty-bearers can help ensure that all men, women and children under their influence, are sharing equally the benefits of human development.

DevInfo 6.0. A database system designed to provide relevant, reliable and timely data at the national and sub-national levels

DevInfo is a database system which harnesses the power of advanced information technology, to compile and disseminate data on human development. In particular, the system has been endorsed by the UN Development Group to assist countries in monitoring achievement of the MDGs. DevInfo provides methods to organize, store and display data in a uniform way, to facilitate data sharing at the country level across government departments, UN agencies and development partners. DevInfo has simple and user-friendly features which produce tables, graphs and maps for inclusion in reports, presentations and advocacy materials. The software supports both standard MDG indicators and user-defined indicators. DevInfo is compliant with international statistical standards to support open access and widespread data exchange. DevInfo is distributed royalty-free to all Member States and UN agencies for deployment on both desktops and the web. The user interface of the system, and the contents of the databases supported by the system, include country-specific branding and packaging options which have been designed for broad ownership by national authorities.

The vision that DevInfo supports is: a day when Member States, as key duty-bearers, use common database standards for tracking national human development indicators; that the database will contain high-quality data with adequate coverage and depth; such that this will sustain good governance around the agenda of ensuring that the development promises of the Millennium Declaration are realized by every child, woman and man.

DevInfo is being used as an advocacy platform to engage a broad spectrum of stakeholders in policy choices for human development. Both duty-bearers and rights-holders are using DevInfo as a tool to help them carry their development agendas forward. Key duty-

bearers – Member States and UN agencies around the world – are using DevInfo to help support the reform of development planning policies. The system is enabling the UN to work together as "One UN," and to effectively deliver as one UN system, based on a common database that leads to a common understanding of how to move forward together, with less duplication of efforts and wasteful delays. Key rights-holders are also able to access the same data and identify disparities and vulnerable sub-populations, giving them the needed data to press for necessary interventions.

The latest release of the database system, DevInfo 6.0, was launched in June 2009 and contains numerous Web 2.0 enhancements to further facilitate access to, and dissemination of, data relevant to evidence-based decision making. These include an online DevInfo user forum; a DevInfo wiki; an online searchable gallery; RSS news feeds; an online scatter plot application; and, enhanced mapping features. The last section of this chapter discusses these various Web 2.0 enhancements in greater detail, and illustrates their power in giving rights-holders, and other stakeholders, access to timely and relevant data.

DevInfo usage in supporting country-led monitoring and evaluation systems is increasing worldwide

As of September 2009, a total of 278 DevInfo adaptations had been launched – an increase of 12% over the 2008 total of 249 DevInfo adaptations. These adaptations are spread across a record 132 countries and regions, up from 123 at the end of 2008.

Country reports indicate that the use of DevInfo at the country and sub-national levels, in supporting national monitoring and evaluation systems, is not only increasing around the world, but it is also becoming institutionalized as its usefulness and relevance is being recognized. Given below are several recent examples of DevInfo use and institutionalization across different continents.

Africa: Rwanda

The National Institute of Statistics (NISR) of Rwanda is forging ahead with the development and implementation of both the DevInfo and Integrated Management Information System (IMIS) databases, to facilitate the monitoring of Rwanda's Economic Development and Poverty Reduction Strategy (EDPRS) and

the MDGs. The databases have been developed in close collaboration with UNICEF and UNFPA, with support from UNDOCO.

The NISR has adopted DevInfo to monitor the EDPRS, MDGs and other indicators linked to human development. The database system currently consists of 96 EDPRS indicators and 146 MDG indicators, and is being used to store aggregated data, with pre-calculated indicator values, for dissemination. Additionally, the One UN initiative in Rwanda is presently developing a DevInfo database for its own use.

Africa: Benin

The latest DevInfo adaptation in Benin – BenInfo 5.0 represents seven years of dedicated progress and successful UN/government collaboration, beginning with the first version of BenInfo launched in 2002. It is being implemented by the National Bureau of Statistics with UN support.

The current version is available in French and contains a comprehensive database with data for 632 indicators; 539 time periods; 283 sources; and, four levels of geography. The database is updated regularly using information collected by the National Institute of Statistics and Economic Analysis (INSAE). As national

Figure 1. BenInfo, national adaptation in Benin, is used to monitor MDGs as well as national development strategies

priority indicators from the Poverty Reduction Strategy Paper have also been integrated into the database, BenInfo helps with reporting on the MDGs as well as on the national priorities.

The BenInfo initiative has been proactive in organizing sufficient training to ensure sustained use and implementation of the database. Five User module training sessions and two Data Administration module training sessions have been organized to date, for various database users and stakeholders. Several Training of Trainers (ToT) sessions have also been held, allowing national partners to organize training sessions by themselves.

With various stakeholders able to use the same database containing indicators at the national, departmental and sometimes communal level, BenInfo has helped reduce duplication of data collection and enhance data reliability provided by various sources.

Benin has also forged ahead in developing a DevInfo adaptation for two of its 12 departments (sub-national areas), Borgou and Alibori. The adaptation is known as BorgAlInfo and is being implemented by the Departmental Direction of Prospective Development (DDPD Borgou/Alibori), along with the National Bureau of Statistics, with the support of UNFPA and UNICEF.

The current version is available in French and consists of a comprehensive database containing data for 2,520 indicators; 34 time periods; 2 sources and 3 levels of geography (department, communes and health districts). The database helps to record data from the Annual Tableau de Bord Social du Borgou/Alibori established with the support of all technical and financial partners involved in Borgou/Alibori (particularly UNFPA), as well as government institutions at the department level. The BorgAlInfo database is being updated by integrating information collected for the Tableau de Bord Social. The indicators of the departmental priorities included in the Tableau de Bord Social have also been integrated in the database.

Africa: Lesotho

With the partnership between the UNCT and the government of Lesotho, Lesotho DevInfo is gaining ground as the Millennium Development Goals (MDGs) progress tracking tool. This DevInfo adaptation is also used for monitoring performance with regard to other national development frameworks, including the national Vision 2020, the

Figure 2. Lesotho DevInfo, national adaptation in Lesotho, is used to monitor MDGs as well as national development strategies

Poverty Reduction Strategy (PRS, and the UNDAF 2008-2012. The database is thus forming the backbone of the National Monitoring and Evaluation System (NMES) and, ultimately, the performance management system of the Government.

Lesotho DevInfo received by key duty-bearers with enthusiasm and appreciation. It has proved to be a very user-friendly tool, and efficient use of it will facilitate production of timely progress reports with enhanced graphical presentations. The database played an important role, for instance, in the production of the latest MDG Status Report. The report predicts that three of the eight goals

(those related to universal primary education, gender equality and a global partnership for development), will probably be met by 2015. Those related to HIV/AIDS and poverty and food insecurity, however, will still require concerted efforts to scale up action. Based on the data it was also determined that the Government and partners need to urgently scale up multi-sector interventions, for achieving a reduction in child mortality and, an improvement in maternal health.

The Lesotho DevInfo database contains more than 300 indicators classified by goal, sector, theme, and source. NMES is one of the key themes in the database and contains 50 indicators categorized under four clusters, which are human development, good governance, cross cutting issues and sustainable and shared economic growth.

A team of two training associates is dedicated to conducting countrywide sensitization and training workshops on the Lesotho DevInfo database. Training workshops target strategic stakeholders in the districts and district planning teams. The plan is to bring on board people who are key actors in the implementation of all government development initiatives. To date, the monitoring and evaluation system team has conducted training workshops in three out of the country's ten districts.

Latin America: Colombia

The 2006-2010 Development Plan of the national government of Colombia (Communitarian State: Development for All), incorporated a strategy called "Facts and Rights: Municipalities and Departments for Childhood and Adolescence." Among other activities of this strategy, UNICEF is providing technical assistance to the public administration of various departments (La Guajira, Boyacá, Antioquia, Tolima and Cauca), so they can create, and operate, their own

Figure 3. ColombiaInfo, national adaptation in Colombia, is being adapted and customized to monitor child and adolescent-related indicators at the sub-national (departmental) level

Children's Observatories. These observatories are interdisciplinary groups which monitor the situation of children and adolescents and, support the formulation, implementation and evaluation of evidence-based public policy for children.

The departmental administrations have been encouraged to support the information activities of these observatories using personalized versions of DevInfo. Although this process is just beginning, DevInfo has emerged as a useful information tool for policy-makers and communities, because it provides them with a simple structure to comprise various indicators and to analyze relevant data regarding the actual situation of children and adolescents.

Asia: Indonesia

The December 2004 Indian Ocean Tsunami off the coast of Sumatra, Indonesia was one of the deadliest natural disasters in history, killing over 225,000 people in 11 countries. The worst-affected part of Indonesia was Aceh.

Officials working on recovery operations had been struggling to get reliable and accurate data to guide their policy-making. The data challenges they faced were typical of those encountered in the aftermath of large-scale natural disasters: lack of data; varying methods of statistical collection from year to year; absence of key planning indicators in the available data; and, lack of a centralized database for data from multiple sources.

That is why Aceh Info 2.0 has come to play a major role in assisting with recovery policy-making. A new user-friendly recovery database, Aceh Info 2.0, was released as a free local adaptation of DevInfo in both desktop and web-based versions. Aceh Info contains data on 280 recovery indicators from over 20 organizations, drawing from regional,

Figure 4. AcehInfo, a sub-national adaptation in Indonesia, is used to monitor recovery indicators for the Aceh region

national and international sources. The product is supported by the Information and Analysis Section (IAS) of the UN Office of the Recovery Coordinator for Aceh and Nias (UNORC).

Aceh Info derives its indicators not only from the Millennium Development Goals (MDGs), but also from the Tsunami Recovery Impact Assessment and Monitoring System (TRIAMS), which was endorsed by the Global Consortium for Tsunami-Affected Countries in June and September 2005. The data, wherever possible, has been disaggregated geographically by administrative area, sex,

age group, and urban-rural classification. With Aceh Info, users can easily access all relevant data for recovery decision-making and repackage it into visually-attractive tables, graphs, and maps – with just a few simple mouse clicks. The fact that the product is freely available in both Indonesian and English makes it useful not only for regional and national policy-makers but also for international ones.

Given the popularity and usefulness of the product in the region, Aceh Info 2.0 has generated a great deal of interest outside Aceh. Leaders from various national Indonesian government agencies met with UNORC officials to discuss the possibility of expanding the product to cover other regions of Indonesia.

DevInfo usage in addressing regional development challenges is increasing worldwide

Reports indicate that the use of DevInfo at the regional level, in supporting regional monitoring and evaluation systems, is also increasing around the world. The following are two recent examples of DevInfo implementation at the regional level.

Asia and Pacific region

UNICEF and UNESCO went online with a one-stop Education for All (EFA) database for Asia and the Pacific that enables governments to better identify which of their population's children are not in school, or not learning, and why. The database brings greater precision in analysing where countries in the region have made significant achievements in meeting the Education for All (EFA) goals. More importantly, the database will help

Figure 5. EFAInfo, a regional adaptation, is used to monitor education indicators for the Asia-Pacific region

identify the un-reached groups in education, in terms of location, ethnicity and socio-economic status, among others.

As a region, Asia and Pacific has made major gains towards all six EFA goals. However, across the region and within countries, significant disparities persist and threaten their achievement. Asia and Pacific EFAInfo provides user-friendly data on those population groups most in need of access to quality education. Government ministries, academic institutions and civil society organizations will have access to regional and national indicators disaggregated by age and sex, ethnicity and socio-economic status. Where data is available, it also probes provincial and district-level settings, providing evidence for effective response to achieving Education for All.

The regional resource goes beyond standard education data to look at issues such as health and poverty, with data drilling below the surface to sub-national data. In this way it allows for an expanded analysis of education, using not only administrative data collected regularly by ministries of education but also more insightful information generated from household surveys. This provides users with the ability to 'triangulate' analysis and look at education from multiple perspectives.

The online EFAInfo database comprises all of the internationally-accepted standardized indicators, as used for the monitoring of EFA and the Millennium Development Goals, together with country-specific indicators, used for the monitoring of national development plans and the implementation of country-specific interventions.

African Union

As a major thrust of the African Union's (AU) efforts to improve educational standards across the continent, the AU adopted a Plan of Action for the Second Decade of Education for Africa (2006-2015). One of the primary goals in this Plan calls for the development of functional national Educational Management Information Systems (EMIS) in member states, which are key for systematic planning, monitoring and evaluation of education-related targets.

The Association for the Development of Education in Africa (ADEA), in conjunction with National Education Statistical Information Systems (NESIS), has been providing technical support to the AU in this priority area of EMIS. One of the main areas of focus in this regard has been to help countries to produce statistical reports to monitor their implementation of the AU's Plan of Action. To this end, ADEA-NESIS decided to create four new adaptations of DevInfo in order to help AU countries, and key AU regions, to effectively

capture, monitor and report their educational sector data. These DevInfo adaptations are as follows:

- NESIS Eye – for the African continent

- SADC Eye – for the 14 member countries of the Southern African Development Community

- ECCAS Eye – for the 10 member countries of the Economic Community of Central African States

- African Union Outlook – which contains all the indicators mentioned in the Plan of Action

New Web 2.0 features in DevInfo facilitate sharing and dissemination of high-quality data

The concept of Web 2.0 refers to "web development and web design that facilitates interactive information sharing, inter-operability, user-centred design and collaboration on the World Wide Web" ("Web 2.0" entry, Wikipedia). The latest release of the DevInfo database system incorporates numerous Web 2.0 features within the package – both in the DevInfo 6.0 application itself as well as in the suite of new DevInfo websites accompanying this release. This section discusses in detail how these various features contribute to the sharing and dissemination of high-quality data, for rights-holders and duty-bearers alike.

di Forum

One of the main characteristics of Web 2.0 is the element of interactive information sharing. To this end, di Forum (www.devinfo.org/di-forum) was launched as a place to share ideas and post updates and suggestions related to DevInfo usage and implementation. The site contains forum categories on a variety of topics, such as: monitoring MDGs; dissemination of census data; rapid assessment; and, follow-up in emergencies, etc.
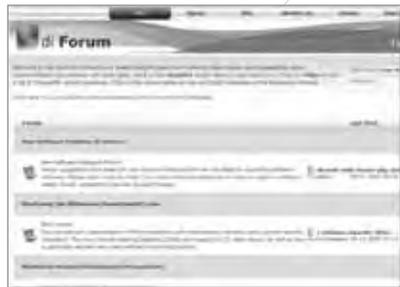
Figure 6. di Forum is a place to share ideas and post updates and suggestions related to DevInfo usage and implementation

The DevInfo initiative is also encouraging the use of di Forum at DevInfo training events worldwide, to facilitate information sharing and networking among training participants. For example, at a recent DevInfo web deployment capacity building event in Rwanda, the National Institute of Statistics created a DevInfo web forum. They intend to update the thread with their experiences of deploying and managing Devinfo on their intranet, to facilitate learning and sharing with other countries and institutions.

di Wiki

Another key feature of Web 2.0 is the concept of authoring – the ability to create constantly updating content. The DevInfo initiative launched di Wiki (www. devinfo.org/di-wiki) as part of its suite of Web 2.0 websites. di Wiki is a free DevInfo encyclopaedia that registered users can edit. Interested contributors can become registered users by simply requesting the DevInfo Support Group for a login ID.

Figure 7. di Wiki is a free DevInfo encyclopedia open for user editing

di Wiki contains wiki-style entries on numerous DevInfo-related topics. The following four categories of information are particularly relevant for decision-makers and are regularly updated:

- *Country fact sheets* – These contain detailed information on the status of DevInfo implementation in countries across the globe and are updated regularly by DevInfo country focal points.

- *Training event summaries* – These contain summaries of recently-completed DevInfo training events held around the world and are updated regularly by DevInfo trainers.

- *DevInfo in Action reports* – These are short articles on specific instances of DevInfo usage and implementation making an impact.

- *Country focal points* – di Wiki maintains the most current list of DevInfo country and regional focal points worldwide.

Online search for gallery objects

Another feature of Web 2.0 is enhanced information retrieval through keyword searches. The new DevInfo website (www. devinfo.org) features a new online searchable gallery of tables, graphs and maps. Users can enter keywords, such as "maternal mortality", to search for gallery objects and then click on a thumbnail image to enlarge

Figure 8. di Gallery features a new online searchable gallery of tables, graphs and maps

the object. These objects can then be copied and pasted into other documents. The DevInfo initiative plans to vastly expand this online gallery by soliciting contributions from DevInfo users worldwide, thereby putting a vast library of ready-made human development presentation objects on a searchable online platform. This will give rights-holders and duty-bearers equal access to presentation objects that can identify disparities, or overlooked subgroups.

RSS feeds

RSS technology is another dimension of Web 2.0 which is based on rapid user notification of content changes. From the DevInfo 6.0 User application, users can launch di Sidebar, which features RSS news feeds with the latest news on DevInfo implementation world-wide as well as weekly Fact of the Week updates. Additionally,

Figure 9. RSS news feeds are available from the DevInfo website

an RSS feed option has been implemented on the main DevInfo website (www.devinfo.org), allowing users to select from a variety of feed options to keep abreast of the latest DevInfo updates.

di Analyzer

In keeping with the Web 2.0 philosophy of interactive informa-tion sharing and user-centered design, the DevInfo initiative launched di Analyzer as part of the global DevInfo 6.0 release. di Analyzer is a web-based application powered by DevInfo database

technology that allows users to display and view data in scatter plots in order to analyze changes and trends over time. Users can view data from DevInfo databases uploaded on the web and optimized for di Analyzer. The optimization process prepares the data for display in scatter plot format.

Figure 10 di Analyzer displaying 1970 human development data for selected African countries

To demonstrate the powerful data analysis and dissemination features of di Analyzer, consider the following example taken from the MDG Info 2008 database. In Figure 10, the scatter plot displays 1970 data for three selected African countries: Mali (blue), Ghana (red), and Egypt (green). The indicator selected for the X-axis represents literacy rate of 15-24 year olds. The indicator selected for the Y-axis represents GDP per capita. The indicator selected for the size of the circle symbol is population below $1 per day.

Figure 11 di Analyzer displaying 1990 human development data for selected African countries

Clicking on instructs the application to begin a time series scatter plot progression, allowing users to view the inter-relationship between multiple indicators and multiple areas over time.
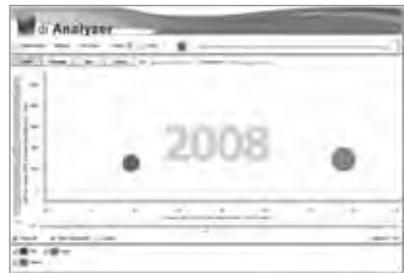
Figure 12 di Analyzer displaying 2008 human development data for selected African countries

Figure 11 displays data for the same indicators for 1990. By comparing these two scatter plots, it can be seen that little progress was achieved in the 20 year period from 1970 to 1990 in all three indicators.

Figure 12 displays data for the same indicators for 2008. By comparing this scatter plot with the 1990 one, it can be seen that while

Mali made some gains in literacy rate, Ghana made significant gains in literacy rate and Egypt made significant gains in both literacy rate and GDP per capita. The shrinking size of the blue circle shows that Mali made significant gains in reducing the percentage of people living on $1 per day. The scatter plots also show the positive correlation between rising GDP per capita and literacy rates.

Enhanced mapping features

In keeping with the Web 2.0 philosophy of interactive information sharing and user-centered design, the new release of DevInfo 6.0 contains enhanced mapping features which enable users to share and disseminate relevant data for informed decision making.

One example of this is the power of enhanced mapping features to display disparities. Consider the following two maps displaying the total literacy rate in India in 1991 and 2001.
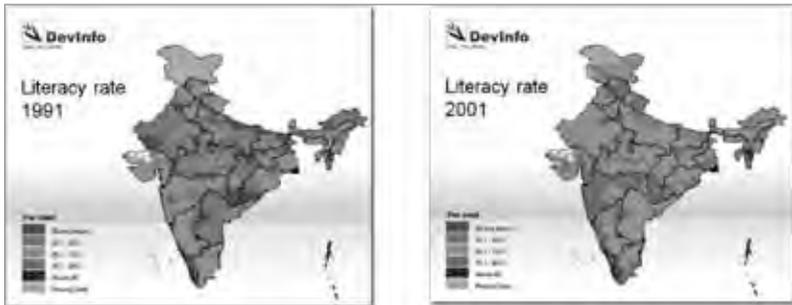
Figure 13 Total literacy rate in India, 1991 and 2001

Note: The boundaries and the names shown and the designations used on these maps do not imply official endorsement or acceptance by the United Nations.

A quick comparison of these two maps shows that overall literacy across India increased from 1991 to 2001.

Now consider the following two maps displaying literacy rate in India disaggregated by gender for 2001.
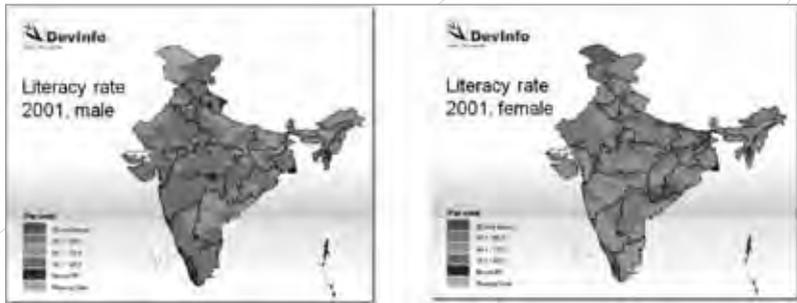
Figure 14 Literacy rate in India disaggregated by gender, 2001

Note: The boundaries and the names shown and the designations used on these maps do not imply official endorsement or acceptance by the United Nations.

A comparison of these two maps with the previous ones shows that while overall literacy rates climbed across India from 1991 to 2001, there are significant gender disparities masked by the maps showing total literacy rate. In fact, the map showing female literacy rates in 2001 resembles the map for total literacy rates in 1991. Based upon the data in these four maps, policy recommendations can be made to target gender equity on the educational front.

Another noteworthy dimension of DevInfo maps is the power of monitoring data series in animated maps. As an example, consider the following series of maps which capture the population density of India over time at selected time periods from 1901 to 2001.
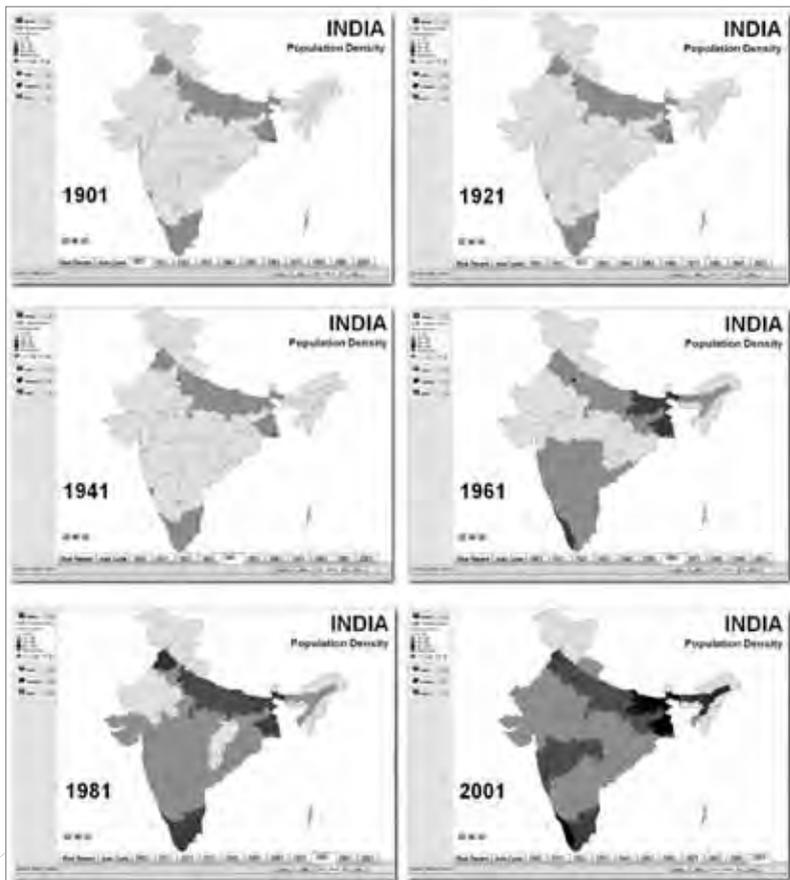
Figure 15 Population density of India over time

Note: The boundaries and the names shown and the designations used on these maps do not imply official endorsement or acceptance by the United Nations.

A comparison of these maps clearly shows the states where population density has grown most rapidly, with obvious ramifications for policy initiatives across all levels of social service delivery at the national and sub-national government levels.

Conclusions

It is clear that DevInfo use at the country, sub-national and regional levels, in supporting national monitoring and evaluation systems, is increasing around the world. Across Africa, Asia and Latin America, many countries are implementing and institutionalizing DevInfo, within their national monitoring and evaluation frameworks, to provide high quality data for policy-making duty bearers across all sectors of human development. This reflects a major drive of duty-bearers towards strengthening their data monitoring, evaluation and assessment capacities, with the goal of ensuring sustained development for all rights-holders.

Additionally, the Web 2.0 features embedded within the latest release of DevInfo promise to equip DevInfo users with a more powerful toolkit to share and disseminate data in a relevant and timely manner to concerned rights-holders, stakeholders and decision-makers.

In light of the ever-increasing global demand for high quality human development data, DevInfo remains poised to continue enabling governments and other duty-bearers, at all levels, to meet this data demand. High quality data, first organized and then appropriately disseminated, will continue to serve as the foundation for the evidence-based decision- making necessary to advance human development across the globe and thereby ensure that the promises of the Millennium Declaration are realized by all people.

References

Pron Nicolas, Oswalt Kris, Segone Marco and Sakvarelidze George. (2008). *Strengthening country data dissemination systems: Good practices in using DevInfo.*

UNDP. (2001). *A Human Rights-based Approach to Development Programming in UNDP – Adding the Missing Link.*

"Web 2.0" article, Wikipedia, http://en.wikipedia.org/wiki/Web_2.0.

THE PROFESSIONALIZATION
OF EVALUATION

by Jean Serge Quesnel,
Senior Facilitator, UNSSC, Adjunct Professor,
ENAP Québec and Carleton University,
and former Director of Evaluation at UNICEF,
IADB and CIDA

"It is not because thing are difficult that we do not try;
It is because we do not try that things are difficult."
SENEQUE

Introduction

Worldwide, there is a growing trend toward professionalization in evaluation. This is due to an exponential demand for high quality evaluations. Meta-evaluations (UNICEF, 2004; CIDA 2007) have shown that a third of evaluations are not worth their investment and another third are of uneven quality. This is somewhat disturbing when it is estimated that, over the last decade, internationally, several billions have been spent on evaluations. Nowadays, there is recognition that evaluation must be conducted at all levels of management. Often managers find themselves at a loss when commissioning evaluations, not knowing how, and where to find competent evaluators. It is however, encouraging to see that the evaluation community is organising itself. During this decade, more than a hundred regional and national evaluation associations/networks[1] have been created on all continents. The aim of these groupings is the greater professionalization of evaluation. Surveys revealed that there are now more and more universities offering graduate courses in evaluation.

Evaluation professionalization

To be professional[2] implies an allegiance to, and a performance of duties in compliance with, the stated norms and ethics. A profession is an occupation practiced for livelihood, which requires

---

1   Based on matrices of evaluation associations and networks kept by the International Organisation for Cooperation in Evaluation and the International Network of the francophonie.

2   Definition extracted from Oxford Dictionary, Oxford, UK.

skills and knowledge of some branch of learning and sciences. Professionalization[3] means to give a professional character or status. It is to establish a profession involving a body of persons engaged in an occupation. These persons must behave according to the prescripts of the profession, reflecting its values and code of conduct. They must practice their art and science according to a professional discipline guiding their methodological approaches and quest for quality products. A professional[4] is a person who earns his/her living in an occupation where amateurs may frequently be engaged, but who is recognized as being expert at her/his work through knowledge, skills and performance.

The rationale for professionalization

There is an exponential demand for professional evaluation because of the global emphasis on good governance and management by results. Evaluation is an important tool for transparency; results assessment; value-for-money; learning-by-doing; participatory decision-making; and, accountability. The International Financial Institutions have been constant, proactive users of evaluation to ensure maximum impact of their portfolios and optimisation of their resource use. The European Commission is another important leader in the use of evaluation to rally divergent countries around results and value-for-money for policy-making. The reform of the English speaking governments has set higher standards for the conduct and use of quality evaluation. Recent global trends for good governance have also, in most countries, generated greater demand for professionally conducted evaluations.

Key components for professionalization

Competencies, combined with ethics, norms and standards, provide the basis for professional credentials. Competencies are the knowledge, skills, and dispositions that evaluators need in order to conduct evaluations effectively, efficiently and of good quality (Stevahn, Ghere, Minnema. 2005).

---

3   Definition extracted from Webster's Encyclopaedic Unabridged Dictionary, Portland House, New York, 1996.

4   Webster's, opus citrate.

Norms and standards for evaluation have been developed by evaluation associations. The standards most referred to, including the Guiding Principles for Evaluators[5] were articulated by the American Evaluation Association. Similar professional standards were approved by the African Evaluation Association[6], national evaluation societies and the United Nations Evaluation Group[7] Codes of professional conduct for ethical professional behaviour have also been developed[8] Many governments have adopted evaluation policies and standards which are paramount to the establishment of professional norms for evaluators. During its General Assembly of December 2009, the International Organisation for Cooperation in Evaluation created a task group to undertake a tally of existing work in these matters and to report on convergences and divergences at the global level.

Complementing national and regional evaluation norms and guidelines with evaluation competencies is a logical subsequent step (Picciotto). A few studies (Doré and Marceau)[9] have been done on competency profiles. Noteworthy are those sponsored by the American Evaluation Association and the Canadian Evaluation Society[10]. In 2007, the United Nations Evaluation Group[11] approved competencies for evaluators. In 2009, the European Evaluation Society completed a survey[12] aiming at defining a competencies framework. The Canadian Evaluation Society (2008) prepared an interesting crosswalk of competencies. The International

5 The Guiding Principles may be found at http://www.eval.org/Publications/aea06. GPBrochure.pdf

6 African Evaluation Guidelines may be found at: http://afrea.org/content/index. cfm?navID=5&itemID=204

7 Norms and Standards for Evaluation in the UN System may be found at: http:// www.uneval.org/normsandstandards/index.jsp?doc_cat_source_id=4

8 For the code of conduct approved by UNEG see: http://search.live.com/results.asp x?q=Evaluation+Code+of+Conduct+UNEG&form=WLMLQB

9 An excellent article on competencies profile for evaluators

10 See Picciotto,; UNEG code of conduct; UN Core competencies at http://www. uneval.org/normsandstandards/index.jsp?doc_cat_source_id=4; competencies survey at http://www.europeanevaluation.org/files/http www. europeanevaluation. pdf ; Canadian Evaluation Society crosswalk of competencies at http://www. evaluationcanada.ca/distribution/20080312_ces_professional_designation_core_ committee.pdf; Perin 2005.

11 UN Core competencies at http://www.uneval.org/normsandstandards/index. jsp?doc_cat_source_id=4.

12 European Evaluation Society competencies survey at http://www. europeanevaluation.org/files/http www. europeanevaluation.pdf

Development Evaluation Association has created a committee to address the need for consolidating and to bring clarity to discussions around credentialing, including the minimum capabilities or credentials for evaluators of development interventions, at entry level; intermediate level; and, at advanced level.

Profile of a professional evaluator

Professional evaluator profiles are emerging quite harmoniously within governmental, non-state, academic and private entities. The profiles delineate the coherent requirement for professional performance. The design and conduct of an evaluation entails: the diagnosis of existing situations; identification of root causes of the problem situation; logic models for impact driven intervention; analysis of strategic action scenarios; assessment of results and their effects; articulation of relevant and valid recommendations to improve interventions; and, the optimisation of resources. An evaluator must have the ability to collect trustworthy quantitative and qualitative information and to ensure valid and rigorous analysis. An evaluator must also be cognizant of the multi-dimensional factors at play, the cross-cultural dimensions and, not least, be an excellent observer and communicator.

From the various efforts at identifying the competencies for evaluators, one can see that a convergence is building around the following competency categories:

- Professional practise – knowing the standards; being involved in practice; ethics; honest conduct and reporting; respecting clients, respondents and other stakeholders.
- Systematic inquiry – relates to technical skills.
- Situational analysis – multi-disciplinary understanding of the environment.
- Management skills – management contexts and conduct of evaluation.
- Reflective practice – being able to step back and understand change dynamics and engage in professional development.
- Interpersonal skills.

Needless to say, the level of various competencies will vary depending on whether the evaluator is junior or senior, and whether they are a manager or a commissioner. Hence the demand for evaluation

competencies is multi-facetted. Evaluation competency information would be welcomed by those universities which provide graduate studies in evaluation. Competency profiles help in the design of curriculum which aim to respond to market needs. Indeed the market is increasingly pointing to a recognition that evaluators should have adequate knowledge and understanding of the evaluation business in order to enhance the independence and integrity of the process and the effective utilization of evaluation results (Perrin, 2005).

Professional Recognition

There is the possibility of having different levels of professional recognition. Four in particular are key in the context of on-going discussion. These are:

Basic competencies: The minimum technical skill sets required to carry out evaluation work are met by an individual.

Credentialing: An attestation is given to an individual, saying that the basic professional requirements are met.

Certification: A statement is issued testifying to an individual's qualifications to conduct evaluations.

Licensing: An authority tests candidates and issues a license giving the holder a legal right to perform evaluations.

Over the decade, many evaluation associations and networks have increasingly been considering the professionalization of evaluators. The matter has also been discussed by all the regional evaluation associations and networks.

In 1999 the *American Journal of Evaluation* devoted a journal issue to the work of a task force which addressed the issue of certification of evaluators (Bickman, 1999). In 2003 the *International Handbook on Educational Evaluation* devoted a chapter to the professionalization of evaluation (Worthen, 2003). The *Gussman report* (2005) and the *Treasury Board discussion paper* provide thorough reviews of the articles written on this matter. In 2006, the Secretariat of the Treasury Board of Canada published a comprehensive discussion paper which addresses possible "strategies and options for government to foster advanced professional development for evaluators in the interests of enhancing quality assurance in the evaluation func-

tion". A literature review (Huse and James, 2006), examined the recent evaluation information related to the challenges, risks and benefits of establishing a credentialing system and, also, the risks of not developing a credentialing system.

Japan has adopted a professional recognition regime for evaluators. As of 2010, the Canadian Evaluation Society will provide the Credentialed Evaluator (CE) professional designation to those of its members who meet the requirements. The United Nations Evaluation Group has been considering the merits of a certification for UN evaluators.

Conclusion

A "certified professional evaluator" designation would be a formal attestation that a person has the appropriate knowledge; academic training; professional skills and conduct; and, the necessary practical experience in a range of evaluation roles and accomplishments. Evaluation associations have begun the long walk on the path leading to: norms and standards for evaluation; codes of professional conduct; competencies profiles; benchmark job descriptions; and, competency-based training programmes. Needless to say, the road will be arduous because evaluators want both the merits of intellectual discipline and the freedom for creative thinking. Some see professionalization as uplifting the contribution of evaluation. Others are afraid that it may be subjecting it to barriers. The challenge is to provide diversity by providing useful guidelines and not to impose formal straight-jackets.

Japan and Canada have already decided to provide professional recognition to evaluators. Many countries of Africa, Asia and Latin America are keen to do so. In the near future, one should not be surprised to see that there may also be various gradations[13] within the professional training, qualification and designation for evaluators. While the supply side of evaluation is organising itself, the demand side is increasing significantly with the advent of learning organisations striving for results; quality; value-for-money; stakeholder satisfaction; accountability; and, learning.

---

13   The Evaluators Institute has three course-based certificates. See: http://www.evaluatorsinstitute.com

References

Bickman, L. (1999). *AEA, bold or timid?* American Journal of Evaluation, 20, 519-520. 1999.

Canadian Evaluation Society (2008). *The crosswalk of competencies*, found at:
http://www.evaluationcanada.ca/distribution/20080312_ces_professional_designation_
core_committee.pdf

CIDA. (2007). *Meta-evaluation*, at http://www.acdi-cida.gc.ca/CIDAWEB/acdicida.nsf/
En/JUD-111795644-KJX

Cousins B. and Aubry T. (2006). Roles for Government in Evaluation Quality Assurance.
See http://www.tbs-sct.gc.ca/eval/tools_outils/QAE-AQE/rgeqa_e.asp

Doré, Guylaine and Marceau, Richard. (2009). *Competencies profile for evaluators,*
in the review *T*élescope of the ENAP, Quebec Vol 13 No1 (2009) See: http://www.enap.
ca/OBSERVATOIRE/fr/telescope-publications.aspx?sortcode=1.16.18.26

European Evaluation Society (2009). *Competencies survey*,
at http://www.europeanevaluation.org/files/http www. europeanevaluation.pdf

Gussman, T. K. (2005). *Improving the professionalism of evaluation*. Ottawa:
Centre for Excellence in Evaluation, Treasury Board Secretariat of Canada. May 2005.

Huse, Irene and McDavid, James C. (2006). *Professionalization of Evaluators*. Literature
Review. University of Victoria. Prepared for the CES Evaluation Professionalization
Project. October 10th 2006. See:http://www.evaluationcanada.ca/txt/2_literature_e.pdf

International Development Evaluation Association. (2009). *The IDEAS initiative
on international competencies for development evaluators*, found at:

http://www.ideas int.org/content/index.cfm?navID=16&itemID=108&CFID=415991&CF
TOKEN=66055133

Perrin, Burt. (2005). *How can information about competencies required for evaluation
be useful?* The Canadian Journal of Programme Evaluation, Volume 20, No 2, 2005

Picciotto, Robert. (2005). *The Value of Evaluation Standards*, Journal of Multi-Disciplinary
Evaluation, JMDE (3) 30, ISSN 1556-8180, 2005.

Quesnel, Jean Serge. (2006). *The Importance of Evaluation Associations and Networks*
published, in Segone, New Trends in Development Evaluation, UNICEF Regional Office
for CEE/CIS and IPEN

Stevahn, L, King, J., Ghere, G., & Minnema, J. (2005). *Establishing Essential
Competencies for Programme Evaluators*, American Journal of Evaluation, Vol. 26 No. 1,
43-59 March 2005. http://aje.sagepub.com/cgi/content/abstract/26/1/43

UNICEF. (2004). *Meta-evaluations*, at http://www.unicef.org/evaldatabase/index_
31120.html.

Smith, M. F. (1999). *Should AEA begin a process of restricting membership in the
professional of evaluation? AJ E*, 20, 521-531. 1999

Worthen, B. R. (1999). Critical *challenges confronting certification of evaluators.
A J E*, 20, 533-555. 1999

Worthen, B. R. (2003). *How can we call evaluation a profession if there are no
qualifications for practice?* In T. Kellaghan & D. L. Stufflebeam (Eds.), International
Handbook of Educational Evaluation (pp. 329-344). Boston: Kluwer. 2003.

PROFESSIONALIZING DEVELOPMENT EVALUATORS. TRAVELLING THE ROAD TO RESULTS

by Linda G. Morra Imas,
Advisor for evaluation capacity development,
The World Bank, and Co-Director, International Program
for Development Evaluation Training (IPDET)

Introduction

In any endeavor, the road to results begins with a sound theory of change. The theory of change in producing monitoring and evaluation training materials and conducting training programmes is first, put succinctly, that we will produce better monitoring and evaluation of interventions than would have been the case in the absence of training and materials. Secondly, that because we have good monitoring information and good evaluations, decision-makers will increasingly use them to make evidence-based decisions about the progress and effectiveness of programmes. So, ineffective interventions cease, those which hold promise improve, and those demonstrating strong positive results will scale-up. That is the road to results being traveled by the development community in pursuit of good monitoring and evaluation of country programmes. Development evaluation has moved away from traditional implementation and output-based evaluation, toward results-based evaluation.

The International Programme for Development Evaluation Training (IPDET)[1], now 10-years-old, and the textbook that has grown out of it, *The Road to Results: Designing and Conducting Effective Development Evaluations (Morra Imas and Rist, 2009)*, break-down monitoring and evaluation practice into some 14 modules in order to provide an overview, knowledge, and some practical skills. The road begins with understanding the foundations of development evaluation, and proceeds with how to develop, implement, and sustain a monitoring and evaluation system. It continues by showing how to prepare, design, and conduct a professional evaluation. Critical is how to manage the evaluation and evaluation teams along the road and how to communicate progress, results, and the learning.

---

1   See www.ipdet.org for information on this programme.

We know that roadmaps do not show how serpentine and complex roads may be. We also know that arriving at "good practice" for monitoring and evaluation of development interventions is not an easy road. Building capacity to conduct monitoring and evaluation, and using the findings effectively, takes more than short-term training and textbooks. Although both are necessary, we know that evaluation capacity development is a complex process involving such things as: the building of national evaluation associations; strengthening of cultures of accountability within public sectors; creating incentives for use of evaluation; growing a pipeline of skilled personnel, through university-based graduate programmes in evaluation, and so on. The competencies: knowledge; skills; and, attitudes required for good evaluators of development interventions are many and complex. Textbooks such as *The Road to Results* are a starting point for identifying competency categories and key knowledge, skills, and attitudes. The goal is not to create barriers to entry, but to encourage and facilitate entry into development monitoring and evaluation.

This chapter focuses on the issue of competencies in relation to the evaluation of development interventions. It explores why competencies are attracting so much attention. It describes briefly some of the efforts being made by national and regional professional evaluation organizations to delineate evaluator competence within their specific contexts. Others efforts are described in other chapters in this book. This chapter, however, focuses on the issue of *global* competencies for those involved in the monitoring and evaluation of development interventions. It describes an effort to develop international standard competencies in monitoring and evaluation and the challenges and specific issues which that presents.

Drivers of evaluation capacity building

An evaluation function that relies on external experts is not likely to meet long-term needs. *The Paris Declaration on Aid Effectiveness and The Accra Agenda for Action* brought a continuing emphasis on multiple actions to strengthen national monitoring and evaluation systems in developing countries. Generally, evaluation capacity building assists countries in:

- using monitoring and evaluation findings to allocate government resources in planning, decision making, and prioritization, particularly in the budget process;

- understanding the performance of ongoing activities at the sector, programme, and project levels for learning and improvement;

- holding managers accountable for the performance of the activities they manage, so that government can be held accountable for performance;

- demonstrating the extent to which development initiatives have been successful in obtaining desired results.

In most countries at the present time, industrialized as well as developing, anyone can decide to take on the identity of evaluator. With few exceptions, there are as yet no standard evaluation competencies to be met through training or practice, nor certifying or credentialing process that give a seal of approval to would-be evaluators. Many people learn on the job without formal training. Managing or directing evaluation units likewise requires no specific competencies: neither knowledge nor experience is a prerequisite in most organizations. Certainly the time-honored method of learning on the job should not be discounted and managers without formal evaluation training can grow into the job. However, the issue for many is that the lack of competencies and a related certifying or credentialing system makes it difficult for purchasers of evaluation services to determine expertise. As pressures increase for government accountability, so do the calls for evaluation. And today increased calls for professionalism are being heard in response to some poorly conceived and executed evaluations.

Competency frameworks and certification, credentialing, and accreditation systems

Often there is confusion in using the terms competency, certification, credentialing, and accreditation. Competency frameworks define the knowledge, skills, and attitudes that members of a profession should have. They may stand alone as general guides to a profession or they may be part of certification, credentialing, and/or accreditation systems. There are differences among certification, credentialing, and accreditation systems. The differences are important in discussion of competencies as they relate to who has the burden of proof, and the depth of the proof, of competency.

Certification systems are voluntary. They are not required for an individual's practice of a profession, but are obtained because they are viewed as providing particular benefits to the individual.

In a certification system, generally, an individual attests to an independent review group that he/she has the knowledge, training, and experience to meet competencies, or some proportion of competencies, in a field. If the certifying group agrees that the combination of training, knowledge, and experience indicated by the individual provides basic competency, it may provide a certificate to the individual to that effect. Sometimes there are levels of competency from entry level, or novice, to expert that can be certified.

In a credentialing system, in contrast to certification, an independent body generally attests or validates that an individual does in fact have the needed knowledge, skills, attitudes, and experience to practice the profession. It provides a credential to that effect to the individual. A passing score on a standard test is commonly required as evidence of knowledge. Government licensing may also be involved. *The Certified Public Accountant* (CPA) is a common example of a credentialing programme. In order to gain the CPA designation in the United States, for example, an individual must obtain a valid CPA certificate issued by the authority of a state; pass the Uniform CPA Exam; complete 120 hours of continuing professional education every three years; practice a minimum amount of time (usually one or two years as determined by the State), in a firm enrolled in an approved practice-monitoring programme, and agree to abide by the Code of Professional Conduct. Board-certified physicians in many countries are also recipients of credentials and have had to pass special exams and often take additional training in order to be deemed "Board-certified".

Accreditation, in contrast to certification or credentialing, is usually provided by a government to an organization which meets government standards for the operation. Educational or training organizations and programmes are often accredited.

Arguments in support of and against evaluation competency frameworks and certification or credentialing systems

The arguments are many in support of competency systems, but also in opposition to them. All of those supporting competency frameworks do not necessarily agree that they should go hand-in-hand with a system either for certifying competency, or providing a credential. Some of the common arguments for the evaluation profession are summarized in the following tables.

COMMON ARGUMENTS IN SUPPORT OF COMPETENCY FRAMEWORKS

Increase credibility of the evaluation function as a profession and increase recognition of evaluation staff as professionals

Increase consistency and methodological rigor within the evaluation community

Increase available training

Increase the skills of those in the evaluation community as more training will become available to complement the competencies

Decrease barriers to entry and broaden the talent pool

Promote self-responsibility for continuing improvement of evaluation skills

Provide a basis for selecting or procuring the services of evaluators

Provide a basis for developing education and training programmes

Promote an ethical code of practice

ADDITIONAL ARGUMENTS IN SUPPORT OF CERTIFICATION OR CREDENTIALING SYSTEMS

Show that one has successfully demonstrated knowledge, skills, training, and experience to an independent board and that one abides by a code of ethics

Provide prestige

Provide incentives by enhancing marketability and salaries

Provides a basis for disciplining those who do not follow ethical codes or who misrepresent their capabilities and/or experience

Avoid narrow competency definitions that may be self-serving to specific organizations

Help prevent poorly qualified persons practicing the profession from undermining the public trust and confidence

COMMON ARGUMENTS IN OPPOSITION TO COMPETENCY FRAMEWORKS

No empirical evidence that specific competencies are necessary or critical to the practice of the profession in all contexts

No evidence that consistency in backgrounds and experience is desirable for the profession

Not enough available education or training programmes to provide the required skills to all

Competency statements cannot keep up with a rapidly changing profession

Given the diversity of contexts in which the evaluation profession is practiced, no one set of competencies can be specified or agreed to

Cost of education and training and access to education and training serve as barriers to entry

Given the diversity of contexts in which the evaluation profession is practiced, no one individual can meet all the required competencies

What is most important to evaluation – the attitudes and disposition of the evaluator – cannot be measured

ADDITIONAL ARGUMENTS IN OPPOSITION TO CERTIFICATION OR CREDENTIALING SYSTEMS

Promotes exclusivity

Works unfairly against those who have learned primarily through doing and have years of experience but little formal training

Little assurance that there would be consistency in application across certifying/credentialing groups

Sufficient competition exists in the workplace and no efforts to guide, constrain, or regulate the profession are necessary

Education and training programmes should be accredited; certification or credentialing of individuals is not needed

Certification and credentialing systems cannot prevent poor practice from those holding certificates or credentials, nor do those graduating from accredited programmes necessarily have the skills and knowledge needed

Does not adequately recognize alternate forms of education such as mentoring, coaching, and twinning

Competency frameworks and national/regional professional evaluation associations

While the debate continues as to whether evaluator competency definitions are needed, as well as related certification or credentialing systems, some associations are moving forward on them. Generally, surveys of their members are the basis of the support for moving forward.

Canada has been a pioneer both in pursuing core competencies and in developing a voluntary system of professional designations for evaluation. Starting in 2000, the *Canadian Evaluation Society* (CES) worked to identify core competencies for Canadian evaluation practice with the CES National Council approving the competencies in 2009.[2] The competencies centre around five categories:

(1) *reflective practice competencies*, which focus on the fundamental norms and values underlying evaluation practice and one's awareness of their evaluation expertise and needs for growth;

(2) *technical practice competencies*, which focus on the specialized aspects of evaluation such as design data collection and analysis;

(3) *situational practice competencies*, which focus on the application of evaluative thinking in analyzing and attending to the unique interests, issues, and contextual circumstances in which evaluation skills are being applied;

(4) *management practice competencies*, which focus on the process of managing a project/evaluation, such as budgeting, coordinating resources and supervising, and

(5) *interpersonal practice competencies*, focusing on people skills, such as communication and conflict resolution.

Also in 2009, the CES National Council approved the related professional designations policy and implementation plan. This professional designation allows members of CES to apply for what is, in our terms, a Certified Evaluator designation, which will mean that its holder "has provided evidence of education and experience to be competent" in the field of evaluation. The designation's requirements focus on the holder's level of education, length of experience in evaluation, and an alignment of this education and commitment to continuously update skills with continuing professional

---

2   A comprehensive history of the Professional Designations Project is found at www.evaluationcanada.ca.

development. The goal is to invite CES members to apply for the designation early in 2010. Notably, not all competencies need to be met for the designation. CES has decided that if 70 percent are met in each of the five domains, that is sufficient.

Unlike Canada, the *American Evaluation Association* (AEA) has never endorsed or adopted competencies for evaluators. However, considerable work was done in identifying competencies, from 2001 to 2006, by Jean King of the University of Minnesota. King is active in AEA and founder and co-chair of AEA's Topical Interest Group on Organizational Learning and Evaluation Capacity Building. Others have also contributed to this effort (King, Stevahn, Ghere, Minnema, 2001; Stevahn, King, Ghere, Minnema, 2005; Ghere, King, Stevahn, Minnema, 2006). While CES drew on and used much of this work, AEA has limited itself to developing *Guiding Principles for Evaluators*. Many other national and regional evaluation associations, such as the regional *African Evaluation Association* (AfrEA), have also developed guidelines and/or ethical standards for evaluation, but have stopped short of evaluation competencies.

The *Japan Evaluation Society*[3] has recently also developed a certification programme for professional evaluators that is tied to a specific Certified Professional Evaluator Training Programme. Three levels of certification are envisioned, but at this time, only the Basic level is developed. The program could be considered to grant a credential in that participants must pass a written exam to be certified. The first such exam has been given and as of December 2009, there is now a cohort of certified professional evaluators.

In Europe, the *German Evaluation Society* (DeGEval) has recently developed recommendations for education and training in evaluation (Caspari, Hennen, Scheffler, Schmitt; Schwab, 2008). Five competency fields are described:

(1) theory and history of evaluation;

(2) methodological competencies;

(3) organizational and subject knowledge;

(4) social and personal competencies; and

(5) evaluation practice.

---

3   Detailed information on the Certified Professional Evaluators Training Program is only available in Japanese: http://www.idcj.or.jp/9evaluation/CPE4_0.html.

Picciotto (2009b), characterizes the DeGEval approach as "input based" in that it focuses on the content of education and training programmes capable of generating the skills, knowledge, and mastery needed to contribute to high quality work. In contrast, "outcome based" approaches present competencies in terms of the results of evaluators' activities.

The *European Evaluation Society* (EES), in an initiative led by Robert Picciotto, recently surveyed their membership about the rationale, scope, and content of a competencies framework suitable for European evaluators. Unfortunately, timing and technical problems made for a low response rate (only 48 respondents). However, almost four fifths gave a high or very high priority score to an initiative on competencies for evaluation professionals, and the many additional comments were valuable (Picciotto, 2009a). A concise draft framework for evaluation competencies focuses on three categories of competence:

(1) evaluation knowledge;

(2) professional practice; and

(3) dispositions and attitudes.

Next steps are yet to be determined.

Competency frameworks in universities and development organizations

A need for information on evaluation competencies has also been reported among six European universities offering master's degrees in evaluation.[4] These universities are working on a detailed list of competencies which may serve as a basis for accrediting evaluation programmes across Europe.

In the development context, no national evaluation professional association has as yet moved to survey their membership on a competency initiative or to draft a framework, however, several development organizations have established competency systems for evaluation personnel. *The Evaluation Capacity Task Force* of the *United Nations Evaluation Groups* (UNEG) has, for example, developed core competencies for evaluation staff at different levels of

---

[4]   Reported by Picciotto (2009a) as Suddansk Universiteit, Odense (Denmark), Universite de Lyon (France), Universitat Sarrebruken (Germany), Universidad Complutense de Madrid (Spain), Universitat Bern (Switzerland) and London Metropolitan University (United Kingdom).

expertise. They were adopted in 2008, as well as core competencies for those who head evaluation units.[5] In the UN, evaluation competencies refer to the qualifications, skills, experience, and attributes required by those employed within the evaluation function, in order to carry out the duties stipulated and, to ensure the credibility of the process. The competencies are more detailed and specific than is often the case. For example, for heads of evaluation, they cover: specific knowledge of the United Nations; technical and professional skills; interpersonal skills; personal attributes; and, management skills.

The United Kingdom's *Department for International Development* (DFID), has recently joined the list of organizations using evaluation competencies.[6] Most interesting is DFID's use of the competencies to establish a pool of accredited evaluation specialists and, the specificity of the competencies. The framework has four levels: expert, skilled, competent, and basic. A competency is, for example, "Best practice in evaluation approaches and methods." In turn, this is defined as "Understands and makes appropriate use of evaluation concepts and methods, helping DFID to achieve best practice in evaluation design and approaches in line with DFID policy, and relevant international and professional technical standards". For each of the framework's four levels, the required knowledge, skills, and/or behaviors are then defined separately for each competency. Other competencies include: gathering and using evidence for evaluation; communicating and sharing evaluation findings, knowledge, and expertise; upholding evaluation independence, quality, and standards; and leading, managing, and delivering evaluations.

---

5   See: Core competencies for evaluators of the UN system UNEG/G/2(2008) and UNEG job descriptions for evaluators in the UN system (P1-P5), UNEG/G/JD/P1-P5 (2008). Also, UNEG core competencies for heads of evaluation offices in the United Nations (UNEG/G(2008)2.

6   DFID. Evaluation Department. Technical Competency Framework. Note these standards were draft as of end-July 2009. See www.ideas-int.org/documents/ Document.cfm?docID=300

Competency frameworks and standard-setting organizations

The International Board of Standards for Training, Performance, and Instruction (IBSTPI) has also established standards for the practice of evaluation.[7] Developed in 2006, the standards cover 14 evaluator competencies in 4 domains:

(1) professional foundations;

(2) planning and designing the evaluation;

(3) implementing the evaluation plan; and

(4) managing the evaluation.

IBSTPI claims to have validated the standards internationally with more than 450 practitioners. However, IBSTPI indicates that these are competencies required by internal staff or external consultants conducting evaluations in organizational settings (such as for-profit and not-for-profit organizations, military, and government agencies), evaluating their own internal programmes. They are not competencies required for evaluators examining the effectiveness of large scale state-wide or national programmes which are often funded by government departments or agencies. This constraint narrows potential use of these standards.

IDEAS: A global initiative to develop competencies for development evaluators

The International Development Evaluation Association (IDEAS) is the only global association for those working on evaluation of development interventions. It is a young organization, but has struck a chord with many, often as a secondary professional membership. In its General Membership Meeting in July 2008, the IDEAS Board approved an initiative to explore the issue of competencies for international development evaluators. A volunteer working group was formed, headed by Linda Morra Imas. Initial interest was overwhelming for a small association with some 60 volunteers. The terms of reference laid out several principles for the working group. The first was that reinventing the wheel was to be avoided. The IDEAS effort would use research and work already done by others to inform and to speed its efforts. *The Canadian Evaluation Society*,

---

7   http://www.ibstpi.org/Competencies/evaluatorcompetencies.htm

for example, had done a large amount of research to link existing professional evaluation knowledge frameworks. Secondly, the discourse with the membership was to be especially valued given the diverse cultures and backgrounds of IDEAS members. The third point, which was the real value added, was that the IDEAS *Competency Working Group* should focus on issues of particular concern to development evaluation and development evaluators. Fourth was that, in tandem with the effort to develop competencies, the working group would develop a code of ethics or guiding principles.

Having established these terms of reference, the first step was to determine if the initiative had from its members, the general support to proceed. With about 250 active IDEAS members[8] at the time of the survey, the 149 partial or complete returns was encouraging and represented a 50-60 percent response rate (most analyses used complete returns which numbered 122). Analysis indicated that members from Africa and Asia were well-represented, but those from Latin America and the Caribbean appeared underrepresented. Only 10 percent viewed evaluation as their major academic background with about 18 percent citing economics as their major discipline and most other respondents affiliating with social sciences such as anthropology, psychology, and sociology. Almost an equal number of respondents worked in non-profits and in governments (42 percent). About 30 percent identified themselves as consultants. Some 40 percent of respondents indicated they were evaluation officers and another 14 percent said they were project officers.

The survey findings showed overwhelming support for a competency framework. Over 90 percent of respondents indicated that a competencies framework for international development evaluators was needed and about the same proportion "strongly agree" or "agree" that such a framework would promote professionalism. Almost all respondents (93 percent) believed that competencies for development evaluators would promote better training. Most also believed that such competencies would improve the quality of evaluations (86 percent), encourage continuous education for professionals (85 percent), and improve managers' hiring decisions (80 percent). The survey found similarly broad support for IDEAS determining the skills that development evaluators should possess

---

8   Encouragingly, IDEAS membership grew to 400 by July 2009 and to over 700 as of December 2009.

(78 percent), and identifying evaluation training related to the competencies (83 percent).

The survey also asked questions about certification which also found support, but scored less high than the competency framework itself. Nevertheless, 74 percent of respondents indicated that they would be motivated to become certified if a competency and credentialing system was established. There was less agreement on whether IDEAS had enough recognition to be in charge of such a credentialing system (40 percent).

Overall, the survey response rate and findings showed:

(1) strong membership support for the initiative to develop competencies for those involved in evaluating development interventions; and

(2) significant support for a certification system.

There was less clarity around which organization should do the certifying.

The survey results were elaborated upon through discussion at the 2008 IDEAS Global Assembly in Johannesburg. This emphasized the need for competencies in development-specific areas. These areas included, for example: gender evaluation; multi-cultural knowledge and sensitivity; evaluation transparency and independence; evaluation of climate change; and, evaluation of natural disaster assistance. While some attendees promoted the evaluator as advocate, there was no general support for the concept. Some questioned how many of the potential competencies any given individual could be expected to meet and, whether it made more sense to think of them in the context of the evaluation team composition meeting the competencies overall. There was strong agreement that the competencies for an evaluation manager and for an evaluator were different and that competencies should be developed for each. Discussion also elicited strong support for the identification of competencies for those who requisition or commission evaluations. Another area of discussion was the need for any certification system to have strong sanctions for those who violated the principles, or fraudulently presented credentials they did not have. Finally, delegates agreed that the financial crisis was hurting development and made it all the more pressing that good evaluation showed what did and what did not work.

Following the IDEAS Global Assembly and its broad mandate, the competency initiative grew to seven subgroups, each with their own head and volunteers. These were: the Theory of Change (Harold Pohoresky); Core Competencies (Cristina Magro); Competencies for Evaluation Managers and Commissioners of Evaluations (Rashmi Agrawal); Ethical Standards and Guidelines (Reinhard Skinner); Certification and Credentialing (on hold); Member Consultations (Luis Bernal); and Communications (Cush Ngonzo). Brief background papers were developed on perspectives on competencies from regions (e.g. South Asia perspective) and countries (e.g. Malawi). Additional information on competencies was gathered, for example, from the doctoral programme at the University of Western Michigan. It quickly became apparent that a tiered system of core competencies would be needed, to ensure competency systems were not a barrier to entry for many from developing countries, where learning by doing is often the only option. By December 2009, the outlines of the various parts were drafted and were being filled in. As with any large volunteer effort, especially one that attempts to reflect the perspectives of many diverse cultures and backgrounds, progress has come in fits and starts. The efforts of those such as CES, which precede us, suggest that a multi-year time frame is not unrealistic. The following are some highlights from key subgroups.

Ethical standards and guidelines

The subgroup is developing guidance around six Canons:

CANON I General Obligations

CANON II Obligations to the Public and Stakeholders

CANON III Obligations to the Client

CANON IV Obligations to the Profession

CANON V Obligations to Colleagues

CANON VI Obligations to the Field of Development

Each Canon begins with Level I stating the general principles of conduct. Canon II, for example, reads:

[Draft] Evaluators should respect the laws and customs of the country in which their work is carried out and protect and respect the welfare, dignity, and rights of all evaluation stakeholders. This includes, though is not limited to, providing a professional level

of service, maintaining objectivity, and ensuring fairness in dealings with the public and stakeholders.

Level II presents the more specific ethical standards to which each member should adhere and Level III the specific rules of conduct. The subgroup views violation of a rule as potential grounds for disciplinary action. Again using Canon II as an example, ethical standards are drafted for the standard categories of Conduct, Competence, Integrity and Objectivity, Transparency, and Accountability. Ethical Standard 2.3 Integrity and Objectivity is:

[Draft] Evaluators shall maintain objectivity in their work and be equitable and fair in their dealings with all members of the public and stakeholders.

Rules for Ethical Standard 2.3 include:

[Draft] Rule 2.3.3: Informed consent must be obtained from respondents before data are collected from them.

Rule 2.3.4: Responses should be kept anonymous unless the informed consent which is given includes agreement to disclosure of identity.

As evident just from this small sample, once fully drafted ensuring the internal consistency of the Development Evaluators' Code of Ethics and Professional Conduct will itself pose a challenge.

Competencies for managers of development evaluations and competencies for commissioners of evaluations

The subgroup is developing competencies for the categories:

(1) Knowledge: Factual information needed;

(2) Skills: Direct application of knowledge; and

(3) Attitudes or Ethics.

This last category may be subsumed later as part of the Code of Ethics. Under "Knowledge Competencies for Managers of Development Evaluations", for example, are knowledge of various evaluation methods and their relative strengths and weaknesses in the particular context. Skill examples are: routinely ensures consideration of gender equity issues in designing evaluations; assesses the adequacy of resources and staff/consultant education, experience, cultural competence, and skills needed for the evaluation.

An example of "Attitudes or Ethics" is: seeks to conduct evaluations jointly with partners. "Competencies for Commissioners of Evaluations" stresses the "Attitudes or Ethics" category with statements such as: refrains from seeking to influence the outcomes of the evaluation.

Core competencies for development evaluators

This subgroup began by reviewing already developed competency frameworks for evaluators, such as the CES Professional Designation. They concluded that short lists of generic competencies are not particularly helpful. The subgroup liked the categories used by CES and defined them as follows:

1. *Reflective Practice* competencies: focus on the fundamental norms and values underlying evaluation practice and one's awareness of their evaluation expertise and needs for growth.

2. *Technical Practice* competencies: focus on the specialized aspects of evaluation, such as design, data collection, analysis, interpretation, and reporting.

3. *Situational Practice* competencies: focus on the application of evaluative thinking in analyzing and attending to the unique interests, issues, and contextual circumstances in which evaluation skills are being applied.

4. *Management Practice* competencies: focus on the process of managing an evaluation as a project, including budgeting, resource coordination, supervision, and the like.

5. *Interpersonal Practice* competencies: focus on people skills, such as communication, negotiation, conflict resolution, collaboration, and diversity.

As part of each review the competencies are examined to see if they are sector specific; could be complemented in the framework by sector specific competencies; are limited to large or small-scale evaluations, internal or external evaluations, or limited in other ways. Each is also examined for strengths and weaknesses and its fit within the developing country context. The next steps for the subgroup are to: pull together the applicable competencies for each category; modify and add to them as necessary for the development context; and, to determine the appropriate level (basic, intermediate, advanced or other labels).

Continuing progress on the *IDEAS Competencies for Development Evaluators* can be tracked by going to the IDEAS website http://www.ideas-int.org and clicking on "Current Projects".

Conclusions

It is clear that, despite some arguments against them, competencies are here to stay for many who work in the evaluation profession. With demand for evaluation grown rapidly in response to increased needs for government accountability and learning, purchasers of evaluation services have not always received good quality evaluations. They are calling for evaluator competencies as a means of determining qualifications up-front before resources are expended. This is particularly the case for those working in the developing country context.

Issues surrounding certification or credentialing appear to be more sensitive than for competency development. There is some fear that such systems in evaluation will freeze-out many of the profession. Again, developing country evaluators who must rely more on learning by doing, rather than from non-existent or scarce training programmes, might be the most concerned. However, some initial data suggest that they are, by and large, supporters of such systems, and of strong sanctions for those who unethically embellish their credentials. Many organizations which are establishing competencies will be watching closely the forthcoming implementation of a professional designation by the pioneer *Canadian Evaluation Society*. However efforts to develop competencies and certification or credentialing systems can be separated to follow two separate action tracks.

The IDEAS competency initiative has much promise as a global initiative for those working in development evaluation. As CES can testify, such efforts are massive undertakings. For CES, ten years in the making. For them, reliance on volunteer efforts has pluses and minuses. It encourages pluralism and dialog, but makes for slow going since most volunteers, although well-intentioned, are constrained by the demands of their day-job. With enormous support for the initiative from its membership, development is proceeding within IDEAS on a code of ethics, core competencies, and competencies for evaluation managers and the commissioners of evaluation. Drafts for circulation and comment within the IDEAS membership and the global development community are eagerly awaited by many.

References

Caspari, A., Hennen, M., Scheffler, D., Schmitt, U., and Schwab, O. (2008). *Recommendations on Education and Training in evaluation: Requirement profiles for evaluators*. DeGEval, Mainz, Germany, www.degevak.de.

DFID. Evaluation Department. *Technical Competency Framework*. See www.ideas-int.org/documents/Document.cfm?docID=300

Ghere, G., King, J.A., Stevahn, L., and and Minnema, J. (2006). *Linking effective professional development and program evaluator competencies.* American Journal of Evaluation, 27(1), 108-123.

King, J.A., Stevahn, L., Ghere, G., and Minnema, J. (2001). *Toward a taxonomy of essential evaluator competencies.* American Journal of Evaluation, 22(2), 229-247.

Morra Imas, Linda G. and Ray C. Rist. (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, DC: The World Bank.

Picciotto, R. (2009a). *Results of a society wide consultation on a proposed evaluation competencies framework for Europe.* Paper prepared for the European Evaluation Society.

Picciotto, R. (2009b). *Towards a competencies framework for evaluation in Europe.* Paper prepared for the European Evaluation Society.

Stevahn, L., King, J.A., Ghere, G., and Minnema, J. (2005). *Establishing essential competencies for program evaluators.* American Journal of Evaluation, 26(1), 43-59.

INSTITUTIONALIZING IMPACT EVALUATION. A KEY ELEMENT IN STRENGTHENING COUNTRY-LED MONITORING AND EVALUATION SYSTEMS[1]

by Michael Bamberger,
independent consultant

Introduction

With the growing emphasis on the assessment of aid effectiveness and the results of development interventions, it is no longer sufficient to report how much money has been invested in programmes or what outputs have been produced. Donor countries and aid recipients are demanding to know how well development interventions achieved their intended objectives, how results compared with alternative uses of scarce resources, and how effectively the programmes contributed to broad development objectives such as the Millennium Development Goals and the eradication of world poverty.

These demands have led to an increase in the number and sophistication of impact evaluations. These have often improved the efficiency and effectiveness of ongoing programmes; helped formulate future policies; strengthened budget planning and financial management; and, provided a more rigorous and transparent rationale for the continuation or termination of particular programmes.[2] However, many impact evaluations are selected in an ad hoc and opportunistic manner, with the selection depending on the availability of funds, or the interest of donors and, although they may have made important contributions to the particular programme or policy being evaluated, their potential contribution to broader development strategies was often not fully achieved.

---

1   This chapter is adapted from the World Bank Independent Evaluation Group (IEG, 2009) publication "Institutionalizing Impact Evaluation Within the Framework of a Monitoring and Evaluation System". This publication drew extensively on MacKay 2007 "How to Build M&E Systems to Support Better Government".

2   See Bamberger, MacKay and Ooi (2004 and 2005) for a discussion of different ways in which impact evaluations have contributed to development management.

Many funding agencies and evaluation specialists have tended to assume that, once a developing country government has seen the benefits of a few well-designed impact evaluations, the process of building a systematic approach for identifying, implementing, and using evaluations at the sector and national levels, will be relatively straightforward. However, many countries with decades of experience with project and programme evaluation have made little progress toward institutionalizing the selection, design, and utilization of impact evaluations.

This chapter describes the progress that is being made in the transition from individual impact evaluations studies to building a systematic approach to identifying, implementing, and using evaluations at sector and national levels. When this is achieved, the benefits of a regular programme of impact evaluations, as a tool for budgetary planning, policy formulation, management and accountability, begin to be appreciated. To date, the institutionalization of impact evaluations has only been achieved in a few developing countries, mainly in Latin America, but others have started the process of institutionalization. This chapter draws lessons on the benefits of an institutionalized approach to impact evaluations; the conditions that favor it; the challenges limiting progress; and, the steps in developing such an approach.

Although this chapter focuses on impact evaluation, it is argued that the institutionalization of impact evaluations can only be achieved within the framework of a comprehensive monitoring and evaluation (M&E) system that provides a menu of evaluations, to cover all the information needs of managers, planners, and policy-makers.

The importance of impact evaluation for Official Development Assistance

The primary goal of Official Development Assistance (ODA) programmes is to contribute to reducing poverty, promoting economic growth and achieving sustainable development. In order to assess the effectiveness of ODA programmes in contributing to these goals it is important to conduct systematic analysis of development effectiveness. Two common ways to do this are the development of monitoring and evaluation systems and Results Based Management (RBM), (Kusek and Rist 2004). However, in most cases these approaches only measure changes in the conditions of the target population (beneficiaries), whom the interventions are intended to affect. They usually do not include a comparison (or control) group not affected by the programme intervention. Consequently it is difficult, based solely

on this information, to determine the extent to which the observed changes can be attributed to the effects of the project and not to other unrelated factors such as changes in the local or national economy, changes in government policies, or similar programmes initiated by government, other donors or NGOs. An unbiased estimate of the impacts of ODA programmes requires the use of a *counterfactual* (the comparison or control group), that can isolate changes attributable to the programme from the effect of these other factors. The purpose of impact evaluations methodology is to provide unbiased estimates of the true impacts of ODA interventions.

Many assessments of ODA effectiveness are based either on data that is only collected from project beneficiaries *after* the project has been implemented, or on the use of monitoring and evaluation or results based management to measure changes that have taken place in the target population over the life of the project. In either case data is only generated on the target population and no information is collected on sectors that do not benefit from the project, or that in some cases may even be worse off as a result of the project (see Box 1). With all of these approaches there is a tendency for the evaluation to over-estimate the true benefits or effects produced by the project (Bamberger 2009b). Typically only project beneficiaries and the government and NGOs actively involved in the project are interviewed, and in most cases they will have a positive opinion of the project (or will not wish to criticize it publicly). None of the families or communities who do not benefit are interviewed and, also, the evaluation does not present any information on the experiences or opinions of these non-beneficiary groups. As a result, ODA agencies are mainly receiving positive feedback and they are lead to believe that their projects are producing more benefits than is really the case. As the results of most evaluations are positive, the ODA agencies may not have an incentive to question the methodological validity of the evaluations – most of which are methodologically weak and often biased. Consequently there is a serious risk that ODA agencies may continue to fund programmes which are producing impacts smaller than are reported and which may even be producing negative consequences for some sectors of the target population. Consequently, there is a danger that unless impact evaluation methodologies are improved, ODA resources may not be allocated in the most cost-effective way.

Although no hard statistics are available, it is quite likely that rigorous impact evaluation designs are only used in perhaps 10 percent of ODA impact evaluations. Given the widespread recognition by ODA agencies of the importance of rigorous impact evaluations,

why are so few conducted? There are many reasons, including: the limited evaluation budgets of many agencies; that most evaluations are not commissioned until late in the project cycle; and, that consultants are only given a very short time to conduct data collection. Also many government agencies see evaluation as a threat, or something that will demand a lot of management time without producing useful findings, and many ODA agencies are more concerned to avoid critical findings (which might create tensions with host country agencies, or prejudice future funding), than they are to ensure a rigorous and impartial assessment of the programmes. Also, as many evaluations produce a *positive bias* (see Box 1) and show programmes in a positive light (or under-emphasize negative aspects), many agencies do not feel the need for more rigorous (as well as more expensive and time-consuming), evaluation methodologies. One of the challenges for the institutionalization of impact evaluation is to convince both ODA agencies and host country governments, that rigorous and objective impact evaluations can become valuable budgetary, policy and management tools.

Box 1. The danger of over-estimating project impact when the evaluation does not collect information on comparison groups who have not benefited from the project intervention

For reasons of budget and time constraints, a high proportion of evaluations commissioned to assess the effectiveness and impacts of ODA projects only interview project beneficiaries and the agencies directly involved in the projects. When this happens there is a danger of a positive bias in which the favorable effects of the interventions are over-estimated, and the negative consequences are ignored or under-estimated. However, if these negative effects and impacts are taken into account, the net positive impact of the project on the total intended target population may be significantly reduced.

An evaluation of a micro-credit programme promoting the manufacture and marketing of traditional carpets in Bolivia, interviewed a sample of carpet makers who had received loans. It was found that on average their income from the sale of carpets had increased significantly. It was concluded that microcredit was an effective instrument for increasing the income of traditional artisans and reducing poverty. However, when carpet manufacturers who had not received loans were interviewed it was found that on average their income had declined. One of the reasons was that loan recipients were able to rent or purchase a vehicle to get their carpets to market more quickly and cheaply, which gave them a competitive advantage. On the basis of this additional information it was estimated that the total sales of carpets had probably not increased very much, but that a larger share of the market was now controlled by loan recipients. The inclusion of a control group, artisans who did not receive loans, radically changed the conclusions of the evaluation.

Consultants commissioned to evaluate the impacts of food-for-work programmes on women's economic and social empowerment were only able to spend an average of 3-4 days visiting project locations and meeting with affected communities in each country. Typically the consultants met with the local government agencies managing the projects, the local NGOs responsible for implementing the projects and residents of the communities where the programmes operated. In each community consultants met with groups of women participating in the food-for-work programmes and with many of their husbands. It was apparent that the project had produced significant increases in the income of the women, that their husbands were very supportive of the economic activities of their wives and that there was convincing evidence of the women's economic empowerment and increase in their self-confidence and social empowerment. In most cases the evaluation ended at this point and a very positive report was produced. However, in several cases the consultants also contacted key informants not involved in the project in order to obtain information on the experiences of other women who had not participated in the project. For example, a local nurse who regularly visited women in these same communities reported that many women who had attended the initial meetings had been forbidden by their husbands to participate in the project and in quite a few cases had been beaten for attending without his permission. Many men were unemployed and felt humiliated that they were not able to fulfill their traditional role of providing economically for their family. This again illustrates that a completely different image of the project was obtained when an effort was made to obtain information on the situation of non-participants and not to simply base the evaluation report on meetings with those who had benefited.

Defining impact evaluation

The primary purpose of impact evaluation is to estimate the magnitude and distribution of changes in outcome and impact indicators among different segments of the target population, and the extent to which these changes can be attributed to the interventions being evaluated. Impact evaluation can be used to assess the impacts of *projects* (clearly defined and time-bound interventions with a defined funding); *programmes* (broader interventions that often comprise a number of projects, with multiple interventions, a wider geographical coverage and often no end date); and *policies* (broad strategies designed to strengthen or change how government agencies operate or to introduce new economic, fiscal, or administrative initiatives). Impact evaluation methodologies were originally developed to assess the impacts of precisely defined project and programme interventions. A major challenge is to adapt these methodologies to evaluate multi-component, multi-donor country-level support packages which are becoming the central focus of development assistance.

A well-designed impact evaluation can help managers, planners, and policy-makers to: *avoid* continued investment in programmes which are not achieving their objectives; *avoid* eliminating programmes which can achieve their objectives; *ensure* that benefits reach all sectors of the target population; *ensure* that programmes are implemented in the most efficient and cost-effective manner *maximizing* both the quantity and the quality of the services and benefits; and, *provide* a decision tool for selecting the best way to invest scarce development resources. Without a good impact evaluation, there is an increased risk of reaching wrong decisions on whether programmes should continue or be terminated, and how resources should be allocated.

Two different definitions of impact evaluation are widely used. The first, that can be called the *technical* or *statistical definition* defines impact evaluation as an evaluation that:

…assesses changes in the well-being of individuals, households, communities or firms that can be attributed to a particular project, programme, or policy. The central impact evaluation question is what would have happened to those receiving the intervention if they had not in fact received the programme. Since we cannot observe this group both with and without the intervention, the key challenge is to develop a counterfactual – that is, a group which is as similar as possible (in observable and unobservable dimensions) to those receiving the intervention. This comparison allows for the establishment of definitive causality – attributing observed changes in welfare to the programme, while removing confounding factors.

Source: World Bank PovertyNet website3

The second, that can be called the *substantive long-term effects* definition, is espoused by the Organization for Economic Co-operation and Development's Development Assistance Committee (OECD/DAC). This defines impact as:

positive and negative, primary and secondary, long-term effects produced by a development intervention, directly or indirectly, intended or unintended.

Source: OECD-DAC 2002, p. 24

---

3   For extensive coverage of the technical/statistical definition of IE and a review of the main quantitative analytical techniques, see the World Bank's Development Impact Evaluation Initiative Web site: http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20205985~menuPK:435951~pagePK:148956~piPK:216618~theSitePK:384329,00.html. For an overview of approaches used by IEG, see White (2006), and for a discussion of strategies for conducting IE (mainly at the *project* level) when working under budget, time, and data constraints see Bamberger (2006)

While the OECD/DAC definition does not require a particular methodology for conducting an impact evaluation, it does specify that impact evaluations only assess *long-term effects.* The *technical* definition requires a particular methodology (the use of a counterfactual, based on a pretest/post-test project/control group comparison, but does not specify a time horizon over which impacts should be measured, and does not specify the kinds of changes (outputs, outcomes or impacts) which can be assessed. This distinction between the two definitions has proved to be important as many evaluators argue that impacts can be estimated using a number of different methodologies (the *substantive* definition), whereas advocates of the *technical* definition argue that impacts can only be assessed using a limited number of strong impact evaluation designs and that randomized control trials should be used wherever possible. In this paper we will use a comprehensive definition of impact evaluation that encompasses both the *technical* and *substantive* definitions.

Impact evaluation is only one of many types of evaluation that provide information to policy-makers, planners, and managers at different stages of a project or programme cycle[4]. Although many impacts cannot be fully assessed until an intervention has been operating for several years, planners and policy-makers cannot wait three or five years before receiving feedback. Consequently, many impact evaluations are combined with formative or process evaluations, while the project is still being implemented, to provide preliminary indications of whether a programme is on track to achieve its intended outcomes and impacts.

There is no one design that fits all impact evaluations. The best design for a particular evaluation will depend on: what is being evaluated (a small project, a large programme, or a nation-wide policy); the purpose of the evaluation; budget, time, and data constraints; and, the time horizon (is the evaluation designed to measure medium and long-term impacts or to make initial estimates of potential future impacts at the time of the mid-term review or the implementation completion report?). Impact evaluation designs can also be classified according to their level of statistical rigour. The most rigorous designs, from the statistical point of view, are the experimental designs commonly known as randomized control trials. These are followed, in descending order of statistical rigour,

---

4   Bamberger 2009a lists 7 different types of evaluations that are commonly conducted at the project planning stage, 9 during project implementation, 5 at the time of project completion and after the project is completed and this list is not exhaustive.

by: strong quasi-experimental designs that use pre-test/post-test control group designs; and, weaker quasi-experimental designs where baseline data has not been collected on either or both of the project and control groups[5].

The least statistically rigorous are the non-experimental designs which do not include a control group and, which may also not include baseline data on the project group. According to the *technical* definition of impact evaluation, the non-experimental designs should not be considered as impact evaluation because they do not include a counterfactual (control group). However, according to the *substantive* definition these can be considered impact evaluation when they are used to assess the long-term project outcomes and impacts. A critical factor in determining the methodological soundness of non-experimental designs is the adequacy of the approach proposed, to examine causality in the absence of a conventional counterfactual (Scriven 2009).

Advocates of the *technical* definition of an impact evaluation often claim that randomized control trials and strong quasi-experimental designs are the "best" and "strongest" designs (some use the term the "gold standard"). However, these designs should only be considered as the "strongest", in an important but narrow statistical sense, as their strength lies in their ability to eliminate, or control for, selection bias. While this is an important advantage, critics point out that these designs are not necessarily stronger than other designs with respect to other criteria, such as: construct validity; the validity and reliability of indicators of outcomes and impacts; and, their ability to collect information on sensitive topics and to identify and interview difficult to reach groups. When used in isolation these designs also ignore the process of project implementation and the economic, political, organizational and cultural context in which each project is implemented.

Policy-makers and planners should also keep in mind that there are relatively few situations in which the most rigorous evaluation designs can be used.[6] While there is an extensive evaluation litera-

---

5   See Bamberger and Rugh 2009 (the previous volume in this series) for a discussion of the different IE designs.

6   Although it is difficult to find statistics, based on discussions with development evaluation experts, this report estimates that randomized control trials have been used in only 1–2 percent of IEs; that strong quasi-experimental designs are used in less than 10 percent, probably not more than 25 percent include baseline surveys, and at least 50 percent and perhaps as high as 75 percent do not use any systematic baseline data.

ture on the small number of cases where strong designs have been used, much less guidance is available on how to strengthen the methodological rigour of the majority of impact evaluations that are forced by budget, time, data, or political constraints, to use methodologically weaker designs[7].

Deciding when an impact evaluation is needed and how much it should cost

An impact evaluation may be justified when decisions have to be made about: the continuation, expansion or replication of a programme; the potential development contribution of an innovative programme; the cost-effectiveness of a particular implementation strategy; or, the contribution of a particular donor to a multi-donor programme. It is important to be aware that there are many situations in which an impact evaluation is not the right choice, and where a different evaluation design is more appropriate to address the questions of interest to stakeholders.

The size and complexity of the programme, and the type of information required by policy-makers and managers, will determine whether a rigorous and expensive impact evaluation is required, or whether it will be sufficient to use a simpler and less expensive design. Also, budget and time-constraints often mean that a simpler and less rigorous design may be the only option available. There is no simple rule for deciding how much an evaluation should cost and when a rigorous and expensive impact evaluation may be justified. However, an important factor to consider is whether the benefits of the evaluation (for example, money saved by making a correct decision or avoiding an incorrect one), are likely to exceed the costs of conducting the evaluation. An expensive impact evaluation that produces important improvements in programme performance can be highly cost-effective. Even minor improvements in a large-scale programme may produce significant savings.

---

7   See Bamberger (2006a), and Bamberger, Rugh & Mabry (2006) for a discussion of strategies for enhancing the quality of impact evaluations conducted under budget, time and data constraints.

Institutionalizing impact evaluation

Institutionalization of impact evaluation at the sector or national level occurs when:

- the evaluation process is country-led and managed by a central government ministry or a major sectoral agency;

- there is strong "buy-in" to the process from key stakeholders;

- there are well-defined procedures for selecting, implementing and using impact evaluations ;

- impact evaluation is integrated into national monitoring and evaluation systems that generate much of the data used in the impact evaluations ;

- impact evaluation is integrated into national budget formulation and development planning;

- there is a focus on evaluation capacity development (ECD).

Institutionalization is a process, and at any given point it is likely to have advanced further in some areas, or sectors, than in others. The way in which impact evaluation is institutionalized will also vary from country to country, reflecting different political and administrative systems and traditions, and historical factors such as strong donor support for programmes and research in particular sectors.

Impact evaluation have not been able to achieve their potential contributions to programme management, budget planning and policy-making because the evaluations were selected and funded in a somewhat ad hoc and opportunistic way, determined by the interests of donor agencies or individual ministries, rather than by national planning priorities. The value of impact evaluation as a policy-making tool can be greatly enhanced once the selection, dissemination and use of the evaluations becomes part of a national impact evaluation system. This requires an annual plan for selection of the government's priority programmes on which important decisions have to be made concerning continuation, modification, or termination, and where the evaluation framework permits the comparison of alternative interventions – in terms of cost-effectiveness and contribution to national development goals.

There is no single strategy that has always proved successful in the institutionalization of impact evaluation. Countries that have made progress have built on existing evaluation experience; political and administrative traditions; and, the interest and capacity of individual ministries, national evaluation champions, or donor agencies. Although some countries (particularly Chile), have pursued a national monitoring and evaluation strategy that has evolved over a period of more than 30 years, most countries have responded in an ad hoc manner, as opportunities have presented themselves.

Figure 1 identifies three alternative pathways to the institutionalization of impact evaluation.

The first pathway (the ad hoc or opportunistic approach), evolves from individual evaluations which were commissioned to take advantage of available funds or from the interest of a senior government official, or a particular donor. Often evaluations were undertaken in different sectors, and the approaches were gradually systematized as experience was gained in selection criteria, effective methodologies, and how to achieve both quality and utilization. A central government agency (usually finance or planning), is either involved from the beginning, or becomes involved as the focus moves toward a national system. Colombia's national monitoring and evaluation system, SINERGIA, is an example of this pathway (Box 2.)

Figure 1: Three pathways for the evolution of institutionalized IE systems*

IE starts through ad hoc studies	IE starts in particular sectors	IE starts at whole-of-government level
Ad hoc opportunistic studies often with strong donor input	Sector management information systems	Whole-of-government M&E system
Increased involvement of national government	Larger-scale, more systematic sector evaluations	Incorporation of government-wide performance indicators
Systematization of evaluation selection and design procedures	Focus on evaluation capacity development and use of evaluations	National system with ad hoc, supply-driven selection and design of IE
	Increased involvement of academic and civil society	Standardized procedures for selection and implementation of IE studies
	Whole-of government standardized IE system	Standardized procedures for dissemination, review and use of IE findings
Examples: Colombia – SINERGIA Ministry of Planning	Examples: • Mexico – PROGRESA Conditional cash transfers • Uganda – Education for All • China Rural-based poverty-reduction strategies	Examples: Chile – Ministry of Finance

*Note: The institutionalized systems may employ either or both the *technical* or the *substantive* definitions of IE. The systems also vary in terms of the level of methodological rigor that they use.
Source: Bamberger 2009a

Box 2. Colombia: Moving from the Ad Hoc commissioning of Impact Evaluations by the Ministry of Planning and sector ministries toward integrating Impact Evaluations into the national M&E system (SINERGIA)

In Colombia the Ministry of Planning is responsible for managing the National System for Evaluation of Public Sector Performance (SINERGIA). The most visible and heavily utilized component is the sub-system for monitoring progress against a total of 320 country development and presidential goals.

Although IE was initiated in 1999, these goals have evolved since 2000 to be commissioned and managed from SINERGIA for a wide range of priority government programmes. To date, SINERGIA has played a major role in the selection of the programmes to be evaluated. Initially it was a somewhat ad hoc process – partly determined by the interest of international funding agencies. As the programme of IE evolved, the range of methodologies was broadened and technical criteria in the selection of programmes to be evaluated were formalized through policy documents (with more demand-side involvement from the agencies managing the programmes being evaluated) and in how the findings are used. Most of the IEs carried out use rigorous econometric evaluation techniques. A World Bank loan is supporting the strengthening of the system with specific activities aiming to further institutionalize IE.

Source: Mackay (2007, pp. 31–36), Bamberger 2009a.

The second pathway is where impact evaluation expertise is developed in a priority sector supported by a dynamic government agency and with one or more champions, and where there are important policy questions to be addressed and strong donor support. Once the operational and policy value of these evaluations has been demonstrated, this becomes a catalyst for developing a national system. The evaluations of the PROGRESA conditional cash transfer pro-

grammes in Mexico are an example of this approach (Box 3).

Box 3. Mexico: Moving from an evaluation system developed in one sector toward a national evaluation system (SEDESOL)

In Mexico a series of rigorous evaluations of the Progresa Conditional Cash Transfer programmes were conducted over a number of years. The evaluations convincingly demonstrated the effectiveness of conditional cash transfers as a way to improve the welfare (particularly education and health) of large numbers of low-income families. The evaluations are considered to have been a major contributing factor in convincing the new government that came to power in 2002, to continue these programmes, which had been started by the previous administration. The evaluations also served to convince policymakers of the technical feasibility and policy value of rigorous IEs and contributed to the passing of a law by Congress in 2007 mandating the evaluation of all social programmes. This law also created the National Commission for the Evaluation of Social Programmes, which was assigned the responsibility for regulating the development of monitoring and evaluation functions in the social sectors. A similar continuity was achieved in Colombia, where progress is also being made toward a national M&E system (see Box 2).

Source: Mackay (2007, p. 56), Bamberger 2008.

Impact evaluation can evolve as a component built into an existing ministry or sector-wide monitoring and evaluation system, or as part of a new monitoring and evaluation system being developed under a project or programme loan funded by one or more donor agencies. This latter approach has not proved an effective way to develop an impact evaluation system, both because much of the data for the impact evaluation studies must be generated by the monitoring and evaluation system that is still in the process of development, and because the system is time bound, with funding ending at the closing of the project loan, which is much too early to assess impacts.

The evaluations of the national *Education for All* programme in Uganda offer a second example of the sector pathway.[8] These evaluations increased interest in the existing national monitoring and evaluation system (the National Integrated Monitoring and Evaluation System, or NIMES), and encouraged agencies to upgrade the quality of the information they submit. The *World*

---

8   A video of a presentation on this evaluation made during the World Bank January 2008 Conference on "Making smart policy: using impact evaluation for policy-making" is available at: http://info.worldbank.org/etools/BSPAN/PresentationView. asp?PID=2257&EID=1006

Bank Africa Impact Evaluation Initiative (AIM) is a broader regional initiative, designed to help governments strengthen their overall monitoring and evaluation capability and systems. AIM is currently supporting some 90 experimental and quasi-experimental impact evaluations in 20 African countries in the areas of education, HIV, malaria, and community-driven development[9]. Similarly, at least 40 countries in Asia, Latin America, and the Middle East are implementing sectoral approaches to impact evaluation with World Bank support. Similar initiatives are being promoted through recently created international collaborative organizations, such as the *Network of Networks for Impact Evaluation* (NONIE),[10] and the *International Initiative for Impact Evaluation* (3IE).[11]

According to Ravallion (2008), China provides a dramatic example of the large-scale and systematic institutionalization, over more than a decade, of impact evaluation as a policy instrument for testing and evaluating potential rural-based poverty-reduction strategies. In 1978 the Communist Party's 11th Congress adopted a more pragmatic approach whereby public action was based on demonstrable success in actual policy experiments on the ground:

> *A newly created research group did field work studying local experiments on the de-collectivization of farming using contracts with individual farmers. This helped convince skeptical policy makers … of the merits of scaling up the local initiatives. The rural reforms that were then implemented nationally helped achieve probably the most dramatic reduction in the extent of poverty the world has yet seen (Ravallion 2008, p. 2).*

The third pathway is where a planned and integrated series of impact evaluations were developed from the start as one component of a whole-of-government system, managed and championed by a strong central government agency, usually the Ministry of Finance or Planning. Chile is a good example of a national monitoring and evaluation system in which there are clearly defined criteria and guidelines for the selection of programmes to be evaluated, their conduct and methodology, and how the findings will be used (Box 4).

---

9   For more information on the Africa Impact Evaluation Initiative see: http://worldbank.org/afr/impact

10   http://www.worldbank.org/ieg/nonie/index.html

11   http://www.3ieimpact.org

Impact evaluations often begin in a somewhat ad hoc and opportunistic way, taking advantage of the interest of key stakeholders and available funding opportunities. The challenge is to build on these experiences to develop capacity to select, conduct, disseminate, and use evaluations. Learning mechanisms, such as debriefings and workshops, can also be a useful way to streamline and standardize procedures at each stage of the impact evaluation process. It is helpful to develop an impact evaluation handbook for agency staff summarizing the procedures, identifying the key decision points, and presenting methodological options (DFID 2005). Box 5 lists steps to institutionalize an impact evaluation system.

> ### Box 4. Chile: Rigorous impact evaluation introduced as part of an integrated Whole-of-Government M&E System
>
> The government of Chile has developed over the past 14 years a whole-of-government M&E system with the objective of improving the quality of public spending. Starting in 1994, a system of performance indicators was developed; rapid evaluations of government programmes were incorporated in 1996; and, in 2001 a programme of rigorous impact evaluations was incorporated. There are two clearly defined impact evaluation products. The first are rapid ex-post evaluations that follow a clearly defined and rapid commissioning process, where the evaluation has to be completed in less than 6 months for consideration by the ministry of finance as part of the annual budget process. The second are more comprehensive evaluations that can take up to 18 months and cost $88,000 on average.
>
> The strength of the system is that it has clearly defined and cost-effective procedures for commissioning, conducting, and reporting of impact evaluations, a clearly defined audience (the Ministry of Finance), and a clearly understood use (the preparation of the annual budget). The disadvantages are that the focus of the studies is quite narrow (only covering issues of interest to the Ministry of Finance) and the involvement and buy-in from the agency being implemented is typically low. Some have also suggested that there may be a need to incorporate some broader and methodologically more rigorous impact evaluations of priority government programmes (similar to the PROGRESA evaluations in Mexico).

Source: Mackay (2007, pp. 25–30). Bamberger (2008)

Integrating impact evaluation into sector and/or national monitoring and evaluation and other data-collection systems

The successful institutionalization of impact evaluation will largely depend on how well the selection, implementation, and use of impact evaluation are integrated into sector and national monitoring and evaluation systems and national data-collection programmes. This is critical for several reasons. First, much of the data required for an impact evaluation can be obtained most efficiently and economically from sector and/or national monitoring and evaluation systems. This includes information on: how programme beneficiaries were selected and how these criteria may have changed over time; how the programme is being implemented; how closely this conforms to the implementation plan; whether all beneficiaries receive the same package of services and of the same quality; the proportion of people who drop out, the reasons for this, and how their characteristics compare with people who remained in the programme; and, how programme outputs compare with the original plan.

Second, impact evaluation findings that are widely disseminated and used to provide an incentive for agencies to improve the quality of monitoring and evaluation data which they collect and report. For example, Ministry of Education staff in Uganda reported that the wide dissemination of the evaluations of the *Education for All* programme made them aware of the importance of carefully collected monitoring data. As a consequence the quality of monitoring reporting improved significantly.

Box 5. Key steps for institutionalizing impact evaluation at the national and sector levels

1. Conduct an initial diagnostic study to understand the context in which the evaluations will be conducted

2. The diagnostic study should take account of local capacity, where this is lacking, define what capacities are required and how they can be developed

3. A key consideration is whether a particular impact evaluation will be a single evaluation, that probably will not be repeated, or whether there will be a continuing demand for such impact evaluations

4. Define the appropriate option for planning, conducting and/or managing the impact evaluation, such as:

   Option 1: Studies are conducted by the agency responsible for implementing the programme being evaluated.

   Option 2: Studies are planned, conducted or managed by a central government agency.

   Option 3: Studies are managed by the sector agency but subcontracted to local or international consultants.

   Option 4: The primary responsibility will rest with the donor agencies.

5. Define a set of standard and transparent criteria for the selection of the impact evaluation to be commissioned each year.

6. Define guidelines for the cost of an impact evaluation and how many impact evaluations should be funded each year.

7. Clarify who will define and manage the impact evaluation

8. Define where responsibility for impact evaluation is located within the organization and ensure that this unit has the necessary authority, resources, and capacity to manage the impact evaluations.

9. Conduct a stakeholder analysis to identify key stakeholders and to understand their interest in the evaluation and how they might become involved.

10. Ensure that users continue to be closely involved throughout the process.

11. Develop strategies to ensure effective dissemination and use of the evaluations.

12. Develop an impact evaluation handbook to guide staff through all stages of the process of an impact evaluation: identifying the programme or project to be evaluated, commissioning, contracting, designing, implementing, disseminating, and using the impact evaluation findings.

13. Develop a list of prequalified consulting firms and consultants eligible to bid on requests for proposals.

Third, access to monitoring data makes it possible for the impact evaluation team to provide periodic feedback to managers and policy makers on interim findings that could not be generated directly from the impact evaluation database. This demonstrates the practical utility of impact evaluation and overcomes one of the major criticisms that clients express about IE – namely the long delay before any results are available.

Fourth, national household survey programmes, such as household income and expenditure surveys, and demographic and health surveys, provide very valuable sources of secondary data for strengthening methodological rigour of impact evaluation design and analysis (for example, the use of propensity score matching to reduce sample selection bias). Some of the more rigorous impact evaluations have used cooperative arrangements with national statistical offices to piggy-back information required for the impact evaluation onto an ongoing household survey, or to add a special module. Although piggy-backing can save money, the required coordination can be very time-consuming.

Creating demand for impact evaluations[12]

Efforts to strengthen the governance of impact evaluation and other kinds of monitoring and evaluation systems are often viewed as technical fixes – mainly involving better data systems and the conduct of good quality evaluations (Mackay, 2007). Although the creation of evaluation capacity needed to provide high-quality evaluation services and reports is important, these *supply-side interventions* will have little effect unless there is sufficient *demand* for quality impact evaluation. Demand for impact evaluation requires that quality impact evaluations are seen as an important policy and management tool in one or more of the following areas:

(a) assisting resource-allocation decisions in the budget and planning process;

(b) to help ministries in their policy formulation and analytical work;

(c) to aid ongoing management and delivery of government services; and

(d) to underpin accountability relationships.

---

12   This section adapts the discussion by Mackay (2007) on how to create broad demand for M&E to the specific consideration here of how to create demand for IE.

Creating demand requires sufficiently powerful incentives within a government to conduct impact evaluation, to create a good level of quality, and to use impact evaluation information intensively. A key factor is to have a public sector environment supportive of the use of evaluation findings as a policy and management tool. If the environment is not supportive or is even hostile to evaluations, raising awareness of the benefits of impact evaluation and the availability of evaluation expertise might not be sufficient to encourage managers to use these resources. To help generate demand for impact evaluation among potential users, MacKay (2007, Table 11.1),[13] suggests some *positive incentives* ("carrots"), such as building ownership and providing funds to build better impact evaluation systems; *sanctions and threats* ("sticks"), such as laws and decrees mandating the planning, conduct and reporting of impact evaluation; and, *positive messages from key figures* ("sermons"), such as high-level statements of endorsement and conferences on good impact evaluation practice.

For several reasons, the incentives are often more difficult to apply to impact evaluation than to promoting general monitoring and evaluation systems. First, impact evaluations are only conducted on selected programmes and at specific points in time; consequently, incentives must be designed to encourage use of impact evaluation in appropriate circumstances but not to encourage its overuse. For example, overuse might be where an impact evaluation would be premature or where similar programmes have already been subject to an impact evaluation. Second, as flexibility is required in the choice of impact evaluation designs, it is not meaningful to propose standard guidelines and approaches, as can often be done for monitoring and evaluation. Finally, many impact evaluations are contracted to consultants so agency staff involvement (and consequently their buy-in), is often more limited.

It is important to involve major national universities and research institutions. In addition to tapping this source of national evaluation expertise, universities (through teaching, conferences, research, and consulting), can play a crucial role in raising awareness of the value and uses of impact evaluation. Part of the broad-based support for the PROGRESA programmes and their research agenda was because they made their data and analysis available, on the Internet, to national and international researchers. This created a

---

13   MacKay's table referred to incentives to generate demand for M&E systems and this was adapted by Bamberger (2009a) to refer specifically to IE.

demand for further research and refined and legitimized the sophisticated methodologies used in the PROGRESA evaluations. Both Mexico's PROGRESA and Colombia's *Familias en Accion* recognized the importance of dissemination (through high-profile conferences, publications, and working with the mass media) in demonstrating the value of evaluation and creating future demand.

Capacity development for impact evaluation

In recent years there has been a significant awareness on the part of ODA agencies of the importance of evaluation capacity development (ECD). For example, the World Bank's Independent Evaluation Group has an ECD website listing World Bank publications and studies on ECD; OECD-DAC has organized a network providing resources and links on ECD and related resources[14]. Many other development agencies also offer similar resources. A recent series of publications by UNICEF presents the broader context within which evaluation capacity must be developed at the country and regional levels. The series documents recent progress – including the growth of national and regional evaluation organizations[15].

The increased attention to ECD is largely due to the recognition that past efforts, for example, to strengthen statistical capacity and national statistical databases, were over-ambitious and had disappointing results. This was often because they focused too narrowly on technical issues without understanding the institutional, and other resource constraints faced, by many countries. Drawing on this earlier experience, ODA agencies have recognized that the successful institutionalization of impact evaluation will require an ECD plan to strengthen the capacity of key stakeholders to fund; commission; design; conduct; disseminate; and, use impact evaluation. On the supply side this involves strengthening the *supply* of resource persons and agencies able to deliver high-quality and operationally relevant impact evaluations and, developing

---

14 http://www.capacity.org/en/resource_corners/learning/useful_links/oecd_dac_network_on_development_evaluatio

15 Segone and Ocampo (editors) 2006. *Creating and Developing Evaluation Organizations: Lessons from Africa, Americas, Asia, Australasia and Europe. IOCE;* Segone (editor) 2008. *Bridging the gap: the role of monitoring and evaluation in evidence-based policymaking,* UNICEF; Segone (editor) 2009. *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results*, UNICEF.

the infrastructure for generating secondary data to complement, or replace, expensive primary data collection. This requires the periodic generation of census, survey, and programme-monitoring data that can be used for constructing baseline data and information on the processes of programme implementation.

Some of the skills and knowledge can be imparted during formal training programmes, but others must be developed over time through gradual changes in the way government programmes and policies are formulated, implemented, assessed, and modified. Many of the most important changes will only occur when managers and staff at all levels come to understand that: impact evaluation can be helpful rather than threatening; that it can improve the quality of programmes and projects; and, that it can be introduced without introducing an excessive burden of work.

An effective capacity-building strategy must target at least five main stakeholder groups: agencies which commission, fund, and disseminate impact evaluations; evaluation practitioners who design, implement, and analyze impact evaluations; evaluation users; groups affected by the programmes being evaluated; and, public opinion. Users include government ministries and agencies which use evaluation results to help formulate policies, allocate resources, and design and implement programmes and projects[16].

The active involvement of leading national universities and research institutions is also critical for capacity development. These institutions can mobilize the leading national researchers (and also have their own networks of international consultants), and they have the resources and incentives to work on refining existing, and developing new, research methodologies. Through their teaching, publications, conferences, and consulting, they can also strengthen the capacity of policy makers to identify the need for evaluation and to commission, disseminate, and use findings. Universities, NGOs, and other civil society organizations can also become involved in action research.

An important role of ECD is to help ministries and executing agencies design "evaluation-ready" programmes and policies. Many programmes generate monitoring and other forms of administrative data which could be used to complement or replace the collection of survey data, or to provide proxy baseline data where an evalu-

---

16   For a fuller discussion of evaluation capacity development see Bamberger (2006) in the earlier UNICEF publication in this series.

ation started too late to have been able to conduct baseline studies. Often, however, the data is not collected or archived in a way which makes it easy to use for evaluation purposes. Closer cooperation between the programme staff and the evaluators can greatly enhance the utility of project data for the evaluation. In other cases, slight changes in how a project is designed or implemented could strengthen the evaluation design. For example, there are many cases where a randomized control trial design could have been used, but the evaluators were not involved until it was too late.

There are a number of different formal and less-structured ways evaluation capacity can be developed, and an effective ECD programme will normally involve a combination of several approaches. These include formal university or training institute programmes ranging from one or more academic semesters to seminars lasting several days or weeks; workshops lasting from a half day to one week; distance learning and online programmes; mentoring; on-the-job training, where evaluation skills are learned as part of a package of work skills; and as part of a community development or community empowerment programme.

Promoting the utilization of impact evaluation

Despite the significant resources devoted to programme evaluation, there is widespread concern that, even for evaluations which are methodologically sound, the utilization of evaluation findings is disappointingly limited (Bamberger, Mackay, and Ooi, 2004). The barriers to evaluation utilization also affect institutionalization, and overcoming the former will contribute to the latter. There are a number of reasons why evaluation findings are underutilized and why the process of impact evaluation is not institutionalized:

- lack of ownership;
- lack of understanding of the purpose and benefits of impact evaluation ;
- bad timing;
- lack of flexibility and responsiveness to the information needs of stakeholders;
- wrong question and irrelevant findings;
- weak methodology;

- cost and number of demands on programme staff;

- lack of local expertise to conduct, review, and use evaluations;

- communication problems;

- factors external to the evaluation;

- lack of a supportive organizational environment.

There are additional problems in promoting the use of impact evaluation. Impact evaluation will often not produce results for several years, making it difficult to maintain the interest of politicians and policy makers who operate with much shorter time-horizons. There is also a danger that key decisions on future programme and policy directions will already have been made before the evaluation results are available. In addition, many impact evaluation designs are quite technical and difficult to understand.

Programme evaluations can be influential in many different ways, not all of which are intended by the evaluator or the client. Some of the different kinds of influence that impact evaluations can have include[17]:

- alerting public service agencies to problems of which they were not fully aware or had not considered important;

- providing objective, quantitative data that civil society can use to pressure agencies to improve services;

- alerting technical agencies to the need to incorporate vulnerable groups and to address community conflicts caused by limited access to services;

- providing political cover to government to take a politically sensitive decision, and showing how to mitigate negative consequences for influential "losers" from the policy change;

- developing methodologies to systematically document problems in the use of public funds that were widely suspected but that have not been possible to document;

- providing strong empirical evidence to convince new administrations to continue major programmes which were initiated by the previous administration.

17   For a fuller discussion of these examples see Bamberger (2008) in the earlier UNICEF publication in this series

Bamberger (2008) suggests various ways to strengthen the utilization of the findings of impact evaluations:

- *Understanding the political context* within which the evaluation is conducted, disseminated and used.

- *Timing of the launch and completion of the evaluation.* Many well-designed evaluations fail to achieve their intended influence because they were completed either too late or too early (before the questions being addressed are on the policy-makers' radar screen).

- *Deciding what to evaluate.* A successful evaluation will focus on a limited number of critical issues and hypotheses based on a clear understanding of the clients' information needs and of how the evaluation findings will be used.

- *Basing the evaluation on a programme theory (logic) model* so that the evaluator shares a common understanding with clients and stakeholders on the problem the programme is addressing, what its objectives are, how they will be achieved, and what criteria the clients will use in assessing success.

- *Creating ownership of the evaluation.* One of the key determinants of evaluation utilization is the extent to which clients and stakeholders are involved throughout the evaluation process. Do clients feel that they "own" the evaluation, or do they not know what the evaluation will produce until they receive the final report? Many clients have strong preferences for or against particular evaluation methodologies, and one of the factors contributing to underutilization of an evaluation may be client disagreement with, or lack of understanding of, the evaluation methodology.

- *Combining impact evaluation with process analysis and formative evaluation*[18]. Even when the primary objective of the evaluation is to assess impacts, it is important to help the client understand how outcomes and impacts are affected by what happens during programme implementation. Process and formative evaluation also provide rapid feedback to the client and increase the sense of ownership and the feeling that the evaluation is useful.

---

18   An evaluation intended to furnish information for guiding programme improvement is called a formative evaluation (Scriven 1991) because its purpose is to help form or shape the programme to perform better (Rossi, Lipsey, and Freeman, 2004, p. 34).

- *Evaluation capacity development* is an essential tool to promote utilization because it not only builds skills, but it also promotes evaluation awareness.

- *Develop a follow-up action plan* to monitor whether agreed decisions are implemented.

Many evaluations have little impact because the findings are not communicated to potential users in a way that they find useful or comprehensible. The following are some guidelines for communicating evaluation findings to enhance utilization:

- clarify what each user wants to know and the amount of detail required.

- understand how different users like to receive information. In a written report? In a group meeting with slide presentation? In an informal, personal briefing?

- do users want statistics or do they prefer photos and narrative? Do they want a global overview, or to understand how the programme affects individual people?

- be prepared to use different communication strategies for different users.

- pitch presentations at the right level of detail or technicality. Do not overwhelm managers with technical details, but do not insult professional audiences by implying that they could not understand the technicalities.

- define the preferred medium for presenting the findings. A written report is not the only way to communicate findings. Other options include oral presentations, videos, photographs, meetings with programme beneficiaries, and visits to programme locations.

- use the right language(s) for multilingual audiences.

Conclusions

ODA agencies are facing increasing demands to account for the effectiveness and impacts of the resources they have invested in development interventions. This has led to an increased interest in more systematic and rigorous evaluations of the outcomes and impacts of the projects, programmes, and policies these agencies support. A number of high-profile and methodologically rigor-

ous impact evaluations have been conducted in countries such as Mexico and Colombia, and many other countries are conducting impact evaluation of priority development programmes and policies – usually with support from international development agencies.

Though many of these evaluations have contributed to improving the programmes they have evaluated, much less progress has been made toward institutionalizing the processes of selection, design, implementation and use of impact evaluation. Consequently most impact evaluations have not achieved their full potential as instruments for budget planning and development policy formulation. Progress toward institutionalization of impact evaluation in a given country can be assessed in terms of six dimensions: (a) Are the studies country-led and managed? (b) Is there strong buy-in from key stakeholders? (c) Have well-defined procedures and methodologies been developed? (d) Are the evaluations integrated into sector and national monitoring and evaluation systems? (e) Is impact evaluation integrated into national budget formulation and development planning? and (f) Are there policies and programmes in place to develop evaluation capacity?

ODA agencies, who continue to provide major financial and technical support for impact evaluation, will continue to play a major role in promoting the institutionalization of impact evaluation. One major responsibility must be an active commitment to move from ad hoc support to individual impact evaluations that are of particular interest to the donor country, to a genuine commitment to helping countries develop an impact evaluation system that serves the interests of national policy-makers and line ministries. This requires a full commitment to the 2005 Paris Declaration on Aid Effectiveness, in particular, to support for country-led evaluation. It also requires a more substantial and systematic commitment to evaluation capacity development, and a recognition of the need to develop and support a wider range of impact evaluation methodologies, designed to respond more directly to country needs, and less on seeking to impose methodologically rigorous evaluation designs that are often of more interest to ODA research institutions than to developing country governments. This emphasizes the importance of recognizing the distinction between the technical and the substantive definitions of impact evaluation, and accepting that most country impact evaluation strategies should encompass both definitions in order to draw on a broader range of methodologies to address a wide range of policy and operational questions.

For their part, developing country governments must invest the necessary time and resources to ensure they fully understand the potential benefits and limitations of impact evaluation and the alternative methodologies that can be used. Usually one or more ODA agencies will be willing to assist governments wishing to acquire this understanding, but governments must seek their own independent advisors so that they do not depend exclusively on the advice of a particular donor who may have an interest in promoting only certain types of impact evaluation methodologies.

References

Bamberger, M (2008a). *Institutionalizing Impact Evaluation Within the Framework of a Monitoring and Evaluation System.* Independent Evaluation Group. Washington, D.C. The World Bank.

Bamberger, M (2006a). *Conducting Quality Impact Evaluations under Budget, Time, and Data Constraints*. Washington, DC. The World Bank.

Bamberger, M (2006b). *Evaluation Capacity Building*. In Creating and Developing Evaluation Organizations: Lessons learned from Africa, Americas, Asia, Australasia and Europe, ed. Segone and Ocampo, Lima, IOCE.

Bamberger, M., K. Mackay, and E. Ooi. (2004). *Influential Evaluations: Evaluations that Improved Performance and Impacts of Development Programmes.* Operations Evaluation Department. Washington, DC. The World Bank.

Bamberger, M., K. Mackay, and E. Ooi. (2005). *Influential Evaluations: Detailed Case Studies.* Operations Evaluation Department. Washington, DC. The World Bank.

Bamberger, M., J. Rugh, and L. Mabry. (2006). *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Sage Publications.

DFID (Department for International Development, UK). (2005). *Guidance on Evaluation and Review for DFID Staff.* http://www.dfid.gov.uk/aboutdfid/performance/files/guidance-evaluation.pdf.

Mackay, K (2007). *How to Build M&E Systems to Support Better Government.* Independent Evaluation Group. Washington, DC. The World Bank.

OECD-DAC (Organization of Economic Co-operation and Development, Development Advisory Committee). (2002). *Glossary of Key terms in Evaluation and Results Based Management.* Paris: OECD.

Ravallion, M. (2008). *Evaluation in the Service of Development*. Policy Research Working Paper No. 4547, Washington, DC, The World Bank.

Rossi, P., M. Lipsey, and H. Freeman. (2004). *Evaluation: a Systematic Approach* (7th ed.). Thousand Oaks, CA: Sage Publications.

Scriven, M (2009). *Demythologizing causation and evidence*. In Donaldson, S; Christie, C; and Mark, M. What counts as credible evidence in applied research and evaluation practice? Sage Publications

Segone, M and Ocampo A.(editors) (2006). Creating *and Developing Evaluation Organizations: Lessons from Africa, Americas, Asia, Australasia and Europe, IOCE*.

Segone, M. (editor). (2008). *Bridging the gap: the role of monitoring and evaluation in evidence-based policymaking.* UNICEF.

Segone, M. (editor). (2009). *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results*. UNICEF.

White, H. (2006). *Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank*. Working Paper No. 38268, World Bank, Washington, DC.

EVALUATION AND DEVELOPMENT.
THE PARTNERSHIP DIMENSION

by Osvaldo Feinstein,
Professor at the Masters in Evaluation,
Universidad Complutense de Madrid (Spain) and
Senior Advisor, Spanish Evaluation Agency

Introduction

Although there have been references to partnerships in develop-
ment practice and discourse for several decades, it is particularly
since the mid-90's, and especially from 2000 onwards, that part-
nerships have become an important dimension of development and
evaluation[1].

Partnerships for development

For development, partnerships have been considered as a means to
promote ownership of development projects, programmes and poli-
cies. There has been a growing consensus that country ownership
is crucial for achieving development results and that partnerships
promote country ownership, particularly country-led partnerships,
for which the *Paris Declaration*, and its ratification in Accra, have
provided an enabling, authorizing environment[2]. This is shown sche-
matically in figure 1.

**Figure 1: Partnerships as one of the key factors leading
to development results**

Partnerships ⟶ Ownership ⟶ Development results

---

[1]   On partnerships in development, see Brinkerhoff (2002); for partnerships and
evaluation, see Liebenthal, Feinstein and Ingram (2004). See also Morra and Rist
(2009)

[2]   See Picciotto (2007), Feinstein (2009) and Segone (2009)

Partnerships are also considered fundamental for scaling-up, which is, moving from pilot projects or experiences to country wide interventions.[3] For example, in middle-income countries, donors and development agencies have realized that they need to establish partnerships with national and sub national governments. This is because in middle-income countries governments, at the national or sub-national levels, can provide the resources and leadership needed to scale-up interventions that have proven to be successful at a small-scale.

Partnerships in evaluation

In the evaluation field, partnerships have been pursued in at least two different ways:

(1) *to complement resources among development evaluation entities*, combining for example the knowledge, experience and/ or reputation of some organizations, with the financial resources of other institutions. This has enabled them to reach a threshold level suitable to conduct appropriate evaluations;

(2) *to conduct joint evaluations, establishing "partnerships for evaluation"*[4] by several partners pooling financial resources and expertise. There have been several instances of this approach.

It is worth mentioning that earlier work on "partners in evaluation" referred to the evaluation of development and community programmes involving participants in those interventions, i.e. to "participatory evaluation". Although this is an important way in which partners are (or can be) involved in evaluation, it is not the focus of this chapter.[5]

Nevertheless, it is useful to note that there is a parallel field of thought and action on "strategic alliances", defined as cooperative agreements between two or more organizations. This work originated mainly in business and private sector management (and also to some extent in international politics). It is very relevant for partnerships in development and evaluation. A recent and valuable review of this literature is provided by Serrat (2009).

---

3   See von Braun, Vargas & Pandya (2009)

4   See. Morra and Rist (2009) and Feinstein (2006)

5   See the pioneer and still useful book by Feuerstein (1986)

Evaluating partnerships

With the growing importance of partnerships, evaluators have faced the challenge of evaluating partnerships and in designing them in the best possible way. It has become clear that partnerships are a modality of collective action, and that there are some general and specific principles which can be applied, in their design and in their evaluation, so as to optimize their effectiveness, efficacy and efficiency.[6]

A crucial assumption concerning the *effectiveness and efficacy* of partnerships, is that partnerships enhance the effectiveness and efficacy of partners. In other words, that there is a potential value added of partnerships in that they may allow partners to achieve results that would not be achievable if they were to act in isolation (or at least that their achievement would be considerably less than through their involvement in a partnership). Evaluations should verify the extent to which this assumed enhanced effectiveness and efficacy, is (or is not) generated by partnerships, and identify the reasons why partnerships yielded (or failed to yield) expected results. Also, evaluations should pay attention to possible unintended results, positive and/or negative, generated by partnerships.

The *efficiency* of partnerships in development has become an important issue. Frequently it has been taken for granted that partnerships necessarily entail an additional cost in comparison to a non-partnership modus operandi. This incremental cost is sometimes justified in terms of the expected value added generated by the partnership. However, there is a remarkable scarcity of data concerning the costs (and benefits) of partnerships. A promising approach to assessing the efficiency of partnerships is through the application of transaction costs for the partners. There are some conceptual papers applying this approach but very few empirical analyses[7]. Given the opportunities for economies of scale to be seized through partnerships (for example, sharing the costs of surveys in the case of partnerships for evaluations), the efficiency of a partnership approach vis-à-vis a non-partnership approach should be explored with an open mind. Being open to the possibility of efficiency gains which can be achieved through partnerships is important, especially since this is opposite to the common view that partnerships will inevitably raise the costs of "doing business" in

---

6   Examples can be found in Liebenthal, Feinstein and Ingram (2004)

7   See for example Jobin (2008)

the development field. In fact, the different ways in which partnerships may operate should be considered. A challenge for those working in evaluation and development is to introduce innovations which may allow for better functioning partnerships, without ruling out the possibility of achieving "more with less" through the introduction of appropriate innovations in how best to engage in partnerships. The following figure, which distills lessons from experience and from the literature on partnerships[8] in evaluation and development, may be useful in developing innovative partnerships and in summarizing the preceding discussion:

Figure 2: Suggestions for effective and efficient partnerships

1. Know your partners.
2. Do not assume that partnerships costs are inevitably high.
3. Search for low-cost modes of partnership.
4. Know the terms of the partnership agreement (the rules of the game).
5. Keep your partners informed.
6. Design partnerships taking into account lessons learned from evaluations.
7. Evaluate your partnership jointly with your partners.

In the previous paragraphs, no reference has been made to the *sustainability* of partnerships. This is because what matters is the sustainability of the results, made possible by partnerships, rather than the sustainability of the partnerships. In fact, it is generally considered good practice to incorporate some kind of "exit strategy" in partnerships, which implies the dissolution of the partnerships after a certain period during which either the results are expected to be achieved, or the mechanisms assumed to generate the expected results would be established.

An important type of development partnerships, for which progress has been made in elaborating good practice standards for their design and evaluation, are the Global and Regional Partnership Programmes (GRPP), dealing with global or regional issues. The first of these programmes was *The Consultative Group for International Agricultural Research* (CGIAR). Much more recently other global programmes have been launched in several areas, particularly

---

8   See also the very recent review of this literature presented in Horton, Prain & Thiele (2009)

in environment, health, water and transportation.[9] *The Independent Evaluation Group* (IEG), of the World Bank, is playing a leading role in developing principles and standards for evaluating GRPPs under the auspices of the OECD/DAC *Network on Development Evaluation*. A free-standing document, *The Sourcebook for Evaluating Global and Regional Partnership Programmes,*[10] builds on principles and standards for evaluating development assistance which have been developed by the OECD/DAC Evaluation Network, the United Nations Evaluation Group, the Evaluation Cooperation Group of the Multilateral Development Banks, evaluation associations, and others. The sourcebook also draws on IEG's experience in reviewing GRPPs as well as the feedback received at the *Stakeholder Consultative Workshop* held for this purpose in Paris in September 2006. In addition, IEG is in the process of preparing good practice guidelines and examples for global and regional partnership programmes. (For information on the ongoing work in this frontier area of the partnership dimension in evaluation and development, see http://www.oecd.org/dataoecd/28/13/42926578.pdf).

Conclusions

Partnerships are an important dimension of evaluation and development. They promote ownership of development interventions, which is a crucial factor for development effectiveness. Partnerships should be evaluated, identifying their value added and crucial success factors, as well as the reasons why they may have eventually failed to yield expected results in given contexts. There is a wealth of experience in development partnerships, and in their evaluation, which has been referred to in this chapter and that can be drawn upon so as to enhance the contribution to development effectiveness of partnerships in evaluation and development.

---

9   See Liebenthal, Feinstein & Ingram (2004) and Morra & Rist (2009).

10   The Sourcebook is available at http://web.worldbank.org/WBSITE/EXTERNAL/ EXTOED/EXTGLOREGPARPROG/0,,contentMDK:21178261~menuPK:4426473 ~pagePK:64829573~piPK:64829550~theSitePK:4426313,00.html This website also provides information on an evaluation community of practice dealing with the evaluation of GRPPs.

References

Brinkerhoff, Jennifer. (2002). *Partnership for international development,* Boulder: Rienner

Feuerstein, Marie-Thérèse. (1986). *Partners in evaluation* London: Macmillan

Feinstein, Osvaldo. (2006. *Evaluation of development interventions and humanitarian action*, in Ian Shaw et.al., eds., (2006): Evaluation Handbook London: Sage

Feinstein, Osvaldo. (2009). *Country-led evaluation systems*, in Marco Segone, ed. (2009): Country-led monitoring and evaluation systems, Geneva: UNICEF

Hartmann, Armtraud & Linn, Johannes F. (2009). *Scaling up. A path to effective development*, in von Braun, Joachim, Vargas Hill, Ruth & Pandya-Lorch, Rajul, eds., (2009) The poorest and hungry Washington D.C.: IFPRI

Horton, Douglas; Prain, Gordon & Thiele, Graham (2009). *Perspectives on Partnership: A literature review* Lima: International Potato Center CIP

Independent Evaluation Group/World Bank. (2009). *Evaluating Global and Regional Partnership Programs (GRPPs).* Progress report to the OECD/DAC Evaluation Network
http://www.oecd.org/dataoecd/28/13/42926578.pdf

Jobin, Denis (2008). *A Transaction Cost-Based Approach to Partnership Performance Evaluation,* Evaluation, Vol. 14, No. 4, 437-465

Liebenthal, Andres; Feinstein, Osvaldo & Ingram, Gregory K., Eds., (2004). *Evaluation and development* New Brunswick: Transaction Publishers

Morra Imas, Linda G. & Rist, Ray C. (2009). *The Road to Results,* Washington D.C.: The World Bank

Picciotto, Robert. (2007). *The new environment for development evaluation,* American Journal of Evaluation Vol.28, No.4

Segone, Marco, ed. (2009). *Country-led monitoring and evaluation systems*, Geneva: UNICEF

Serrat, Oliver. (2009). *Learning in strategic alliances,* Manila: Asian Development Bank, http://www.adb.org/Documents/Information/Knowledge-Solutions/Learning-in-Strategic-Alliances.pdf

EXPLORING EFFECTIVE STRATEGIES FOR FACILITATING EVALUATION CAPACITY BUILDING

by Hallie Preskill, Executive Director,
Strategic Learning and Evaluation Center, FSG Social Impact
Advisors,and former President of the American Evaluation Association
by Shanelle Boyle, EVALCORP

Introduction

Evaluation capacity building (ECB) is at the heart of those who work in the international development arena. For more than a decade, efforts have been underway to integrate evaluative thinking and practice into the decision-making processes of governments, corporations, and not-for-profit organizations. Endorsed by the 2005 *Paris Declaration*, "which reflects a commitment by the world's ministers, heads of agencies and senior officials, to increase efforts to harmonize, align and mange 'aid for results' with a set of monitorable actions and indicators", development agencies' commitment to evaluation capacity building has been gaining significant momentum. The 2008 *Accra Agenda for Action* further strengthened the commitments made in Paris for monitoring, documenting and evaluating progress. While it is exciting to have such a strong world wide commitment to monitoring and evaluation, it is critically important that those engaged in evaluations have the necessary knowledge, skills, and attitudes to support effective practice; that is, evaluation capacity.

While various definitions of evaluation capacity building exist in the literature, most writers tend to agree that ECB is about building the knowledge, skills, and attitudes of organization members; increasing the sustainability of professional evaluation practice; and, providing adequate resources and motivations to engage in ongoing evaluation work (Gibbs, Napp, Jolly, Westover, and Uhl, 2002; Milstein and Cotton, 2000; Stockdill, Baizerman, and Compton, 2002). However, we believe that current definitions provide only a partial explanation for what it means to help others to learn about, and engage in, meaningful evaluation practice. To fill this gap, we offer the following definition:

Evaluation capacity building involves the design and implementation of teaching and learning strategies to help individuals, groups, and organizations learn about what constitutes effective, useful, and professional evaluation practice. The ultimate goal of evaluation capacity building is sustainable evaluation practice – where members continuously ask questions that matter, collect, analyze, and interpret data, and use evaluation findings for decision-making and action. For evaluation practice to be sustainable, organization members must be provided leadership support, incentives, resources, and opportunities to transfer their learning about evaluation to their everyday work. Sustainable evaluation practice also requires the development of systems, processes, policies, and plans which help embed evaluation work into the way the organization accomplishes its mission and strategic goals. (Preskill and Boyle, 2008, p. 444)

The fundamental premise underlying this definition is that the organization is committed to internalizing evaluation processes, systems, policies, and procedures which are self-renewing and evolving. It also suggests that the organization must foster a culture that values: trust; risk-taking; openness; curiosity; inquiry; and, experimentation, and that champions the ongoing learning of all its members. As such, an evaluation culture would be reflected by shared and persisting positive beliefs about the value of evaluation to the organization, and an enduring commitment to evaluation practice and the use of findings among its members.

Designing and implementing evaluation capacity building activities and processes

Evaluation capacity building has been a topic on which many presentations and workshops have been given, and articles written. However, little attention has been paid to the ways in which evaluators design and implement ECB activities and processes. As part of a larger research study, we explored evaluators' learning objectives as well as the teaching and learning strategies they used to design and deliver capacity building experiences[1].

---

1   For a copy of the Executive Summary, please contact either of the authors: Hallie. preskill@fsg-impact.org or Shanelle.boyle@gmail.com.

Evaluation capacity building objectives

As illustrated in the definition provided earlier, evaluation capacity building reflects an intention to increase participants' knowledge and skills, and to generate more positive attitudes towards evaluation. Based on the evaluation capacity building literature, we identified 36 knowledge, skills, and belief objectives or competencies which could be used to design and implement any evaluation capacity building effort (Table 1).

Table 1: Evaluation capacity building objectives

Knowledge (Cognitive)
ECB participants understand:
• that evaluation involves purposeful, planned, and systematic activities;
• evaluation terms and concepts;
• the relationship between research and evaluation;
• how evaluation processes and findings can contribute to decision-making;
• the strengths and weaknesses of different evaluation approaches;
• the strengths and weaknesses of different data collection methods;
• how to apply basic statistical analyses to quantitative data;
• how to apply basic content and thematic analyses to qualitative data;
• how politics can affect evaluation processes and findings;
• the importance of using culturally appropriate and responsive evaluation approaches and methods;
• what constitutes ethical evaluation practice;
• that various stakeholders may have differing opinions, experiences, and perspectives about an evaluation;
• the relationship between a programme's goals, objectives, activities, and expected outcomes;
• the knowledge, skills, and experiences to look for when hiring an evaluator.

Skills (Behaviors)

ECB participants are able to:

- develop a programme logic model;
- develop key evaluation questions;
- write an evaluation plan;
- design data collection instruments;
- choose appropriate and relevant data collection methods;
- collect credible and reliable data;
- analyze quantitative data;
- analyze qualitative data;
- interpret results and draw conclusions;
- develop an evaluation budget;
- communicate and report evaluation processes and findings using a variety of strategies;
- use the *Programme Evaluation Standards* and/ or the *AEA Guiding Principles for Evaluators;*
- teach others about evaluation;
- develop an evaluation strategic plan;
- manage the evaluation process.

Beliefs (Affective)

ECB participants believe that:

- evaluation yields useful information;
- evaluation can be a positive experience;
- evaluation should be part of a programme's design process;
- evaluation contributes to a programme's success;
- evaluation adds value to the organization;
- evaluation is an important part of their work;
- evaluation is worth the time and money.

Beginning the ECB design process with clear, specific objectives is key to selecting the most appropriate ECB strategies for those who will be learning from and about evaluation and supporting evaluation activities. The objectives/competencies in Table 1 can also be used to help determine what evaluation capacity already exists within a programme or organization. The ECB strategies chosen should build on and leverage existing capacity to expand particular objectives/competencies to a larger group of organization members.

Types of learning environments

When we think about designing and delivering evaluation capacity building experiences, it is useful to consider the extent to which the learner has a say in the choice of content and delivery methods. For example, we are all familiar with *formal learning* environments. In these situations, the content is chosen by others and presented to participants. The most common example is a training workshop or other type of classroom instruction where participants learn new knowledge and skills and the goal is for participants to transfer their learning to another context.

Informal Learning environments are somewhat different, in that the learning experience is typically unstructured, experiential and non-institutional. It differs from *formal learning* by the degree of control exercised by the learner, location, and predictability of outcomes (Marsick, Volpe, and Watkins, 1999). Informal learning is also a result from the natural opportunities that occur in a person's working life when the person controls his or her own learning (Cseh, Watkins, and Marsick, 1999). Interestingly, in spite of the billions of dollars spent on training every year across sectors, is has been shown that 70-80% of what people know about their jobs, has been learned informally from the people with whom they work (Cross, 2007).

Once the ECB facilitator has chosen the learning objectives, and has considered the type of learning environments best suited for his/her participants, the choice of teaching/learning strategies can be considered. We have provided a worksheet at the end of this chapter (Appendix A), which can be used as a tool for matching the type of learning environment with the teaching/learning strategies.

Determining evaluation capacity building strategies

Our scan of the literature and personal experiences led us to develop a list of ten teaching and learning strategies that can be used to help people learn about and conduct evaluations (Preskill and Boyle, 2008). These are: involvement in an evaluation process; training; technical assistance; written materials; communities of practice; Appreciative Inquiry; technology; internship/apprentice-ship; mentoring/coaching; and, meetings.

While they may be labeled differently in some organizations, and are often used in overlapping ways, the following describes each of the strategies, examples of use, and guiding questions for determining when to use them.

Involvement in an evaluation process: participating in the design and/or implementation of an evaluation. It is generally believed that the mere act of participating in the design and/or implementation of an evaluation is a learning experience (Patton, 2008; Preskill and Torres, 1999). What is learned, how much is learned, and the degree of transferability of that learning will often depend on the extent to which the evaluator has shared responsibility for various evaluation activities and how well dialogue; reflection; asking questions; identifying and exploring values beliefs, and assumptions; and, feedback, have been embedded in the evaluation process.

Guiding questions for involving stakeholders in an evaluation process:

- Which stakeholders should be involved in which parts of the evaluation design and implementation?

- What roles should stakeholders play? How much responsibility should they assume?

- How and where should we build in opportunities to engage in the facilitative learning processes?

- How can we capture and share our learning along the way?

- What are the benefits and risks of engaging stakeholders?

Training: classes/workshops/seminars on evaluation. A workshop is a short-term learning experience that encourages active, experiential learning and uses a variety of learning activities to meet the needs of diverse learners (Brooks-Harris and Stock-Ward, 1999). This teaching/learning strategy is one of the most often employed to build evaluation capacity. Examples include a four hour workshop for programme staff on how to design and conduct evaluations, two

hour workshop on developing logic models, or university course on an evaluation topic.

Guiding questions for developing ECB related workshops:

- What is the purpose and focus of the training? (e.g. skill building; problem solving; increasing knowledge; personal awareness; self improvement; systemic change.)

- What are the goals and objectives?

- Who should be involved? What experience do they have with evaluation? How many people will be trained?

- Where will the training take place?

- How long will the training be?

- How will the trainee's learning be shared with others?

Technical assistance: receiving help from an internal or external evaluator. While the phrase "technical assistance" is often used as an umbrella term for many different evaluation capacity building strategies, it is a specific teaching/learning strategy. Technical assistance has been defined as, "A means of using knowledge to improve the adoption and implementation of some type of educational practice or procedure" (Yin and White, 1984), and "Connections to resources and power, the creation of networks and communities for action and support, and the development of knowledge to solve particular local problems" (Fruchter, Cahill and, Wahl, 1998).

Technical assistance typically involves one-on-one or small group support in the form of:

- telephone consultations;

- newsletters;

- specific knowledge and skill training;

- information dissemination (via website, mailings);

- resources (databases, links, referrals); and,

- facilitation of meetings.

An example is when programme staff are matched to experienced evaluators who are available to answer questions about the evaluation's design and implementation as needed. Communications are typically through email and the phone.

Guiding questions for using technical assistance as an ECB strategy:

- Who might provide technical assistance?
- What resources are available for providing technical assistance?
- What is the scope of technical assistance services?
- Who would have access to technical assistance?

Written materials: using a variety of written documents to learn about evaluation. Organizations can design, develop, and use a variety of written materials to build evaluation capacity (e.g. how-to manuals; newsletters; books; brochures; evaluation plans and reports). Examples may include:

- a checklist of things to include in an evaluation plan;
- evaluation textbooks made available in the organization's library; and,
- newsletter articles about a completed evaluation with descriptive information on what was studied and why; its methods; findings; recommendations; and, how the results will be (or are being) used.

Guiding questions for using written materials as an ECB strategy:

- What kinds of written materials already exist in the organization?
- What kinds of written materials do organization members pay attention to?
- Who will write the new materials?
- What resources exist to design and develop new written materials?
- What might these written materials focus on?
- How will these written materials be distributed?

Communities of practice: sharing evaluation experiences; practices; information; and, readings at self-organized meetings by those who have an interest and/or need to engage in evaluation work. Communities of practice are groups of people who share information, insight, experience, and tools about an area of common interest. They span institutional structures and hierarchies and exist in every organization. They form spontaneously as people seek help, try to solve problems, and develop new ideas and approaches. They dissolve when there is no further reason to exist. (McDermott, 2000; Wenger, 1998). Examples include:

- Several organization members have expressed concern about a particular programme's effectiveness. After several informal conversations about the need to answer some critical questions about the programme, they decide to design and conduct an evaluation. Once the evaluation has been completed, the group disbands.

- After serving on a cross-functional team that has decided to survey a particular department on an organizational issue, team members decide they want to learn more about how to develop good survey questions before going any further. They start a study group whereby each reads an article or book on how to design effective surveys, and meet once a week for four weeks sharing what they have learned.

Guiding questions for using communities of practice as an ECB strategy:

- What communities of practice already exist in the organization?

- How could communities of practice that focus on evaluation be encouraged?

- Who might initiate the development of such a community?

- What kinds of organizational support exist to foster an evaluation community of practice?

- What additional kinds of support would be necessary?

- How might evaluation communities of practice share their learning throughout the organization?

Appreciative Inquiry: using an assets-based, collaborative approach to learning about evaluation that focuses on what has worked well, as well as strengths within the organization or community. Appreciative Inquiry is:

> "The study and exploration of what gives life to human systems when they function at their best. This approach to personal change and organization change is based on the assumption that questions and dialogue about strengths, successes, values, hopes, and dreams are themselves transformational. Appreciative Inquiry suggests that human organizing and change, at its best, is a relational process of inquiry, grounded in affirmation and appreciation." (Whitney and Trosten-Bloom, 2003, p. 1)

Appreciative Inquiry is a particularly useful strategy for increasing positive beliefs about evaluation, and developing motivation and buy-in from participants to learn about evaluation and engage in evaluation practices (Preskill and Catsambas, 2006). The Appreciative Inquiry process can also be used to collaboratively identify ways to improve and/or sustain evaluation practices. The final phase of the Appreciative Inquiry process can lead to the development of a strategic plan for evaluation capacity building when designed with this goal in mind (Boyle, 2009).

Guiding questions for using Appreciative Inquiry as an ECB strategy:

- Has the organization or community ever engaged in an Appreciative Inquiry process? If yes, what happened? What impact did it have?

- To what extent will framing ECB within an Appreciative Inquiry framework be welcomed? Will it be supported?

- To what extent is it possible to bring people together to engage in Appreciative Inquiry processes? (Inquire, Imagine, Innovate, Implement)

- Are there individuals internal to the organization or community who can facilitate an Appreciative Inquiry process geared towards evaluation? If not, what are the implications of hiring an external consultant?

Technology: using online sources such as websites, blogs, and/or e-learning programmes to learn about evaluation. The use of synchronous and asynchronous technologies can provide information and feedback on various aspects of evaluation practice (e.g. email; knowledge management systems; websites; video-conferencing, web conferencing; videotapes; and, social media). Examples include:

- A knowledge management database of previous evaluations has been developed that is accessible to all members in the organization. Employees may, at any time, review previous evaluation studies to learn more about an evaluation's methodology, findings, and recommendations.

- A list of websites which provide information on how to design and conduct evaluations is made available to all organization members.

Guiding questions for using technology as an ECB strategy:

- What are we trying to achieve with using the various technologies?

- What are the most appropriate technologies for these various objectives?

- What are the strengths and limitations of each technology?

- What resources are available to build and use these technologies?

- What experiences do organization members have with using these technologies?

- How likely are members to use these technologies?

Internship/apprenticeship: participating in a formal programme that provides practical evaluation experience for novice evaluators. It involves establishing a formalized relationship with an individual who is considered an expert in the occupation. An example might be a graduate student who works as a research assistant in an evaluation center. The student engages in evaluation activities while she/he takes courses, makes presentations to local and national evaluation organizations, and perhaps even teaches evaluation courses and workshops under the guidance of one or more established evaluators.

Guiding questions for using internship/apprenticeship as an ECB strategy:

- Who are the experts?

- To what extent are experts willing and able to work with the interns or apprentices?

- Who are the interns or apprentices?

- What resources are available to support the interns or apprentice and the expert?

- To what extent does the organization have the systems and structures to support an internship or apprenticeship programme?

- How does the intern or apprentice share his/her learning with others?

Mentoring/coaching: building a relationship with an evaluation expert who provides individualized technical and professional support. This teaching/learning strategy establishes an ongoing one-to-one relationship in which a more experienced individual offers advice, counsel, or guidance to someone less experienced. Mentors provide emotional and psychological support, assistance with career and professional development, and role modeling (Davis, 2001). Examples include:

- University faculty/external evaluator invites graduate student to work on an evaluation project after the student has completed an evaluation course.

- Internal evaluator invites programme staff to work on an evaluation study after they have taken a workshop on evaluation design and implementation.

- Mentors and coaches help individuals:

  - develop knowledge and skills;

  - learn from others' experiences;

  - develop relationships;

  - develop political awareness;

  - build commitment;

  - create networks;

> – understand the larger picture;
>
> – practice according to a set of professional ethics; and,
>
> – integrate and use constructive feedback.

Guiding questions for using mentoring/coaching as an ECB strategy:

- Who is available to be a mentor?
- Who are the potential mentees?
- What structures are there to support a mentoring programme?
- To what extent is a "programme" necessary – should it be more informal?
- What will be expected of mentees and mentors?
- Who will oversee the mentoring "programme"?
- What benefits might be achieved from mentoring individuals?

Meetings: allocating time and space to discuss evaluation activities specifically for the purpose of learning from and about evaluation. This strategy involves incorporating evaluation as a standing agenda item at meetings which already occur on a regular basis (e.g. a weekly staff meeting or a monthly all managers meeting). Whether the organization is in the beginning stages of framing an evaluation, midway through data collection, or even finalizing an evaluation report, it can be useful to have frequent discussions about evaluation to help organization members reflect on and share what they are learning. This can also provide time to answer lingering questions and resolve issues outside the evaluation process itself. Open dialogue about evaluation during meetings builds familiarity and develops an evaluation-friendly culture.

Guiding questions for using meetings as an ECB strategy:

- What types of meetings already occur on a regular basis?
- How often is evaluation on the agenda of these meetings?
- How long are the meetings? Would there be sufficient time to discuss and reflect on what members are learning from and about evaluation?
- Who attends the meetings? Are they an appropriate audience for this type of discussion?

- Is there someone internal who already attends the meetings who could facilitate discussions about evaluation and answer any questions?

- What evaluation topics might organization members be interested in discussing at meetings?

Conclusion

It is our hope that the ECB guidelines discussed in this chapter will help continue the forward momentum for monitoring, documenting and evaluating the progress of international development projects. We believe that evaluation capacity building should be done in an intentional, thoughtful manner in order to make the best use of resources and maximize the impacts of such efforts. The definitions, examples and guiding questions provided for each strategy are intended to help evaluation capacity building facilitators not only decide which strategies make the most sense, given the organization's existing infrastructure and resources as well as cultural context, but also to emphasize that the design and implementation of each strategy is critical for successful capacity building. When we take the time to determine which objectives are most important for our development work, and then carefully consider our options for implementing the strategies which best fit these objectives, we are in a better position to help individuals, groups and organizations learn about and engage in effective evaluation practice.

References

Boyle, S. (2009). *Using Appreciative Inquiry to build evaluation capacity at three non-profit organizations.* Dissertation Abstracts International, 70 (06), Dec 2009. (UMI No. 3363869).

Brooks-Harris, J. E. and Stock-Ward, S. R. (1999). *Workshops: Designing and facilitating experiential learning.* Thousand Oaks, CA: Sage.

Cross, J. (2007). *Informal learning: Rediscovering the natural pathways that inspire innovation and performance.* San Francisco: Pfeiffer.

Davis, B. G. (2001). *Tools for teaching.* San Francisco: Jossey-Bass.

Fruchter, N., Cahill, M. and Wahl, E. (1998). Building to Local Capacity: A Review of Technical Assistance Strategies. New York University Institute for Education and Social Policy.

Gibbs, D., Napp, D., Jolly, D., Westover, B, and Uhl, G. (2002). *Increasing evaluation capacity within community-based HIV prevention programs.* Evaluation and Program Planning, 25, 261-269.

Marsick, V. J., Volpe, M., Watkins, K.E. (1999*). Theory and practice of informal learning in the knowledge era*. Advances in Developing Human Resources, 3, 80-95.

McDermott, R. (1999). *Learning across teams: The role of communities of practice in team organizations*. Knowledge Management Review, May/June.

Milstein, B., Chapel, T. J., Wetterhall, S. F., and Cotton, D. A. (2002). *Building capacity for program evaluation at the Centers for Disease Control.* New Directions for Evaluation, 93, 27-46. San Francisco: Jossey-Bass.

Patton, M. Q. (2008). *Utilization-focused Evaluation.* 4th Edition. Thousand Oaks, CA: Sage.

Preskill, H., and Boyle, S. (2008). *A multidisciplinary model of evaluation capacity building.* American Journal of Evaluation, 29, (4), 443-459.

Preskill, H., and Catsambas, T. (2006). *Reframing evaluation through appreciative inquir*y. Thousand Oaks, CA: Sage.

Preskill, H. and Torres, R. T. (1999). *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage.

Stockdill, S. H., Baizerman, M., and Compton, D. (2002). *Toward a definition of the ECB process: A conversation with the ECB literature.* New Directions for Evaluation, 93, 7-25. San Francisco: Jossey-Bass.

Wenger, E. (1998). *Communities of practice.* Cambridge: Cambridge University Press.

Whitney, D., Trosten-Bloom, A., and Cooperrider, D. (2003). *The power of Appreciative Inquiry: A practical guide to positive change.* San Francisco: Berrett-Koehler.

Yin, R., and White, J. (1984). *Federal Technical Assistance Efforts: Lessons and Improvement in Education for 1984 and Beyond.* Washington, DC: Cosmos Corporation.

Annex A

Strategies for Building Evaluation Capacity by Focus and Type of Learning Planning Worksheet

Strategy	Focus	Individual	Team/Group	Organization
Stakeholder Involvement in Evaluation *Formal & Informal*		X	X	X
Training *Formal*		X	X	X
Mentoring & Coaching *Formal Informal*		X		
Internship & Apprentice-ship *Formal*		X		
Technical Assistance *Informal*		X	X	
Technology *Informal*		X	X	X
Written Materials *Informal*		X		
Community of Practice *Informal*		X	X	X
Appreciative Inquiry *Informal*		X	X	X

USE OF EVALUATION TRAINING
IN EVALUATION CAPACITY BUILDING

by Alexey Kuzmin,
President, Process Consulting Company

Training is important but not enough
for developing evaluation capacity

Numerous publications[1] confirm that *evaluation training is an important component of evaluation capacity building* (ECB). Patton considers evaluators' responsibility to train users in evaluation processes and uses of information, to be one of the fundamental premises of utilization-focused evaluation. He points out that "training stakeholders in evaluation methods and processes attends to both short-term and long-term evaluation uses" (Patton, 1997).

Nevertheless there seems to be a consensus that *although training is important in the ECB process, it is not sufficient for building a sustainable evaluation capacity in organizations.* Stockdill, Baizerman, and Compton (2002) emphasize that even high quality, short evaluation training is not necessarily a part of evaluation capacity building, and should not in any case be considered a synonym for the latter. Milstein et al. (2002) assert that "people and organizations learn evaluation by doing it. No amount of education or exhortation fully prepares one for the task". Horton et al. (2003) develop this idea even further: 'Learning by doing' is fundamental to capacity development. Therefore, managers who wish to develop their organizations' capacities should seek to create an environment that is open to self-criticism, reflection, and improvement".

In most cases, evaluation training is used in organizations along with other interventions, namely:

- Training in results-based management (K. Mackay, 1999; Milstein, Chapel, Wetterhall, & Cotton, 2002; Rist, 2002; UNDP, 1997).

---

1   Boyle & Lemaire, 1999; D. Compton, Baizerman, M., Stockdill, S.H., 2002; Duttweiler & Grogan, 1998; Guerrero, 1999; Hauge, 1998; D. Horton, 2002; D. A. Horton et al., 2003; K. Mackay, 1999; Keith Mackay, 2003; R. Mackay & Horton, 2002; Malik & Roth, 2000; Milstein & Cotton, 2000; Picciotto, 1998; Preskill H.S. & Boyle, 2008; Preskill & Russ-Eft, 2005; Sonnichsen, 1999.

- Supplying additional financial and physical resources (D. Horton, 2002; K. Mackay, 1999; R. Mackay & Horton, 2002; Milstein et al., 2002).

- Information dissemination (D. Horton, 2002; D. A. Horton et al., 2003; K. Mackay, 1999; Milstein et al., 2002).

- Supervised learning by doing, mentoring (D. A. Horton et al., 2003; Milstein et al., 2002).

- Peer learning and cooperation (D. W. Compton, Glover-Kudon, Smith, & Avery, 2002; Gariba, 2003; K. Mackay, 1999).

- Strengthening audit and accounting (K. Mackay, 1999).

- Identification of various intended users and dissemination of evaluation results (Lusthaus, Adrien, & Perstinger, 1999; K. Mackay, 1999; R. Mackay & Horton, 2002).

- Consultations on design and implementation of monitoring and evaluation systems (Haarich & del Castillo Hermosa, 2004; D. A. Horton et al., 2003; Lusthaus et al., 1999; K. Mackay, 1999; Preskill H.S. & Boyle, 2008; Preskill & Russ-Eft, 2005).

- Facilitation of organizational change (Duttweiler & Grogan, 1998; D. A. Horton et al., 2003; R. Mackay & Horton, 2002; Schaumburg-Muller, 1996).

- Connecting employees working in the evaluation field with local, national and international evaluation associations, universities, evaluation consulting groups (D. W. Compton et al., 2002; Gariba, 2003; Kuzmin, 2003; K. Mackay, 1999; Segone, Patel, Rouge, & Russon, 2003).

- Study tours (Gariba, 2003; K. Mackay, 1999).

What factors affect the use of evaluation training in evaluation capacity building?

Based on evaluation literature, our research (Kuzmin, 2004) and practical experience, we identified the following five groups of factors which affect the use of evaluation training in ECB in organizations:

Factors which cause organizations to become involved in evaluation training include:

- external pressure that creates a demand for conducting evaluation, as in a request from a client or donor;

- internal pressure, as in the need to better control field operations, monitor programmes, or inform strategic decisions;

- needs related to the current state of the organization's development, as in "start-up", "renewal", or "formalization";

- the motivation of the top managers caused by dis-confirmation, lack of confirmation; feelings of anxiety, personal interest, curiosity, and so forth.

Factors related to the utilization focus of the training include:

- clarity prior to the training, about the intended use of the training and the primary intended users;

- involvement of the primary intended users in the training design and/or the training.

Based on Patton's definition of utilization-focused evaluation (Patton, 1997) we define utilization-focused evaluation training (as opposed to evaluation training in general) as "evaluation training conducted for and with specific, intended primary users for specific, intended uses". The utilization-focused approach, developed by Patton (1997) for evaluation in general, is applicable to evaluation training in particular. In order for there to be immediate and effective use of evaluation training, it is necessary to identify the primary intended users and the specific intended uses of the training, prior to the training event. It may be helpful, and might even be essential, for the primary users to participate in the training. If a projected specific training use is not identified beforehand, the actual use of a training may, at the least, be delayed.

Factors related to the training itself include:

- consideration of the actual needs of organizations, especially their existing and required evaluation capacities, in the training design;

- duration of the programme;

- incorporation of practice (doing a real evaluation) in the training design;

- use of a variety of training methods, including a combination of lectures and interactive methods;

- the personal involvement, competence and enthusiasm of the trainer or trainers.

Factors related to the complementary learning strategies which reinforce use of the training include:

- **Learning by doing and mentoring**: mentoring and supervised learning by doing can be an effective follow-up of the training. These interventions reinforce the early use of what people learned in training and further develop the participants' ability to conduct quality evaluations.

- **Peer learning**: peer learning can be used as a part of evaluation training to improve its effectiveness and even as a system for transferring evaluation knowledge in organizations. It reinforces the use of evaluation training.

- **Self-education**: self-education is an effective strategy for evaluation capacity building. It extends the use of evaluation training through the dissemination of training information among people who did not attend the actual training. It also allows training participants to refresh their knowledge and to remember details when needed. Availability of evaluation publications and/or the Internet are necessary conditions for self-education.

- **Learning by sharing**: learning by sharing is a powerful way of deepening evaluation knowledge and can be considered one of the best possible outcomes of evaluation training, i.e., people using their new learning to teach other people. Part of the new learning in these cases comes from the reflection and conceptualization needed to prepare evaluation training or to write an article. The questions asked by people receiving information provide another occasion for learning.

Factors related to the sustainability of training outcomes include:

- support from the top managers;

- the regularity and intensity of evaluation practice;

- research background of both employees involved in evaluation and the leaders of organizations;

- the degree of localization[2] of evaluation capacity in the organization;

- the inclusion of evaluation functions in the responsibilities of a person or a group.

- the availability of financial resources is an important condition for all five groups of factors;

- the availability of professional resources is an important condition for training and complementary learning.

Evaluation training as a part of evaluation capacity building intervention

Factors mentioned above which affect the use of evaluation training in evaluation capacity development are related not only to the training itself, but also to the pre-training activities and training follow-up. This means that evaluation training can be effectively used in ECB only when the training becomes a part of an ECB intervention that includes pre-training and post-training activities and involves a certain level of organizational change.

While evaluation training is mainly focused on developing evaluation competence, ECB interventions also necessarily include development of evaluation practice. This is the essential difference between the evaluation training itself, and an ECB intervention that includes evaluation training. As Preskill and Boyle (2008) propose, "the ultimate goal of ECB is sustainable evaluation practice, where members continuously ask questions that matter; collect; analyze and interpret data; and, use evaluation findings for decision-making and action. For evaluation practice to be sustained, participants must be provided with leadership support, incentives, resources, and opportunities to transfer their learning about evaluation to their

---

2   "Localization," the sense in which we are using the word here, refers to the parts of an organization, i.e., the people and the departments, which accumulate evaluation knowledge and skills and that have taken on evaluation as a one of their responsibilities. Sonnichsen (1999b) says that the probability of success in establishing an evaluation function is increased when the "location of evaluation" in the organization is strategically chosen. He argues that an evaluation office should be located where it reports to the head of the organization. We are discussing another phenomenon to which we have given a different name. Evaluation knowledge and skills are often concentrated in certain parts of an organization, i.e. in particular people and department). This localization of evaluation capacity in an organization may not be even. Spreading evaluation capacity and preventing its narrow localization helps to minimize the risk of loosing capacity when people leave the organization.

everyday work. Sustainable evaluation practice also requires the development of systems, processes, policies, and plans which help embed evaluation work into the way the organization accomplishes its mission and strategic goals".

This understanding of evaluation capacity could be presented as an "Evaluation Capacity Formula":

$$EC=C*P$$

where *EC* is evaluation capacity; *C* is evaluation competence; and *P* is ongoing evaluation practice. The EC formula suggests that an organization has evaluation capacity only when it *has* evaluation competence and, *uses* this competence as a part of its ongoing practices. Evaluation capacity depends on the use of competence in practice.

Evaluation capacity development, in turn, can be defined as an organizational intervention aimed at building an organization's evaluation competence and providing effective use of the evaluation competence in an organization's ongoing practices. When an intervention is not balanced and either the competence or practice components are not considered, evaluation capacity will not be developed.

Table 1 presents a theoretical framework for an ECB intervention that includes training. This framework is a summative statement taken from our research (Kuzmin, 2004) and practice, and from the literature. It includes five stages of an ECB intervention which are identical to the groups of factors (described above) affecting the use of evaluation training in evaluation capacity development.

Since an ECB intervention is aimed at increasing evaluation competence and using this competence in the evaluation practice, each stage of an ECB intervention can potentially contribute to both outcomes. The first two stages (deciding about the training and focusing the training) contribute mainly to the evaluation practice. The third stage (designing and conducting the training) contributes mainly to the development of evaluation competence. Each of the last two stages (complementary training opportunities and training follow-up) contributes to both competence and practice outcomes. Each cell contains actions relevant to that stage of ECB intervention.

Conclusion

Evaluation training is an important component of ECB. To maximize effectiveness of training in ECB intervention it is important to consider the following groups of factors which affect the use of evaluation training in building sustainable evaluation capacity in organizations:

1. Factors that cause involvement of organizations in evaluation training;

2. Factors related to the utilization focus of the training;

3. Factors related to the training itself;

4. Factors related to complementary learning that enforces the use of training;

5. Factors related to the sustainability of the training outcomes.

A good utilization-focused evaluation training contributes both to the development of evaluation competence and to the use of evaluation competence in the ongoing practice.

Table 1: Toward a theoretical framework for an ECB intervention that includes a training component

Major contributions of activities undertaken	*Stages of an ECB intervention*					*Outcomes*
	Make thoughtful decisions about evaluation training	*Focus training on the intended use*	*Design and conduct a quality training*	*Provide complementary learning opportunities*	*Sustain training outcomes*	
Contributes to the development of evaluation competence	Get acquainted with evaluation basics. Find out what evaluation is and how organizations could benefit from it.	Find out how evaluation training is used by other organizations. Who are the primary users?	Consider the organization's needs. Consider the existing and required levels of the organization's evaluation competence. Make the training long enough. Use various training methods. Invite competent, committed and enthusiastic trainers. Use evaluation exercises built on the real life situations.	Provide complementary learning opportunities in the course of the training and afterwards. Consider learning by doing; peer learning; mentoring; self-education; learning by sharing. Have a quality control system in place to assure the adequacy of the participants' performance when they are learning by doing and/or by sharing.	Support use of the training results. Practice evaluation and provide opportunities for trainees to do evaluation work. Distribute evaluation competence widely enough in organization. Include evaluation functions in the formal job descriptions or informal responsibilities.	Increased evaluation competence
Contributes to the use of evaluation competence in the ongoing practice	Consider organizational context: the current stage of organizational life cycle; internal and external pressures; and, motivation of the managers.	Identify intended use and intended users of the training. Make sure that they are involved in training design and the training.	Use exercises which enforce use of new skills and knowledge in the participants' everyday practice. Reflect on their immediate experience and help them learn from it.			Effective use of the evaluation competence in the ongoing practice

References

Boyle, R., & Lemaire, D. (Eds.). (1999). *Building effective evaluation capacity: Lessons from practice.* New Brunswick, N.J.: Transaction Publishers.

Compton, D., Baizerman, M., Stockdill, S.H. (Ed.). (2002). *The art, craft, and science of evaluation capacity building.* San Francisco, CA: Jossey-Bass.

Compton, D. W., Glover-Kudon, R., Smith, I. E., & Avery, M. E. (2002). *Ongoing Capacity Building in the American Cancer Society (ACS)* 1995-2001. In D. Compton, Baizerman M. & S. H. Stockdill (Eds.), The art, craft, and science of evaluation capacity building (pp. 47-61). San Francisco, CA: Jossey-Bass.

Duttweiler, M. W., & Grogan, S. (1998). *An Action-Research Approach to Evaluation Capacity Building.* Paper presented at the American Evaluation Association annual conference "Evaluation'98", Chicago, Illinois.

Gariba, S. (2003). *Broadening the Constituency for Evaluation.* In P. G. Grasso, S. S. Wasty & R. V. Weaving (Eds.), World Bank Operations Evaluation Department: The first 30 years (pp. 133-135). Washington, D.C.: The World Bank.

Guerrero, R. P. (1999). *Evaluation Capacity Development in Developing Countries: Applying the Lessons from Experience.* In R. Boyle, and Lemaire, D. (Ed.), Building Effective Evaluation Capacity (pp. 177-194). New Brunswick, N.J.: Transaction Publishers.

Haarich, S. N., & del Castillo Hermosa, J. (2004). *Development of Evaluation Systems - Evaluation Capacity Building in the Framework of the New Challenges of EU Structural Policy.* Paper presented at the ERSA 2004 Conference.

Hauge, A. O. (1998). *Evaluation capacity development: Lessons learned by the UNDP.* In Evaluation Capacity Development in Africa: Selected Proceedings from a Seminar in Abijan. Washington, D.C.: World Bank.

Horton, D. (2002). *Planning, implementing, and evaluating capacity development.* ISNAR Briefing Paper (50).

Horton, D. A., Alexaki, S., Bennett-Lartey, K. N., Brice, D., Campilan, F., Carden, J., et al. (2003). *Evaluating capacity development: experiences from research and development organizations around the world.* The Netherlands: International Service for National Agricultural Research (ISNAR); Canada: International Development Research Centre (IDRC), the Netherlands: ACP-EU Technical Centre for Agricultural and Rural Cooperation (CTA).

Kuzmin, A. (2003). *Regional Evaluation Network and its role in Evaluation Capacity Development in the Newly Independent States.* Paper presented at the Annual conference of the American Evaluation Association - Evaluation' 2003, Reno, Nevada.

Kuzmin, A. (2004). *Exploration of factors that affect the use of evaluation training in evaluation capacity development* (Doctorate Dissertation). Cincinnati, Ohio Union Institute and University.

Lusthaus, C., Adrien, M.-H., & Perstinger, M. (1999). *Capacity Development: Definitions, Issues and Implications for Planning, Monitoring and Evaluation.* Universalia Occasional Paper, 2002, from www.universalia.com

Mackay, K. (1999). *Evaluation capacity development: A diagnostic guide and action framework.* Washington, D.C.: World Bank.

Mackay, K. (2003). *Evaluation Capacity Development.* In P. G. Grasso, S. S. Wasty & R. V. Weaving (Eds.), World Bank Operations Evaluation Department: The first 30 years (pp. 105-114). Washington, D.C.: The World Bank.

Mackay, R., & Horton, D. (2002). *Capacity development in planning, monitoring, and evaluation: Results of an evaluation.* Briefing Paper, 2002, from www.isnar.cgiar.org

Malik, K., & Roth, C. (2000). *Evaluation capacity development in Asia: Selected proceedings from the International conference, Beijing, October 1999.* New York: UNDP, Chinese National Center for Science and Technology Evaluation, and the World Bank.

Milstein, B., Chapel, T. J., Wetterhall, S. F., & Cotton, D. A. (2002). *Building Capacity for Programme Evaluation at the Centers for Disease Control and Prevention.* In D. Compton, Baizerman M. & S. H. Stockdill (Eds.), The art, craft, and science of evaluation capacity building (pp. 27-46). San Francisco, CA: Jossey-Bass.

Milstein, B., & Cotton, D. (2000). *Defining concepts for the presidential strand on building evaluation capacity.* from www.eval.org/eval2000/public/presstrand.pdf

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

Picciotto, R. (1998). *Evaluation capacity development: Issues and challenges.* In Evaluation Capacity Development in Africa: Selected Proceedings from a Seminar in Abijan. Washington, D.C.: World Bank.

Preskill H.S., & Boyle, S. (2008). *A Multidisciplinary Model of Evaluation Capacity Building.* American Journal of Evaluation, *29*(4), 443-459.

Preskill, H. S., & Russ-Eft, D. F. (2005). *Building evaluation capacity: 72 activities for teaching and training.* Thousand Oaks, Calif.: Sage Publications.

Rist, R. C. (2002). *Programme Evaluation - International Development.* In J.-E. Furubo, R. C. Rist & R. Sandahl (Eds.), International atlas of evaluation. New Brunswick: Transaction Publishers.

Schaumburg-Muller, H. (1996). *Evaluating capacity building: Donor support and experiences. Report for the DAC (Development Assistance Committee) Expert Group on Aid Evaluation, OECD (Organization for Economic Cooperation and Development).* Copenhagen, Denmark: DANIDA (Danish International Development Assistance).

Segone, M., Patel, M., Rouge, J. C., & Russon, C. (2003). *Facilitating sustainable monitoring and evaluation capacity development.* Paper presented at the Evaluation' 2003, Reno, Nevada.

Sonnichsen, R. C. (1999). *Building Evaluation Capacity within Organizations.* In R. Boyle, and Lemaire, D. (Ed.), Building Effective Evaluation Capacity: Lessons from practice (pp. 53-74). New Brunswick, N.J.: Transaction Publishers.

Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). *Toward a definition of the ECB process: A conversation with the ECB literature.* In D. Compton, Baizerman M., Stockdill, S.H. (Ed.), The art, craft, and science of evaluation capacity building (pp. 7-25). San Francisco, CA: Jossey-Bass.

UNDP. (1997). *Results-Oriented Monitoring and Evaluation.* New York: UNDP.

Annex 1: Recommendations for practical use: Use of training in ECB checklist

The following checklist based on our research will help practitioners make evaluation training more useful in their evaluation capacity development work in organizations.

Understand the needs

- Does the organization experience any external or internal pressures to develop evaluation capacity? If so, what specifically are the pressures?

- What kind of evaluation capacity does the organization need to respond adequately, and in a timely manner, to these external and/or internal pressures? How will evaluation knowledge and skill help the organization to respond?

- Where is the organization in its lifecycle and what are its developmental needs in consequence of being at that stage? How can evaluation training meet those needs?

- What are the motivations of the top manager(s)? Why are they interested in using evaluation training in their organization? How can training meet their individual interests?

Make sure that the training is utilization-focused

- What is the specific intended use of the training?
- Who are the primary intended users[3]?
- How can these users be involved in the training design?
- How can they be involved in the training?

Design and conduct a high-quality training

- Are the organization's needs reflected in the training design?
- Does the training design fit the existing and required levels of the organization's evaluation capacity? Is it neither too basic nor too advanced?
- Is the training long enough to achieve its goals?

---

3   If the top managers are not among the intended users, decide if the evaluation training will be a good thing to do.

- Does the training include practicing evaluation in real life situations? Does it include reflection on the participants' experience?

- Are various training methods used in the training? Is the balance between lectures and interactive methods appropriate?

- Are the trainers competent, committed and enthusiastic?

Provide complementary learning opportunities

- Will the training participants have complementary learning opportunities in the course of the training? After the training?

- Will these opportunities include learning by doing? Peer learning? Mentoring? Self-education? Learning by sharing?

- Is there a quality control system in place to assure the adequacy of the participants' performance when they are learning by doing and/or by sharing?

Make the training outcomes sustainable

- Do the top managers support use of the training results? Did they make specific decisions regarding this use? Did they involve the evaluation training participants in these decisions?

- Does the organization practice evaluation on a regular basis? Does it provide opportunities for trainees to do evaluation work on a regular basis?

- How is evaluation capacity localized in the organization? Is this capacity distributed widely enough to prevent its loss if people competent in evaluation leave the organization?

- Are the evaluation functions included in the formal job descriptions or informal responsibilities of a person or a group?

Use financial resources effectively

- Are sufficient funds available for a useful evaluation training?

- What is the most appropriate allocation of available funds between the training and related activities, i.e., follow up training, complementary learning, evaluation practice, information dissemination, etc.)?

UTILIZATION-FOCUSED EVALUATION[1]

by Michael Quinn Patton,
Founder and Director, Utilization-focused evaluation,
and former President of the American Evaluation Association

Utilization-focused evaluation begins with the premise that evaluations should be judged by their utility and actual use, therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done, *from beginning to end*, will affect use. Utilization-focused evaluation is concerned with how real people, in the real world apply evaluation findings and experience the evaluation process. Therefore, the *focus* in Utilization-focused evaluation is on intended use by intended users.

In any evaluation there are many potential stakeholders and an array of possible uses. Utilization-focused evaluation requires moving from the general and abstract, i.e. possible audiences and potential uses, to the real and specific: actual primary intended users and their explicit commitments to concrete, specific uses. The evaluator facilitates judgment and decision-making by intended users rather than acting solely as a distant, independent judge. Since no evaluation can be value-free, Utilization-focused evaluation answers the question of whose values will frame the evaluation, by working with clearly identified primary intended users, who have the responsibility to apply evaluation findings and implement recommendations. In essence, Utilization-focused evaluation is premised on the understanding that evaluation use is too important to be merely hoped for or assumed. Use must be planned for, and facilitated.

Utilization-focused evaluation is highly personal and situational. The evaluation facilitator develops a working relationship with intended users to help them determine what kind of evaluation they need. This requires negotiation in which the evaluator offers a menu of possibilities. Utilization-focused evaluation does not depend on or advocate any particular evaluation content, model, method, theory, or even use. Rather, it is a process for helping primary intended users to select the most appropriate content, model, methods,

---

1   Utilization-focused evaluation was originally published in "Оценка программ: методология и практика". Reprinted with the permission of Process Consulting and the author.

theory, and uses for their particular situation. Situational responsiveness guides the interactive process between evaluator and primary intended users. As the entries in this book demonstrate, many options are now available in the feast that has become the field of evaluation. In considering the rich and varied menu of evaluation, Utilization-focused evaluation can include any evaluative purpose (formative, summative, developmental), any kind of data (quantitative, qualitative, mixed), any kind of design (e.g., naturalistic, experimental), and any kind of focus (processes, outcomes, impacts, costs, and cost-benefit, among many possibilities). Utilization-focused evaluation is a process for making decisions about these issues, in collaboration with an identified group of primary users, focusing on their intended uses of evaluation.

A psychology of use underpins and informs Utilization-focused evaluation. In essence, research on evaluation use (cf. Patton, 2008), indicates that intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings. They are more likely to understand and feel ownership if they have been actively involved, and by actively involving primary intended users, the evaluator is training users in use, preparing the groundwork for use, and reinforcing the intended utility of the evaluation every step along the way. While concern about utility drives Utilization -focused evaluation, the evaluator must also attend to the evaluation's accuracy, feasibility and propriety (Joint Committee on Standards, 1994). Moreover, as a professional, the evaluator has a responsibility to act in accordance with the profession's adopted principles of conducting systematic, data-based inquiries; performing competently; ensuring the honesty and integrity of the entire evaluation process; respecting the people involved in, and affected by, the evaluation; and being sensitive to the diversity of interests and values that may be related to the general and public welfare (Shadish et al, 1995).

Basic definitions

Programme evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programmes in order to make judgments about the programme, improve programme effectiveness and/or inform decisions about future programming. *Utilization-focused programme evaluation* (as opposed to programme evaluation in general) is evaluation done for, and with, specific intended primary users for specific intended uses.

The general definition above has three interrelated components:

(1) the systematic collection of information, about

(2) a potentially broad range of topics, for

(3) a variety of possible judgments and uses.

The definition of Utilization-focused evaluation adds the requirement to specify intended use by intended users. This matter of defining evaluation is of considerable importance because different evaluation approaches rest on different definitions. The use-oriented definition offered above contrasts in significant ways with other approaches which define evaluation as measuring goal attainment, providing accountability, or emphasize applying social science methods, in order to judge programme effectiveness.

Involving intended users in making evaluation decisions: the personal factor

Many decisions must be made in any evaluation. The purpose of the evaluation must be determined. Concrete evaluative criteria for judging programme success will usually have to be established. Methods will have to be selected and timelines agreed on. All of these are important issues in any evaluation. The question is: Who will decide these issues? The Utilization-focused answer is: *primary intended users of the evaluation*.

Clearly and explicitly identifying people who can benefit from an evaluation is so important that evaluators have adopted a special term for potential evaluation users: *stakeholders*. Evaluation stakeholders are people who have a stake – a vested interest – in evaluation findings. For any evaluation there are multiple possible stakeholders: programme funders; staff; administrators; and, clients or programme participants. Others with a direct, or even indirect, interest in programme effectiveness may be considered stakeholders, including journalists and members of the general public, or, more specifically, taxpayers, in the case of public programmes. Stakeholders include any one who makes decisions or desires information about a programme. However, stakeholders typically have diverse and often competing interests. No evaluation can answer all potential questions equally well. This means that some process is necessary for narrowing the range of possible questions to focus the evaluation. In Utilization-focused evaluation this process begins by narrowing the list of potential stakeholders to a much shorter,

more specific group of primary intended users. Their information needs, i.e. their *intended uses*, focus the evaluation.

Different people see things differently and have varying interests and needs. This can be taken as a truism. The point is that this truism is regularly and consistently ignored in the design of evaluation studies. To target an evaluation at the information needs of a specific person, or a group of identifiable and interacting persons, is quite different from what has been traditionally recommended as "identifying the audience" for an evaluation. Audiences are amorphous, anonymous entities. Nor is it sufficient to identify an agency or organization as a recipient of the evaluation report. Organizations are an impersonal collection of hierarchical positions. People, not organizations, use evaluation information – thus the importance of the personal factor.

The personal factor is the presence of an identifiable individual or group of people who personally care about the evaluation and the findings it generates. Research on use (Patton, 2008), has shown that where a person or group is actively involved with and interested in an evaluation, evaluations are more likely to be used. Where the personal factor is absent, there was a correspondingly marked absence of evaluation impact.

The personal factor represents the leadership; interest; enthusiasm; determination; commitment; assertiveness; and caring of specific, individual people. These are people who actively seek information to make judgments and reduce decision uncertainties. They want to increase their ability to predict the outcomes of programmatic activity and thereby enhance their own discretion as decision makers, policy makers, consumers, programme participants, funders, or whatever roles they play. These are the primary users of evaluation.

Though the specifics vary from case to case, the pattern is markedly clear: Where the personal factor emerges, where some individuals take direct, personal responsibility for getting findings to the right people, evaluations have an impact. Where the personal factor is absent, there is a marked absence of impact. Use is not simply determined by some configuration of abstract factors; it is determined in large part by real, live, caring human beings.

"Nothing makes a larger difference in the use of evaluations than *the personal factor* – the interest of officials in learning from the evaluation and the desire of the evaluator to get attention for what he knows" (Cronbach et al., 1980: 6; emphasis added).

The importance of the personal factor in explaining and predicting evaluation use leads directly to the emphasis in Utilization-focused evaluation, on working with intended users to specify intended uses. The personal factor directs us to attend to specific people who understand, value and care about evaluation, and further directs us to attend to their interests. This is the primary lesson the profession has learned about enhancing use, and it is wisdom now widely acknowledged by practicing evaluators (Cousins et al, 1996; Preskill and Caracelli, 1997).

User-focused

In essence, Utilization-focused evaluation is user-focused (Alkin, 1995). Since no evaluation can serve all potential stakeholders' interests equally well, Utilization-focused evaluation makes explicit whose interests are served – those of explicitly identified primary intended users.

Attending to primary intended users is not just an academic exercise performed for its own sake. Involving specific people who can and will use information enables them to establish direction for, commitment to, and ownership of the evaluation every step along the way, from initiation of the study through the design and data collection stages right through to the final report and dissemination process. If decision makers have shown little interest in the study in its earlier stages, they are not likely to suddenly show an interest in using the findings at the end. They will not be sufficiently *prepared* for use.

The steps in Utilization-focused evaluation process

First, intended users of the evaluation are identified. These intended users are brought together or organized in some fashion, if possible (e.g., an evaluation task force of primary stakeholders), to work with the evaluator and share in making major decisions about the evaluation.

Second, the evaluator and intended users commit to the intended uses of the evaluation and determine the focus the evaluation, for example, formative, summative, or knowledge generating. Prioritizing evaluation questions will often include considering the relative importance of focusing on attainment of goals, programme

implementation, and/or the programme's theory of action (logic model). The menu of evaluation possibilities is vast, so many different types of evaluations may need to be discussed. The evaluator works with intended users to determine priority uses with attention to political and ethical considerations. In a style that is interactive and situationally responsive, the evaluator helps intended users to answer the question: Given expected uses, is the evaluation worth doing? To what extent and in what ways are intended users committed to the intended use?

The third overall stage of the process involves methods, measurement and design decisions. Primary intended users are involved in making decisions about methods so that they fully understand the strengths and weaknesses of the findings they will use. A variety of options may be considered: qualitative and quantitative data; naturalistic; experimental and, quasi-experimental designs; purposeful and probabilistic sampling approaches; greater and lesser emphasis on generalizations; and, alternative ways of dealing with potential threats to validity, reliability and utility. More specifically, the discussion at this stage will include attention to issues of methodological appropriateness; believability of the data; understandability; accuracy; balance; practicality; propriety and cost. As always, the overriding concern will be utility. Will results obtained from these methods be useful – and actually used?

Once data have been collected and organized for analysis, the fourth stage of the Utilization-focused process begins. Intended users are actively and directly involved in interpreting findings, making judgments based on the data, and generating recommendations. Specific strategies for use can then be formalized in light of actual findings and the evaluator can facilitate following through on actual use.

Finally, decisions about dissemination of the evaluation report can be made beyond whatever initial commitments were made earlier in planning for intended use. This reinforces the distinction between intended use by intended users (planned utilization) versus more general dissemination for broad public accountability (where both hoped for and unintended uses may occur).

While in principle there is a straightforward, one-step-at-a-time logic to the unfolding of a Utilization-focused evaluation, in reality the process is seldom simple or linear. For example, the evaluator may find that new users become important or new questions emerge in the midst of methods decisions. Nor is there

necessarily a clear and clean distinction between the processes of focusing evaluation questions and making methods decisions; questions inform methods, and methodological preferences can inform questions.

Negotiating evaluations to fit specific situations

Utilization-focused evaluation involves negotiations between the evaluator and intended users throughout the evaluation process. This is most obvious, perhaps, at the design stage. The design of a particular evaluation depends on the people involved and their situation. *Situational evaluation* is like situation ethics (Fletcher, 1966), situational leadership (Hersey, 1985), or situated learning: "action is grounded in the concrete situation in which it occurs" (Anderson et al, 1996: 5). The standards and principles of evaluation provide overall direction, a foundation of ethical guidance, and a commitment to professional competence and integrity, but there are no absolute rules an evaluator can follow to know exactly what to do with specific users in a particular situation. That's why Newcomer and Wholey (1989) concluded in their synthesis of knowledge about evaluation strategies for building high-performance programmes, that: "Prior to an evaluation, evaluators and programme managers should work together to define the ideal final product" (p.202). This means negotiating the evaluation's intended and expected uses.

Every evaluation situation is unique. A successful evaluation (one that is useful, practical, ethical, and accurate) emerges from the special characteristics and conditions of a particular situation: a mixture of people; politics; history; context; resources; constraints; values; needs; interests; and chance. Despite the rather obvious, almost trite, and basically common sense nature of this observation, it is not at all obvious to most stakeholders who worry a great deal about whether an evaluation is being done "right." Indeed, one common objection stakeholders make to getting actively involved in designing an evaluation is that they lack the knowledge to do it "right." The notion that there is one right way to do things dies hard. The right way, from a Utilization-focused perspective, is the way that will be meaningful and useful to the specific evaluators and intended users involved, and finding that way requires interaction, negotiation, and situational analysis.

The phrase "active-reactive-interactive-adaptive" describes the nature of the consultative interactions that go on between Utilization-focused evaluators and intended users. The phrase is meant to be both descriptive and prescriptive. It describes how real-world decision-making actually unfolds. Yet, it is prescriptive in alerting evaluators to consciously and deliberately act, react and adapt in order to increase their effectiveness in working with primary intended users.

Utilization-focused evaluators are, first of all, active in deliberately and calculatedly identifying intended users and focusing useful questions. They are reactive in listening to intended users and responding to what they learn about the particular situation in which the evaluation unfolds. They are interactive in the back-and-forth process of negotiating. They are adaptive in altering evaluation questions and designs in the light of their increased understanding of the situation and changing conditions. Active-reactive-interactive-adaptive evaluators do not impose cookbook designs. They do not do the same thing time after time. They are genuinely immersed in the challenges of each new setting and authentically responsive to the intended users of each new evaluation.

This active-reactive-interactive-adaptive stance characterizes all phases of evaluator-user interactions from initially identifying primary intended users, to focusing relevant questions, choosing methods, and analyzing results. All phases involve collaborative processes of action-reaction-adaptation as evaluators and intended users consider their options. The menu of choices includes a broad range of methods, evaluation ingredients from bland to spicy, and a variety of evaluator roles: collaborator; trainer; group facilitator; technician; politician; organizational analyst; internal colleague; external expert; methodologist; information broker; communicator; change agent; diplomat; problem solver; and, creative consultant. The roles played by an evaluator in any given situation will depend on the evaluation's purpose, the unique constellation of conditions with which the evaluator is faced *and the evaluator's own personal knowledge, skills, style, values, and ethics*.

Being active-reactive-interactive-adaptive explicitly recognizes the importance of the individual evaluator's experience, orientation, and contribution by placing the mandate to be "active" first in this consulting triangle. Situational responsiveness does not mean rolling over and playing dead (or passive) in the face of stakeholder interests or perceived needs. Just as the evaluator in Utilization-focused

evaluation does not unilaterally impose a focus and set of methods on a programme, so too, the stakeholders are not set up to impose their initial predilections unilaterally or dogmatically. Arriving at the final evaluation design is a negotiated process that allows the values and capabilities of the evaluator to intermingle with those of intended users.

The Utilization-focused evaluator, in being active-reactive-interactive-adaptive, is one among many at the negotiating table. At times there may be discord in the negotiating process; at other times harmony. Whatever the sounds, and whatever the themes, the Utilization-focused evaluator does not act alone.

Process use

Most discussions about evaluation use focus on use of findings. However, being engaged in the processes of evaluation can be useful quite apart from the findings which may emerge from those processes. Reasoning processes are evaluation's donkeys; they carry the load. If, as a result of being involved in an evaluation, primary intended users learn to reason like an evaluator and operate in accordance with evaluation's values, then the evaluation has generated more than findings. It has been useful beyond the findings in that it has increased the participants' capacity to use evaluative logic and reasoning. "Process use," then, refers to using the logic, employing the reasoning, and being guided by the values that undergird the profession of evaluation.

Those trained in the methods of research and evaluation can easily take for granted the logic that underpins those methods. Like people living daily inside any culture, the way of thinking of those inside the research culture seems to them natural and easy. However, to practitioners, decision makers, and policy makers, this logic can be hard to grasp and quite unnatural. Thinking in terms of what is clear, specific, concrete and observable does not come easily to people who thrive on, even depend on, vagueness, generalities and untested beliefs as the basis for action. Learning to see the world as an evaluator sees it often has a lasting impact on those who participate in an evaluation – an impact that can be greater and last longer than the findings that result from that same evaluation.

Process use refers to, and is indicated by, individual changes in thinking and behavior, and programme or organizational changes in procedures and culture, which occur among those involved in

evaluation as a result of the learning that occurs during the evaluation process. Evidence of process use is represented by the following kind of statement after an evaluation: "The impact on our programme came not so much from the findings but from going through the thinking process that the evaluation required."

Any evaluation can, and often does, have these kinds of effects. What is different about Utilization-focused evaluation is that the process of actively involving intended users increases these kinds of evaluation impacts. Furthermore, the possibility and desirability of learning from evaluation processes, as well as findings, can be made intentional and purposeful. In other words, instead of treating process use as an informal offshoot, explicit and up-front attention to the potential impacts of evaluation logic and processes can increase those impacts and make them a planned purpose for undertaking the evaluation. In that way the evaluation's overall utility is increased.

The groundwork for process use is laid in working with intended users to help them think about the potential and desired impacts of how the evaluation will be conducted. Questions about who will be involved take on a different degree of importance when considering that those most directly involved will not only play a critical role in determining the content of the evaluation, and therefore the focus of findings, but they will also be the people most affected by exposure to evaluation logic and processes. The degree of internal involvement, engagement and ownership will affect the nature and degree of impact on the programme's culture.

Also affected is how funders and users of evaluation think about and calculate the costs and benefits of evaluation. The cost-benefit ratio changes on both sides of the equation when the evaluation produces not only findings, but also serves longer term programmatic needs like staff development and organizational learning.

Six primary types of process use have been differentiated:

(1) infusing evaluative thinking into the programme or organizational culture;

(2) enhancing shared understandings, especially about results;

(3) supporting and reinforcing the programme through intervention-oriented evaluation;

(4) increasing participants' engagement, sense of ownership and self-determination (participatory and empowerment evaluation);

(5) instrumentation effects, that is, what gets measured gets done; and,

(6) programme or organizational development (Patton, 2008).

An example of process use can be found in the framework of Cousins and Earl (1995) who have advocated participatory and collaborative approaches primarily to increase use of findings. Yet, they go beyond increased use of findings when they discuss how involvement in evaluation can help create a learning organization. Viewing participatory evaluation as a means of creating an organizational culture committed to ongoing learning has become an important theme in recent literature linking evaluation to "learning organizations" (e.g., King, 1995; Sonnichsen, 1993).

Utilization-focused evaluation is inherently participatory and collaborative in actively involving primary intended users in all aspects of the evaluation, as a strategy for increasing use of findings. The added attention to process use is how participation and collaboration can lead to an ongoing, longer term commitment to using evaluation logic and building a culture of learning in a programme or organization. Making this kind of process use explicit enlarges the menu of potential evaluation uses. How important this use of evaluation should be in any given evaluation is a matter for negotiation with intended users. The practical implication of an explicit emphasis on creating a learning culture as part of the process will mean building attention to, and training in, evaluation logic and skills into the evaluation.

Summary premises of Utilization-focused evaluation

In summary, the premises of utilization-focused evaluation are:

(1) *Commitment to intended use by intended users* should be the driving force in an evaluation. At every decision point – whether the decision concerns purpose; focus; design; methods; measurement; analysis; or reporting – the evaluator asks intended users, "How would that affect your use of this evaluation?"

(2) *Strategizing about use is ongoing and continuous* from the very beginning of the evaluation. Use isn't something one becomes interested in at the end of an evaluation. By the end of the evaluation, the potential for use has been largely determined.

From the moment stakeholders and evaluators begin interacting and conceptualizing the evaluation, decisions are being made which will affect use in major ways.

(3) *The personal factor contributes significantly to use*. The personal factor refers to the research finding that the personal interests and commitments of those involved in an evaluation, underpin its use. Thus, evaluations should be *specifically* user-oriented – aimed at the interests and information needs of specific identifiable people, not vague passive audiences.

(4) *Careful and thoughtful stakeholder analysis* should inform identification of primary intended users, taking into account the varied and multiple interests that surround any programme, and therefore, any evaluation. Staff; programme participants; directors; public officials; funders; and, community leaders all have an interest in evaluation, but the degree and nature of their interests will vary. Political sensitivity and ethical judgments are involved in identifying primary intended users and uses.

(5) *Evaluations must be focused in some way*. Focusing on intended use by intended users is the most useful way. Resource and time constraints will make it impossible for any single evaluation to answer everyone's questions, or to give full attention to all possible issues. Because no evaluation can serve all potential stakeholders' interests equally well, stakeholders representing various constituencies should come together to negotiate what issues and questions deserve priority.

(6) *Focusing on intended use requires making deliberate and thoughtful choices*. Purposes for evaluation vary and include: judging merit or worth (summative evaluation); improving programmes (instrumental use); and, generating knowledge (conceptual use). Primary information needs and evaluation uses can change and evolve over time as a programme matures.

(7) *Useful evaluations must be designed and adapted situationally*. Standardized recipe approaches won't work. The relative value of a particular utilization focus can only be judged in the context of a specific programme and the interests of intended users. Situational factors affect use. These factors include: community variables; organizational characteristics; the nature of the evaluation; evaluator credibility; political considerations; and, resource constraints. In conducting a Utilization-focused evaluation, the active-reactive-adaptive evaluator works with

intended users to assess how various factors and conditions may affect the potential for use.

(8) *Intended users' commitment to use* can be nurtured and enhanced by actively involving them in making significant decisions about the evaluation. Involvement increases relevance, understanding and ownership of the evaluation, all of which facilitate informed and appropriate use.

(9) *High quality participation is the goal, not high quantity participation.* The quantity of group interaction time can be inversely related to the quality of the process. Evaluators conducting Utilization-focused evaluations must be skilled group facilitators.

(10) *High quality involvement of intended users* will result in high quality, useful evaluations. Many researchers worry that methodological rigor may be sacrificed if non-scientists collaborate in making methods decisions. But, decision-makers want data that are useful *and* accurate. Validity and utility are interdependent. Threats to utility are as important to counter as threats to validity. Skilled evaluation facilitators can help non-scientists understand methodological issues so that they can judge for themselves the trade-offs involved in choosing among the strengths and weaknesses of design options and methods alternatives.

(11) *Evaluators have a rightful stake* in an evaluation in that their credibility and integrity are always at risk, thus the mandate for evaluators to be active-reactive-adaptive. Evaluators are active in presenting to intended users their own best judgments about appropriate evaluation focus and methods; they are reactive in listening attentively and respectful to others' concerns; and they are adaptive in finding ways to design evaluations which incorporate diverse interests, including their own, while meeting high standards of professional practice. Evaluators' credibility and integrity are factors affecting use as well as the foundation of the profession. In this regard, evaluators should be guided by the profession's standards and principles.

(12) *Evaluators committed to enhancing use have a responsibility to train users* in evaluation processes and the uses of information. Training stakeholders in evaluation methods and processes attends to both short-term and long-term evaluation uses. Making decision makers more sophisticated about evaluation

can contribute to greater use of evaluation over time. Any particular evaluation, then offers opportunities to train evaluation users and enhance organizational capacity for use – what has come to be called "process use" – using the evaluation process to support longer term programme and organizational development.

(13) *Use is different from reporting and dissemination.* Reporting and dissemination may be means to facilitate use, but they should not be confused with such intended uses as making decisions, improving programmes, changing thinking, and generating knowledge.

(14) *Serious attention to use involves financial and time costs that are far from trivial.* The benefits of these costs are manifested in greater use. These costs should be made explicit in evaluation proposals and budgets so that utilization follow through is not neglected for lack of resources.

Issues in doing Utilization-focused evaluation

A number of special issues arise in doing Utilization-focused evaluation.

User responsiveness and technical quality

Being responsive to and actively involving primary intended users in an evaluation should not mean a sacrifice of technical quality. A beginning point is to recognize that standards of technical quality vary for different users and varying situations. The issue is not meeting some absolute research standards of technical quality but, rather, making sure that methods and measures are *appropriate* to the validity and credibility needs of a particular evaluation purpose and specific intended users.

Jennifer Greene (1990,) examined in depth the debate about "technical quality, versus user responsiveness." She found general agreement that both are important, but disagreements about the relative priority of each. She concluded that the debate is really about how much to recognize and deal with evaluation's political inherency: "Evaluators should recognize that tension and conflict in evaluation practice are virtually inevitable, that the demands imposed by most if not all definitions of responsiveness and technical quality (not

to mention feasibility and propriety), will characteristically reflect the competing politics and values of the setting (p.273). She then recommended that evaluators "explicate the politics and values" that underpin decisions about purpose, audience, design, and methods. Her recommendation is consistent with Utilization-focused evaluation.

User turnover: the Achilles heel of Utilization-focused evaluation

The Achilles heel of Utilization-focused evaluation, its point of greatest vulnerability, is turnover of primary intended users. The process so depends on the active engagement of intended users that to lose users along the way to job transitions, reorganizations, reassignments and elections can undermine eventual use. Replacement users who join the evaluation late in the process seldom come with the same agenda as those who were present at the beginning. The best antidote involves working with a task force of multiple intended users so that the departure of one or two is less critical. Still, when substantial turnover of primary intended users occurs, it may be necessary to re-ignite the process by renegotiating the design and use commitments with the new arrivals on the scene.

Many challenges exist in selecting the right stakeholders, getting them to commit time and attention to the evaluation, dealing with political dynamics, building credibility and conducting the evaluation in an ethical manner. All of these challenges revolve around the relationship between the evaluator and intended users. When new intended users replace those who depart, new relationships must be built. That may mean delays in original timelines, but such delays pay off in eventual use by attending to the foundation of understandings and relationships upon which Utilization-focused evaluation is built.

Building capacity for evaluation use

Just as students need experience and practice to learn to do evaluations, programmes and organizations need experience and practice to become adept at using evaluations for programme improvement and organizational learning. The field of evaluation is paying more and more attention to ways of building evaluation capacity into programmes and organizations (Kuzmin, 2005; Patton, 1994). Openness to evaluation increases when organizations have positive experiences with evaluation – and learn to reflect on, and take les-

sons from, those experiences. A common problem in introducing evaluation to organizations has been doing too much (large scale efforts and universal mandates), before capacity was sufficient to support useful evaluation. That capacity includes developing administrative and staff understanding of the logic and values of evaluation, developing organization-specific processes for integrating evaluation into planning and programme development, and connecting evaluation to the latest understandings about organizational learning (Sonnichsen, 2000; Preskill and Torres, 1998).

A quarter-century of research on "readiness for evaluation" (Preskill and Torres, 2000; Seiden, 2000; Mayer, 1975) has found that valuing evaluation and learning are necessary conditions for evaluation use. Valuing evaluation cannot be taken for granted. Nor does it happen naturally. Users' commitment to evaluation is typically fragile, often whimsical, and must be cultivated like a hybrid plant that has the potential for enormous yields, but only if properly cared for, nourished, and appropriately managed. Utilization-focused evaluation makes such nurturing a priority, not only to increase use of a particular evaluation but also to build capacity (process use), for utilization of future evaluations.

Variable evaluator roles linked to variable evaluation purposes

Different purposes for evaluation call for varying evaluator roles. Three types reflect evaluation's historical development from three different traditions:

(1) social science research;

(2) pragmatic field practice, especially by internal evaluators and consultants; and

(3) programme and financial auditing.

When evaluation research aims to generate generalizable knowledge about causal linkages between a programme intervention and outcomes, rigorous application of social science methods is called for and the evaluator's role, as methodological expert, will be primary. When the emphasis is on determining a programme's overall merit or worth, the evaluator's role as judge, takes center stage. If an evaluation has been commissioned because of, and is driven by, public accountability concerns, the evaluator's role as independent auditor, inspector, or investigator will be spotlighted for policymakers and the general public. When programme improvement is

the primary purpose, the evaluator plays an advisory and facilitative role with programme staff. As a member of a design team, a developmental evaluator will play a consultative role. If an evaluation has a social justice agenda, the evaluator becomes a change agent (Patton, 2010).

In Utilization-focused evaluation, the evaluator is always a negotiator – negotiating with primary intended users what other roles he or she will play. Beyond that, all roles are on the table, just as all methods are options. Role selection follows from and is dependent on intended use by intended users.

Consider, for example, an international evaluation of food aid to feed rural people during a sever drought. For purposes of accountability and policy review, the primary intended users are members of the programme's oversight committees in the international funding agency. In the midst of an international food crisis, the programme will be highly visible, costly, and possibly controversial, especially because special interest groups often disagree about how food should be distributed and which needy groups should get priority. Under such conditions, the evaluation's credibility and utility will depend heavily on the evaluators' independence, ideological neutrality, methodological expertise, and political savvy.

Contrast such an international accountability evaluation with an evaluator's role in helping a small, rural leadership increase its impact. The programme operates in a few local communities. The primary intended users are school teachers, local officials, and local health professionals who have helped design the programme, with support from an international donor. The evaluation focuses on programme improvement to increase participant satisfaction and to support desired increases in knowledge, as well as behavior changes. Under these conditions, the evaluation's use will depend heavily on the evaluator's relationship with local programme team members. The evaluator will need to build a close, trusting, and mutually respectful relationship to effectively facilitate the team's decisions about evaluation priorities and methods of data collection. Then the evaluator will take them through a consensus-building process as results are interpreted and changes agreed on.

These contrasting case examples illustrate the range of contexts in which programme evaluations occur. The evaluator's role in any particular study will depend on matching her or his role with the context and purposes of the evaluation as negotiated with primary intended users. This is especially true where the Utilization-focused

evaluator and primary intended users agree to include explicit attention to process use. Process use goes beyond the traditional focus on findings and reports as the primary vehicles for evaluation impact. Any evaluation can, and often does, have effects unintentionally or as an offshoot of using findings. What is different about Utilization-focused evaluation is that the possibility and desirability of learning from evaluation processes as well as from findings can be made intentional and purposeful – an option for intended users to consider building in from the beginning. In other words, instead of treating process use as an informal ripple effect, explicit and up-front attention to the potential impacts of evaluation logic and processes can increase those impacts and make them a planned purpose for undertaking the evaluation. In this way the evaluation's overall utility is increased and future evaluation capacity is being built.

However, the Utilization-focused evaluator who presents to intended users options that go beyond narrow and traditional uses of findings, has an obligation to disclose and discuss objections to such approaches. As evaluators explore new and innovative options, they must be clear that dishonesty; corruption; data distortion; and selling out, are not on the menu. Where primary intended users want and need an independent, summative evaluation, that is what they should get. Where they want the evaluator to act independently in bringing forward improvement-oriented findings for formative evaluation, that is what they should get. But those are no longer the only options on the menu of evaluation uses. New participatory, collaborative, intervention-oriented, and developmental approaches are already being used. In Utilization-focused evaluation the new challenge is working with primary intended users to understand when such approaches are appropriate, and helping intended users make informed decisions about their appropriateness, for a specific evaluation endeavor.

Political underpinnings of Utilization-focused evaluation

Utilization-focused evaluation requires astute political sensitivity in identifying both intended uses and intended users for evaluation design. Use always occurs within a political context. Some lessons from practice are:

(1) *Not all information is useful.*

To be power-laden, information must be relevant and in a form that is understandable to users. Organizational sociologist Michael Crozier has observed: "People and organizations will care only about what they can recognize as affecting them and, in turn, what is possibly within their control" (1964: 158).

(2) *Not all people are information users.*

Individuals vary in their aptitude for engaging evaluative information and processes. Differential socialization, education, and experience magnify such differences. In the political practice of evaluation, this means that information is most powerful in the hands of people who know how to use it and are open to using it. The challenge of use is one of matching: getting the right information to the right people. What about people who are not inclined to use information – people who are intimidated by, indifferent to, or even hostile to evaluation? A Utilization-focused evaluator looks for opportunities and strategies for creating and training information users. Thus, the challenge of increasing use consists of two parts: (a) finding and involving those who are, by inclination, information users and (b) training those not so inclined.

(3) *Information targeted at use is more likely to hit the target.*

It is difficult knowing in advance of a decision precisely what information will be most valuable. Utilization-focused evaluation aims to increase the probability of gathering appropriate and relevant information by focusing on real issues, with real timelines, aimed at real decisions. In that way, Utilization-focused evaluation aims at closing the gap between potential and actual use, between knowledge and action. Targeting an evaluation at intended use by intended users increases the odds of hitting the target.

(4) *Only credible information is ultimately powerful.*

Alkin et al (1979), found that the characteristics of both an evaluation and an evaluator affect use and one of the most important characteristics of each is credibility. Eleanor Chelimsky, one of the profession's most experienced and successful evaluators in dealing with Congress, has emphasized this point: "Whether the issue is fairness, balance, methodological quality, or accuracy, no effort to establish credibility is ever wasted. The memory of poor quality lingers long...." (Chelimsky, 1987: 14). The more politicized the context in which an evaluation is conducted and the more visible an evaluation will be in that politicized environment, the more impor-

tant to credibility will be an independent assessment of evaluation quality to establish credibility. This amounts to a form of Utilization-focused evaluation matching-in which safeguards of evaluation credibility are designed to anticipate, and counter, specific political intrusions within particular political environments.

Where possible and practical, an evaluation task force can be organized to make major decisions about the focus, methods and purpose of the evaluation. The task force is a vehicle for actively involving key stakeholders in the evaluation. Moreover, the very processes involved in making decisions about an evaluation will typically increase stakeholders' commitment to use results while also increasing their knowledge about evaluation, their sophistication in conducting evaluations, and their ability to interpret findings. The task force allows the evaluator to share responsibility for decision-making by providing a forum for the political and practical perspectives that come best from those stakeholders who will ultimately be involved in using the evaluation.

Utilization-focused evaluators need special skills

To nurture evaluation use and keep an evaluation from getting caught up in destructive group processes or power politics, a Utilization-focused evaluator needs to be politically savvy, skillful in group facilitation, able to decipher relevant internal organizational dynamics, and a user-friendly communicator (Patton, 2008; Torres et al, 1996). This makes explicit that Utilization-focused evaluators need not only technical and methodological skills, but also group process skills and political astuteness -- what are sometimes called "people skills for evaluators" (Ghere et al, 2006).

Evaluation misuse

Utilization-focused evaluation strives to facilitate *appropriate use* of evaluation findings and processes, so Utilization-focused evaluators must also be concerned about misuse. Evaluation processes and findings can be misrepresented and misused in the search for political advantage. Alkin and Coyle (1988) have made a critical distinction between "*misevaluation*," in which an evaluator performs poorly or fails to adhere to standards and principles, and "*misuse*," in which users manipulate the evaluation in ways that distort the findings or corrupt the inquiry. King (1982) has argued that intentional non-use of poorly conducted studies should be viewed as appropriate and responsible. Here are some premises with regard to misuse.

As use increases, misuse may also increase, so Utilization-focused evaluators must be attentive lest their efforts to call greater attention to evaluations backfire. When people ignore evaluations, they ignore their potential uses as well as abuses. As evaluators successfully focus greater attention on evaluation data and increase actual use, there may be a corresponding increase in abuse, often within the same evaluation experience. Donald T. Campbell made a similar prediction in formulating "a discouraging law that seems to be emerging: the more any social indicator is used for social decision making, the greater the corruption pressures upon it" (1988: 306).

Working with multiple users who understand and value an evaluation is one of the best preventatives against misuse. Allies in use are allies against misuse. Indeed, misuse can be mitigated by working to have intended users take so much ownership of the evaluation that they become the champions of appropriate use, the guardians against misuse, and the defenders of the evaluation's credibility when misuse occurs.

Policing misuse is sometimes beyond the evaluator's control, but what is always squarely within an evaluator's domain of direct responsibility and accountability is misevaluation: failures of conduct by the evaluator, which brings this discussion to evaluation ethics.

Ethics of being user-focused

Sometimes there is concern that in facilitating Utilization-focused evaluation, the evaluator may become co-opted by stakeholders. How can evaluators maintain their integrity if they become involved in close, collaborative relationships with stakeholders? How does the evaluator take politics into account without becoming a political tool of only one partisan interest?

The nature of the relationship between evaluators and the people with whom they work is a complex one. On the one hand, evaluators are urged to maintain a respectful distance from the people they study to safeguard objectivity and minimize personal and political bias. On the other hand, the human relations perspective emphasizes that close, interpersonal contact is a necessary condition for building mutual understanding. Evaluators thus find themselves on the proverbial horns of a dilemma: getting too close to decision makers may jeopardize scientific credibility; remaining distant may undermine use.

One way to handle concerns about co-optation is to stay focused on evaluation's empirical foundation. The empirical basis of evaluation involves making assumptions and values explicit, testing the validity of assumptions and carefully examining a programme to find out what is actually occurring. The integrity of an evaluation depends on its empirical orientation – that is, its commitment to systematic and credible data collection and reporting. Likewise, the integrity of an evaluation group process depends on helping participants adopt an empirical perspective. A commitment must be engendered to really find out what is happening, at least as nearly as one can given the limitations of research methods and scarce resources. Engendering such commitment involves teaching and facilitating. The savvy evaluator will monitor the empirical orientation of intended users and, in an active-reactive-adaptive mode of situational responsiveness, take appropriate steps to keep the evaluation on an empirical and useful path.

Evaluators encounter all kinds of situations that require a strong grounding in ethics that may demand courage. Beyond general ethical sensitivity, however, the ethics of Utilization-focused evaluators are most likely to be called into question around two essential aspects of Utilization-focused evaluation:

(1) limiting stakeholder involvement to primary intended users; and

(2) working closely with those users.

The ethics of limiting and focusing stakeholder involvement concerns who has access to the power of evaluation knowledge. The ethics of building close relationships concerns the integrity, neutrality and corruptibility of the evaluator. Both of these concerns center on the fundamental ethical question: Who does an evaluation – and an evaluator – serve?

First, evaluators need to be deliberative and intentional about their own moral groundings and attend thoughtfully to concerns about whose interests are represented in the questions asked and who will have access to the findings. The active part of being active-reactive-interactive-adaptive invites evaluators to bring their own concerns, issues and values to the negotiating table of evaluation. The evaluator is also a stakeholder – not the primary stakeholder – but, in every evaluation, an evaluator's reputation, credibility and beliefs are on the line. A Utilization-focused evaluator is not passive in simply accepting and buying into whatever an intended user initially desires. The active-reactive-adaptive process implies an obligation on the part of the evaluator to represent the standards

and principles of the profession, as well as his or her own sense of morality and integrity, while also attending to and respecting the beliefs and concerns of other primary users.

A second issue concerns how the interests of various stakeholder groups are represented in a Utilization-focused process. The preferred solution is to work to get participants in affected groups representing themselves as part of the evaluation negotiating process. As noted earlier, user-focused evaluation involves real people, not just attention to vague, abstract audiences. Thus, where the interests of disadvantaged people are at stake, ways of hearing from or involving them directly should be explored, not just to have them represented, in a potentially patronizing manner, by the advantaged. Whether and how to do this may be part of what the evaluator attends to during active-reactive-interactive-adaptive interactions.

A different concern about Utilization-focused evaluation is raised by those who worry that the varied roles available to Utilization-focused evaluators may undermine what some consider evaluation's central (or only) purpose – rendering independent judgments about merit or worth. If evaluators take on roles beyond judging merit or worth, such as creating learning organizations or facilitating judgments by intended users, does that confuse what evaluation is?

Michael Scriven, for example, argues that evaluators do not serve specific people. They serve truth. Truth may be a victim, he believes, when evaluators form close working relationships with programme staff. Scriven admonishes evaluators to guard their independence scrupulously. Involving intended users would only risk weakening the hard-hitting judgments the evaluator must render. Evaluators, he has observed, must be able to deal with the loneliness that may accompany independence and guard against "going native," the tendency to be co-opted by and become an advocate for the programme being evaluated (1991a: 182). Going native leads to "incestuous relations" in which the "evaluator is `in bed' with the programme being evaluated" (p. 192). He has condemned any failure to render independent judgment as "the abrogation of the professional responsibility of the evaluator...." (1991: 32). He has derided what he mockingly called "a kinder, gentler approach" to evaluation (p. 39). His concerns stem from what he has experienced as the resistance of evaluation clients to negative findings and the difficulty evaluators have – psychologically – providing negative feedback. Thus, he has admonished evaluators to be uncompromising in reporting negative results. "The main reason that evaluators avoid negative conclusions is that they haven't the courage for it..." (p. 42).

My experience as a Utilization-focused evaluator has been different from Scriven's, so I reach different conclusions. I choose to work with clients who are hungry for quality information to improve programmes. They are people of great competence and integrity who are able to use and balance both positive and negative information to make informed decisions. I take it as part of my responsibility to work with them in ways that they can hear the results, both positive and negative, and use them for intended purposes. I do not find them resistant. I find them quite eager to get quality information that they can use to develop the programmes to which they have dedicated their energies. I try to render judgments, when we have negotiated my taking that role, in ways that can be heard, and I work with intended users to facilitate their arriving at their own conclusions. They are often harsher on themselves than I would be.

In my experience, it does not so much require courage to provide negative feedback as it requires skill. Nor do evaluation clients have to be unusually enlightened for negative feedback to be heard and used if, through skilled facilitation, the evaluator has built a foundation for such feedback so that it is welcomed for long-term effectiveness. Dedicated programme staff don't want to waste their time doing things that don't work.

Conclusion

The fundamental focus of Utilization-focused evaluation (working with primary intended users to achieve intended use), has become central to the practice of most professional evaluators. Cousins and his colleagues surveyed a sample of 564 evaluators and 68 practitioners drawn from the membership lists of professional evaluation associations in the United States and Canada. The survey included a list of possible beliefs that respondents could agree or disagree with. Greatest consensus centered on the statement: "Evaluators should formulate recommendations from the study." The item eliciting the next highest agreement (71%) was: "The evaluator's primary function is to maximize intended uses by intended users of evaluation data" (Cousins et al, 1996: 215). Preskill and Caracelli (1997), reported similar results from a 1996 survey of American Evaluation Association members. Thus, in thirty years, since the first edition of *Utilization-focused evaluation* (Patton, 1978), its basic premise has moved from a controversial idea (cf. Alkin, 1990), to mainstream evaluation philosophy.

References

Alkin, Marvin (ed.). (1990). *Debates on Evaluation.* Newbury Park, CA: Sage.

Alkin, Marvin. (1995). *Lessons Learned About Evaluation Use.* Panel presentation, International Evaluation Conference, American Evaluation Association, Vancouver, November 2

Alkin, Marvin and Karin Coyle. (1988). *Thoughts on Evaluation Misutilization.* Studies in Educational Evaluation

Alkin, Marvin C., Richard Daillak, and Peter White. (1979). *Using Evaluations: Does Evaluation Make a Difference?* Beverly Hills, CA: Sage.

Anderson, J., L. Reder, and H. Simon. (1996). *Situated Learning and Education.* Educational Researcher., 25, 4: 5-21.

Campbell, Donald T. (1988). *Methodology and Epistemology for Social Science.* Selected Papers, edited by E.S.Overman. Chicago: University of Chicago Press.

Chelimsky, Eleanor. (1987). *The Politics of Programme Evaluation.* Pp. 5-22 in D.S. Cordray, H.S. Bloom, and R.J. Light (eds.), *Evaluation Practice in Review.* New Directions for Programme Evaluation. No. 34, Summer. San Francisco: Jossey-Bass.

Cousins, J. B., J. Donohue, and G. Bloom. (1996). *Collaborative Evaluation in North America: Evaluators Self-reported Opinions, Practices and Consequences.* Evaluation Practice, 17, 3: 207-226.

Cousins, J. Bradley and Lorna M. Earl (eds.). (1995). *Participatory Evaluation in Education: Studies in evaluation use and organizational learning.* London: Falmer Press.

Cronbach, Lee J. and associates. (1980). *Toward Reform of Programme Evaluation.* San Francisco: Jossey-Bass.

Crozier, Michel. (1964). *The Bureaucratic Phenomenon.* Chicago: University of Chicago Press.

Fletcher, Joseph. (1966). *Situation Ethics: The New Morality.* London: Westminster John Knox.

Ghere, Gail, Jean King, Laurie Stevahn, and Jane Minnema. (2006). *A Professional Development Unit for Reflecting on Programme Evaluation Competencies.* American Journal of Evaluation 27(1): 108-123.

Greene, Jennifer C. (1990). *Technical Quality Versus User Responsiveness in Evaluation Practice.* Evaluation and Programme Planning. 13 (3): 267-74.

Hersey, Paul. (1985). *Situational Leader.* North Carolina: Center for Leadership.

Joint Committee on Standards for Educational Evaluation. (1994). *The Programme Evaluation Standards.* Thousand Oaks, CA: Sage.

King, Jean A. (1982). *Studying the Local Use of Evaluation: A Discussion of Theoretical Issues and an Empirical Study.* Studies in Educational Evaluation, 8: 175-183.

King, Jean A. (1995). *Involving Practitioners in Evaluation Studies: How Viable is Collaborative Evaluation in Schools.* Pp. 86-102 in Cousins, J. Bradley and Lorna Earl (eds.). 1995b. Participatory Evaluation in Education: Studies in evaluation use and organizational learning. London: Falmer Press.

Kuzmin, Alexey. (2005). *Exploration of Factors That Affect the Use of Evaluation Training in Evaluation Capacity Development*. Doctoral dissertation, Union Institute and University, Cincinnati, Ohio.

Newcomer, Kathryn E. and Joseph S. Wholey. (1989). *Conclusion: Evaluation Strategies for Building High-Performance Programmes*. Pp. 195-208 in J.S. Wholey and K. E. Newcomer (eds.), Improving Government Performance: Evaluation Strategies for Strengthening Public Agencies and Programmes. San Francisco: Jossey-Bass.

Patton, Michael Quinn. (1978). *Utilization-Focused Evaluation.* Beverly Hills, Ca: Sage.

Patton, Michael Quinn. (1994). *Developmental Evaluation.* Evaluation Practice, 15,3 (October): 311-20.

Patton, Michael Quinn. (2008). *Utilization-Focused Evaluation,* 4th edition. Thousand Oaks, Ca: Sage Publications.

Patton, Michael Quinn. (2010). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press.

Preskill, Hallie and Valerie Caracelli. (1997). *Current and Developing Conceptions of Evaluation Use: Evaluation Use TIG Survey Results.* Evaluation Practice, 18, 3: 209-225.

Preskill, Hallie and Rosalie Torres. (1998). *Evaluative Inquiry for Learning in Organizations.* Thousand Oaks, Ca: Sage Publications.

Preskill, Hallie and Rosalie Torres. (2000). *The Readiness for Organizational Learning and Evaluation Instrument*. Oakland, Ca: Developmental Studies Center.

Scriven, Michael. (1991). *Beyond Formative and Summative Evaluation.* Pp.18-64 in M.W.McLaughlin and D.C. Phillips (eds.), *Evaluation and Education: At Quarter Century*. 90th Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

Scriven, Michael. (1991a). *Evaluation Thesaurus*. 4th edition. Newbury Park, CA: Sage.

Seiden, Katherine (2000), *Development and Validation of the `Organizational Readiness for Evaluation' Survey Instrument*. Unpublished doctoral dissertation, University of Minnesota.

Shadish, William R., Jr., Dianna L. Newman, Mary Ann Scheirer, Christopher Wye. (1995). *Guiding Principles for Evaluators*. New Directions for Programme Evaluation. Number 66. San Francisco: Jossey-Bass.

Sonnichsen, Richard C. (1993). *Can Governments Learn?* In F.Leeuw, R. Rist, and R.

Sonnichsen (eds.) Comparative Perspectives on Evaluation and Organizational Learning. New Brunswick, N.J.: Transaction.

Sonnichsen, Richard C. (2000). *High Impact Internal Evaluation.* Thousand Oaks, Ca: Sage Publications.

Torres, Rosalie; Hallie, Preskill, and Mary E. Piontek. (1996). *Evaluation Strategies for Communicating and Reporting: Enhancing Learning in Organizations.* Thousand Oaks, Ca: Sage.

EMPOWERMENT EVALUATION: AN EXERCISE IN EFFICIENCY

by David Fetterman,
President, Fetterman and Associates

Introduction

Empowerment evaluation is a widely recognized landmark on the intellectual landscape of evaluation. It is a global phenomenon. It has been used in: Australia, Brazil, Canada, Ethiopia, Finland, Israel, Japan, Mexico, Nepal, New Zealand, South Africa, Spain, the United Kingdom, and the United States. Sponsors include government agencies and ministries, foundations, and corporations. It has been successfully applied in many settings, ranging from American Boards of Trustees to South African townships. The approach has been used in public schools; higher education; including medical schools (Fetterman, 2009 and in press); Native American reservations, as well as rural development; environmental protection; adult probation; adolescent pregnancy prevention; substance abuse prevention; and, tobacco prevention programmes[1].

The definition of empowerment evaluation is simple: the use of evaluation concepts, techniques, and findings to foster improvement and self-determination. It is aimed at increasing the probability of achieving programme or curricular success by providing people with tools for:

- assessing the planning;

- implementation, and self-evaluation of their programmes; and

- mainstreaming evaluation as part of their planning and management (Fetterman and Wandersman, 2005).

Theories, principles, concepts and steps

Empowerment evaluation is guided by specific theories, principles, concepts, and steps (Fetterman, 2001; Fetterman and Wandersman, 2005). The most important theories include process use and theories of action and use.

---

[1] See http://www.davidfetterman.com for case examples and videos of empowerment evaluation projects.

Theories guiding empowerment evaluation

One of the most significant theories guiding empowerment evaluation is 'process use'. According to the theory of process use, the more that people engage in the act of conducting their own evaluations the more likely it is that they will find the results credible and act on the recommendations. The reason is because they own them. This enhances knowledge utilization. (See Fetterman, 2005, p. 8-9.)

The 'theory of action' and the 'theory of use' are two additional theories guiding empowerment evaluation practice. The 'theory of action' is the espoused theory of the organization or group of people. It is what they say they are all about. This is compared with the 'theory of use' or what people actually do in practice. Often the theories of action and use are not in alignment. Fundamentally, empowerment evaluation is designed to build feedback loops which help people align what they say they are doing or hope to do with what they are actually doing in practice. (See Fetterman, 2005, p. 13-15.).

Principles of empowerment evaluation

These theories work in conjunction with 10 specific principles of empowerment evaluation. The empowerment evaluation principles (Fetterman and Wandersman, 2005, pp. 1-2, 27-41,42-72) are: improvement; community ownership; inclusion; democratic participation; social justice; community knowledge; evidence-based strategies; capacity building; organizational learning; and, accountability. Each of these principles is detailed below:

Improvement: empowerment evaluation is not a neutral experiment; it is designed to use evaluation to help people accomplish their objectives.

Community ownership: communities have the right to make decisions about actions which affect their lives; a sense of ownership (of the evaluation) is key to knowledge utilization.

Inclusion: a broad representation of participants should be recruited and asked to participate; diversity is additive, not subtractive.

Democratic participation: empowerment evaluation has faith in the capacity of human beings for intelligent judgment and action if proper conditions are furnished; decisions should invite participation and be transparent.

Social justice: recognition of the fact that there are basic social inequities in society and striving to ameliorate these conditions by helping people use evaluation to help improve programmes which impact their social conditions; a commitment to a fair, equitable allocation of resources, opportunities, obligations, and bargaining power.

Community knowledge: empowerment evaluation respects and values community knowledge – the tacit "know-how" knowledge of stakeholders.

Evidence-based strategies: empowerment evaluation respects and uses the knowledge base of scholars; communities are encouraged to adapt evidence based knowledge, rather than adopt it; use of the literature helps to avoid reinventing the wheel.

Capacity building: enhance stakeholders' ability to conduct evaluation and to improve programme planning and implementation.

Organizational learning: use data to inform decision-making, implement practices based on the data, and evaluate new practices; inquire into the systemic consequences of actions rather than settle for short-term solutions which do not address the underlying problem.

Accountability: did the programme accomplish its objectives; stakeholders work together, holding each other accountable and building on each other's strengths to achieve results; empowerment evaluations exists within the context of the existing policies and standards which the programme is already being held accountable to for the project/programme.

According to Fetterman (2005), these principles guide every part of empowerment evaluation, from conceptualization to implementation. The principles of empowerment evaluation serve as a lens to focus an evaluation. The principle of inclusion, for example, recommends erring on the side of including rather than excluding members of the community, even though fiscal and scheduling constraints might suggest otherwise. The capacity building principle reminds the evaluator to provide community members with the opportunity to collect their own data, even though it might initially be faster and easier for the evaluator to collect that same information. The accountability principle guides community members to hold one another accountable. It also places the evaluation within

the context of external requirements. The community is accountable for reaching specific standards or delivering specific results, products, and/or outcomes.

Concepts in empowerment evaluation

Empowerment evaluation is also guided by key concepts including: critical friends; culture of evidence; cycles of reflection and action; community of learners; and, reflective practitioner.[2] A critical friend is an evaluator who believes in the purpose of the programme, but is critical and analytical. They pose questions diplomatically to ensure rigor and honesty, because they want the programme to be more effective and to accomplish its objectives. Empowerment evaluations are conducted by programme staff members, participants, and/or community members. An empowerment evaluator is a critical friend helping to facilitate the process, rather than an external expert controlling it.

A culture of evidence is created by asking people why they believe what they believe. They are asked for evidence or documentation at every stage, so that it becomes normal and expected to have data to support one's opinions and views. Cycles of reflection and action consist of the process of using evaluation data to think about programme practices and then using the data to inform decision-making, e.g. implementing new strategies, eliminating ineffective ones, and so on. The concept emphasizes the cyclical nature of the process, rather than a unilinear approach. Data are continually fed into the decision-making system with the understanding that the programme is dynamic, not static, and will require continual feedback as the programme changes and evolves (and periodically stabilizes). Empowerment evaluation is successful when it is institutionalized and becomes a normal part of the planning and management of the programme, rather than a separate and parasitic entity operating in a "parallel universe." Once institutionalized the cycle of reflection and action is complete because it creates a continual routinized organizational feedback loop.

---

2   These concepts are influenced by traditional organizational development and transformation theorists including Argyris and Schon (1978) and Senge (1994), as well as evaluators associated with organizational learning (Preskill and Torres, 1999).

Empowerment evaluation facilitates an existing community of learners and cultivates new ones[3]. Empowerment evaluation is driven by the group, by design. The group members learn from each other, serving as their own peer review group, critical friend, resource, and norming mechanism. A community of learners is reinforcing, relying on group peer pressure. The group has values held in common and they hold each other accountable on progress toward stated goals. A community of learners also helps to focus the group and keep it on track. Finally, empowerment evaluations produce and then rely on reflective practitioners. Community members learn to use data to inform their decisions and actions concerning their own daily activities. This produces a self-aware and self-actualized individual who has the capacity to apply this world-view to all aspects of their life.

Steps in empowerment evaluation

Empowerment evaluation has accumulated a warehouse of useful tools. The three-step approach to empowerment evaluation is one of the most popular tools in the collection (Fetterman, 2001). It includes helping a group:

1) establish their mission;

2) take stock of their current status; and

3) plan for the future.

The three steps are detailed as follows:

Mission. The empowerment evaluator helps the group come to a consensus concerning their mission or values. The empowerment evaluator asks participants to suggest ideas for the mission statement and records them on a poster sheet of paper. A member of the group, in cooperation with the empowerment evaluator, drafts a mission statement, which is circulated among the group. They are asked to suggest specific changes in wording as needed. It sets the tone for the entire exercise.

Taking stock. The taking stock step has two parts: prioritization and ratings. The empowerment evaluator helps people develop a list of the most important activities required to accomplish organizational or programmatic goals, which are rooted in the mission statement.

---

3   This is critical to an accreditation effort because accrediting agencies are looking for wide-spread faculty, student, and staff member involvement in curricular development, review, and refinement.

A list of 20 to 25 activities is sufficient. The empowerment evaluator gives each participant five dot stickers, and asks the participants to place them by the activities they think are the most important to accomplish programmatic and organizational goals (and thus the most important to evaluate as a group from that point on). They can put one sticker on five different activities or all five on one activity if they are concerned that activity will not get enough votes. The top 10 items with the most dots are selected for the rating exercise which follows.

The empowerment evaluator asks participants in the group to rate how well they are doing concerning each of the activities selected, using a 1 (low) to 10 (high) scale. The columns are averaged horizontally and vertically. The empowerment evaluator facilitates a discussion and dialogue about the ratings, asking participants why they gave a certain activity a 3 or 7 for instance. This information is recorded to document the assessment and facilitate the next step: the planning for the future discussions.

Planning for the future. The planning for the future step involves generating goals, strategies, and credible evidence (to determine if the strategies are being implemented and if they are effective). The goals are directly related to the activities selected in the taking stock step. For example, if communication was selected, rated, and discussed, then communication (or improving communication) should be one of the goals. The strategies emerge from the taking stock discussion, as well, as noted earlier. For example, if communication received a low rating and one of the reasons was because the group never had agendas for their meetings, then preparing agendas might become a recommended strategy in the planning for the future exercise.

Monitoring the strategies

These new strategies are monitored to determine if they are effective. Conventional and innovative evaluation tools are used to monitor the strategies, including online surveys, focus groups, interviews, photo-journaling, and the use of a quasi-experimental design. In addition, programme specific metrics are developed, using baselines, benchmarks, and goals. For example, for the empowerment evaluations of the tobacco prevention programme, the following steps are used:

1) determine the number of people using tobacco in their community (baseline);

2) agree upon the number of people they plan to help stop using tobacco by the end of the year (goal);

3) establish benchmarks concerning how many people they expect to help stop using tobacco each month (benchmarks);

4) describe the implementation of the programme (process);

5) record the number of people who stop using tobacco (outcomes);

6) compare the number of people who stop using tobacco each month with the benchmarks and end-of-year goals (monitoring and evaluation).

The data is fed back to the organization to enable decision-makers to make mid-course corrections and substitute ineffective strategies for potentially more effective ones in real time – before it is too late to get back on course. A second taking stock exercise is conducted 3 to 6 months later, to enable the group to compare their baseline assessments with their second taking stock ratings or second data point. This enables them to document change over time.

The data on change over time are often translated into policy decision-making language, such as, in the tobacco prevention example, the amount of money saved in excess medical expenses. The cycle is unending until the issue is resolved and replaced with a new issue, activity, and/or concern.

The nature of empowerment evaluation

A discussion about the more nuanced aspects of the mission, taking stock, and planning for the future steps, helps to shed light on the nature of empowerment evaluation. In general, empowerment evaluation is primarily about principles and secondarily concerned with specific methods and tools. There are a variety of tools to facilitate an empowerment evaluation, in addition to the three-step model highlighted in this discussion. Some groups prefer to circle key words such as communication, and to place red dots (negative) and green dots (positive) around the circle to signify where they think things are going well and where they merit improvement. Wandersman's 10-step model is another popular tool to facilitate the process (Chinman, M., Imm, P., and Wandersman, A., 2004). In addition, the role of the critical friend is critical to the process. Their success depends on the degree to which they can successfully foster a constructive and trusting climate.

The remaining comments or discussion focus on the three-step approach to empowerment evaluation: mission; taking stock; and, planning for the future.

Mission. It is important to take time to conduct the mission step in the process. A pre-existing mission may exist and should be acknowledged and respected. However, an older mission statement rarely represents the values of the current group of people in the organization. This exercise provides people currently "sitting at the table" with an opportunity to own the programme. Typically, participants will censor themselves when something completely outside of, or antithetical to, the tenants of the organization is suggested. Participation in the construction of a mission statement gives people a voice and helps them chart the future direction of the organization. It mentally sets the tone for taking stock of the organization. It is a form of intellectual scaffolding, as one step prepares people for the next. The process of crafting a formal mission statement, however, should not delay the 'taking stock' step.

Taking stock. It is not feasible or desirable to evaluate everything. The use of sticky dots is fast and transparent. The dots help people "vote" for the activities requiring assessment right now; during that phase of the organization's developmental life cycle. Other approaches are less efficient and cast doubt on the process and final selection of activities. The dots represent another mechanism to get people involved in conducting the evaluation and in a tactile manner.

There are a number of reasons why the ratings are averaged vertically and horizontally, as discussed earlier. Vertically, they highlight a person's degree of skepticism or optimism from an evaluative perspective. This is a form of norming. For example, the next time a person with a low average rating says something positive, people take notice. Horizontally, the ratings are averaged to produce a picture of how well or how poorly the group thinks they are doing. They can discuss their group ratings first by activity and then as a whole by taking an average of all the individual activity averages. These activity averages create a baseline for future comparison.

The ratings are not confidential. Participants place their initials on the top of their column of ratings. The point of the exercise is to engage in a dialogue, specifically asking people about their ratings. The reasons participants provide for their ratings during the dialogue represent the evidence required to support their views. They also provide an insight into the solutions. It reduces ambiguity

about the nature of the problem by asking people to present concrete examples. The reasons for a specific rating about communication, for example, points to the type of strategy needed to improve it in the 'planning for the future' step. The dialogue also highlights what's working which enables the group to build on strengths and can also be applied to specific strategies in the 'planning for the future' exercise.

The dialogue aspect of the taking stock step typically surfaces the "elephant in the room" or the issue that significantly impacts programme operations, but that no one talks about. The issue may be politically sensitive, but this framework provides a "safe environment" for these concerns to be openly examined. The ratings dialogue is also illuminative. Colleagues often learn more about the resources and activities in their own organization as a function of people explaining and supporting their ratings. For example, all but one person might rate public relations activities low. When that person explains how much they do in that area, it not only provides a solid foundation for the rating, it also informs and educates the rest of the group about a new resource they can use in their own organization.

Planning for the future. This last step creates a thread of intellectual coherence and continuity in the evaluation. It is the crystallization of the group's or community's action steps, which are rooted in their self-assessment, which is rooted in their mission. Empowerment evaluators can only determine the methods they plan to recommend after the group has mapped out its plans for the future. It represents the organization's intervention.

One of the most important contributions planning for the future makes in an empowerment evaluation is institutionalization. At the end of this step, individuals are asked to volunteer to lead groups on each of the goals. The next time the group meets, at a staff or community meeting, that person's supervisor or leader asks that person for an update. This update and discussion is requested during a normal meeting (not a special evaluation meeting), making the question a part of the normal planning and management of the group. Internalizing or institutionalizing evaluation enhances the sustainability of self-assessment and reflection.

Empowerment evaluation as a process of reduction.

Empowerment evaluation is a process of reduction. The process helps people visualize the "big picture" or mission. It takes people from a 30,000 foot panoramic view of the landscape to a more focused view of critical activities associated with programme performance. Empowerment evaluation further reduces the scope by focusing on specific strengths and weaknesses associated with each major activity. This step helps people prepare very specific, focused, relevant, and meaningful strategies and solutions, instead of generating a host of logical but irrelevant ideas, which misdirect efforts and drain the organization of vital resources. Each step in the empowerment evaluation process helps participants and staff members to narrow their scope and to focus on the most significant activities required to improve programme performance and actualize its goals. It is, fundamentally, an exercise in efficiency since the process narrows the scope; surgically pinpoints organizational strengths, weaknesses, and solutions; and, maps out a specific plan of action.

Conclusion

Empowerment evaluation is an approach that has captured the imagination of community members, programme staff, donors, and evaluators throughout the world[4]. Empowerment evaluation is simple, powerful, and honest. In addition, it is an evaluation approach designed to enable people to help themselves by internalizing and institutionalizing evaluation as part of their normal day-to-day planning and operations. Empowerment evaluation helps to build capacity and enhance self-determination. It leaves a legacy of evaluation practice behind which contributes to sustainability.

---

[4]   (See Fetterman and Wandersman, 2007. See also the empowerment evaluation web page at: http://www.davidfetterman.com and select "empowerment evaluation". See also the empowerment evaluation blog at: http://eevaluation.blogspot.com for American and international case examples.)

References

Chinman, M., Imm, P., and Wandersman, A. (2004). *Getting To Outcomes 2004*: *Promoting Accountability Through Methods and Tools for Planning, Implementation, and Evaluation*. Santa Monica, CA: Rand Corporation.

Fetterman, D.M. (2001). *Foundations of Empowerment Evaluation*. Thousand Oaks, CA: Sage.

Fetterman, D.M. (2005). *A window into the heart and soul of empowerment evaluation: looking through the lens of empowerment evaluation principles.* In Fetterman, D.M. and Wandersman, A. (2005). Empowerment evaluation principles in practice. New York: Guilford Publications.

Fetterman, D.M. (2009). *Empowerment evaluation at the Stanford University School of Medicine: using a critical friend to improve the clerkship experience.* Ensaio: aval. Pol. Publ. Educ., Rio de Janeiro, 17(63): 197-204. http://www. davidfetterman.com/Ensaioclerkships.pdf

Fetterman, D.M. (in press). *Empowerment evaluation: a collaborative approach to evaluating and transforming a medical school curriculum*. Academic Medicine.

Fetterman, D.M. and Wandersman, A. (2005). *Empowerment evaluation principles in practice*. New York: Guilford Publications.

Fetterman, D.M. and Wandersman, A. (2007). *Empowerment evaluation: yesterday, today, and tomorrow.* American Journal of Evaluation, 28(2):179-198.

COLLABORATIVE EVALUATION. CREATING ENVIRONMENTS TO ENGAGE STAKEHOLDERS IN EVALUATION[1]

Rita O'Sullivan,
University of North Carolina,
Associate Professor and EvAP Director

Introduction

Collaborative evaluation invites and engages stakeholders systematically in programme evaluation planning and implementation. Unlike "distanced" evaluation approaches, which reject stakeholder participation as members of evaluation teams, Collaborative evaluation assumes that active, on-going engagement between evaluators and programme staff, result in stronger evaluation designs, enhanced data collection and analysis, and results which stakeholders understand and use.

As with any of the many evaluation approaches (Fitzpatrick, Sanders, and Worthen, 2004), Collaborative evaluation consistently follows a typical programme evaluation processes. In this the client/supervisor is interested in asking questions about a programme, which will require the systematic collection of information to answer those questions. In conducting the evaluation, a competent evaluator is expected to follow appropriate professional guidelines (see for example the *Guiding principles for evaluators* of the American Evaluation Association; *Evaluation guidelines* from the International Programme Evaluation Network; or, the United Nations' *Standards for evaluation*). Such guidelines help to assure that the evaluation is of high quality. The collaborative aspect of the evaluation is found in how the evaluators go about conducting the evaluation, which includes a variety of efforts to engage programme stakeholders in the evaluation.

---

1   "Collaborative evaluation: creating environments to engage stakeholders in evaluation" was originally published in "Оценка программ: методология и практика". Reprinted with the permission of Process Consulting and the author.

Among similar "participant-oriented" evaluation approaches (Fitzpatrick, Sanders, and Worthen, 2004), Collaborative evaluation distinguishes itself in that it uses a sliding scale for levels of collaboration. This means that different programme evaluations will experience different levels of collaborative activity. The sliding scale is applied as the evaluator considers each programme's evaluation needs, readiness, and resources.

Liliana Rodriguez-Campos, another Collaborative evaluation proponent wrote, "Collaborative evaluation is an evaluation in which there is a significant degree of collaboration between the evaluator and stakeholders in the evaluation process" (2005, p. 1). Thus, a Collaborative evaluation stance requires evaluators to enhance evaluation activities by creating environments which invite and allow stakeholder involvement. Additionally, collaborative evaluators need to understand and assess barriers to collaboration and create opportunities to overcome them. All this must be done cognisant of the evaluation being conducted and sensitive to the organizational context of the programme.

Theoretical perspectives

While Collaborative evaluation is a term widely used in evaluation, its meaning varies considerably. Often used interchangeably with participatory and/or empowerment evaluation, the terms can be used to mean different things, which can be confusing. The Topical Interest Group, representing evaluators following this "participant-oriented" approach in the American Evaluation Association, have entitled themselves, "Collaborative, Participatory, and Empowerment Evaluation." They have made no formal attempt at defining and/or distinguishing among terms. Therefore, for the purposes of this presentation, to reduce confusion, a Collaborative evaluation approach is defined as an evaluation that actively engages stakeholders of the programme as members of the evaluation team, to the extent that they are able and willing. The work of O'Sullivan (2004), and Rodriguez-Campos (2005), best represent how such evaluations are planned and implemented.

From a broad, theoretical perspective, Collaborative evaluation belongs to the "Use" branch of evaluation theory as described by Alkin (2004), in *Evaluation Roots*. It has much in common with Participatory evaluation (Cousins, 2004; Cousins and Earl, 1995; King, 1998, 2004), Utilization-focused evaluation (Alkin, 2004; Patton, 2004, 2008), and Empowerment evaluation (Fetterman,

2004; Fetterman and Wandersman, 2005). Consistent among these approaches is a strong appreciation of stakeholder involvement in evaluation and a desire for the evaluation results to be useful.

Collaborative evaluation distinguishes itself from these other approaches by its inclusion of all stakeholders as potential collaborators and the sliding-scale nature of that collaboration. There is not a set amount of collaboration, or step-by-step directions, about how the collaboration should proceed. Neither is there a set assumption about who has the ultimate authority in conducting the evaluation The assumption is that collaboration with programme staff and other stakeholders enhances the evaluation, and so Collaborative evaluators seek this participation to the extent that programme stakeholders are willing and able to become involved.

Prior to beginning an evaluation, levels of collaboration are explored and negotiated. The negotiation could lead to high levels of involvement in the evaluation by stakeholders to virtually no involvement by stakeholders. Ideally, the negotiation assesses the evaluation capacity of stakeholders and creates an evaluation plan that will allow for the collection of quality information to answer evaluation questions. It also aspires to enhance the evaluation expertise of participating stakeholders. This adds appreciable importance to the steps surrounding the initial clarification of an evaluation request.

Levels of decision-making power are also negotiable (within the confines of ethical evaluation procedures and guidelines), with a Collaborative evaluation approach, which is not true of the other approaches. Fetterman (2001), distinguished empowerment evaluation from participatory evaluation (as represented by Brunner and Guzman, 1989), using the dimension of stakeholder control. He saw Empowerment evaluation assuming greater stakeholder control than Participatory evaluation. Later Fetterman (2005) likened the evaluator to a facilitator who coaches those community members who are engaged in and in charge of the evaluation. For Utilization-focused evaluation, Patton (1997) portrays external evaluators as facilitators, collaborators, and resources, while participants are decision-makers and evaluators. Cousins and Earl (1995), in contrast, maintain that with Participatory evaluation, the external evaluator leads the evaluation.

Collaborative evaluation sequences

Collaborative evaluation projects follow paths equivalent to distanced (e.g. pre-ordinate) evaluations, or any other applied empirical research methods. Evaluation questions (hypotheses) are identified, and then data are gathered to answer the evaluation questions (test the hypotheses). Techniques to gather data are selected from the common research toolbox of sampling and data collection strategies. Once the data are collected, the analysis techniques used and data reporting options selected also are very much the same as they would be for most evaluation or applied social science research project. The key differences with Collaborative evaluations are: how the techniques are implemented, and by whom. Further, the sequencing of events and the communication around events are extremely important.

Rodriguez-Campos (2005), has proposed a "Model for Collaborative Evaluations" that exemplifies the importance of building the evaluation team and meeting collaborative challenges. Her six part system includes:

a) identifying the situation;

b) clarifying the expectations;

c) establishing a shared commitment;

d) ensuring open communication;

e) encouraging best practices; and

f) following specific guidelines.

She details the elements of each of the components, explains their importance, and supports them with examples of how to collaboratively engage stakeholders in the evaluation process. She emphasizes that the six components are to be used as needed to promote collaboration and should not be considered as a menu of Collaborative evaluation practice.

At Evaluation, Assessment and Policy Connections (EvAP) in the School of Education at the University of North Carolina at Chapel Hill, O'Sullivan (2007), has identified a set of Collaborative evaluation techniques which have worked well with multi-site, multi-year evaluations (see for example: Jay, O'Sullivan, and Costello, 2006; O'Sullivan, Hudson, Jay, Anderson, Steiner, Muller, and Powers, (2006); O'Sullivan, Skaga, and Chernow (2005)). While multi-

year evaluations are not essential, they certainly provide a stable base upon which to build enhancements in on-going collaboration and evaluation expertise. With multi-year evaluations, of course, the cycle repeats. In general there are four steps in the cycle:

a) review programme status;

b) develop evaluation plans;

c) implement the evaluation; and

d) share evaluation findings.

Each of these steps has recommended evaluation techniques that support the Collaborative process. Throughout, communication between evaluator and key programme stakeholders is essential. Figure 1 below depicts the annual cycle of activities associated with such a Collaborative evaluation, including recommended collaborative techniques. The Collaborative evaluation activities listed in Figure 1 are expanded in the section that follows.

Figure 1: Annual Collaborative evaluation cycle

Review programme status:
• Understanding the context
• Asking for available evidence
• Probing Evaluation Culture

Share evaluation findings:
• Evaluation fairs/conferences/ learning communities/ networks
• Cross-site summaries

COMMUNICATION

Develop evaluation plans:
• Programme develops logic model
• Evaluator feedback on logic model plans

Implement the evaluation:
• Identification of Common Assessment Needs
• Programmes help develop and review instrument drafts
• Member checking of data
• Stakeholder assist with analysis data

Collaborative evaluation techniques: steps in conducting multi-site collaborative programme evaluation

The four steps in conducting multi-site collaborative programme evaluation are:

Step One: Review programme status. Among things to consider at this beginning stage are: understanding the context of the programme, which would include the cultural and political contexts within which the programme operates as well as the particulars of the programme content and its status vis-à-vis similar programmes; asking for any available evaluation evidence, which would include previous evaluation and monitoring reports, proposals, internal reports, etc.; probing the evaluation culture, which would include previous experiences with evaluation, other evaluation requirements across the agency, evaluation expertise and other resources available to assist with the evaluation, reluctance or willingness or programme personnel to engage in evaluation, etc.

Step Two: Develop evaluation plans. These techniques start with Collaborative evaluation planning, which usually begins with training around the format of the desired evaluation planning and a discussion of the cross-site evaluation needs. Coupled with this discussion is an emphasis on grantees adding site-specific evaluation questions to their evaluation plans. In most cases grantees/sites are asked to develop a logic model for their evaluation, whose format has been developed in consultation with the client organization. The evaluator provides feedback on the grantee/site logic models and then asks grantees to work on subsequent drafts, if necessary. Essential here is that the plan remains that of the grantee. During this phase of the Collaborative evaluation, evaluators usually find that about one-third of the grantees are able to do this activity with very little assistance beyond the initial planning meeting. Another third will require 1-2 drafts to complete the task, and the final third may require extensive technical assistance to complete their logic models.

Step Three: Implement the evaluation. As evaluators have already identified the cross-site evaluation requirements and seen drafts of all the logic models in the previous step above, they are in a position to identify common assessment needs across sites. This is something that the skill sets of most evaluators cover and grantees/sites appreciate, not to mention the increased potential for developing

more valid and reliable instruments. As part of common instrument construction, programmes help to develop and review instrument drafts. This technique places more programme experts around the review table to safeguard content validity, language appropriateness, and realistic data collection procedures. With qualitative data collection, evaluators' practice and sites are strongly encouraged to use member checking of data. This provides opportunities for participants to comment and amend draft data summaries from focus groups and interviews. This effort makes the data collection activity more transparent and once again actively engages stakeholders. Finally, involving grantees/sites in data analysis is a technique that strengthens the quality of information by validating interpretations which can be made about the programme.

Step Four: Share evaluation findings. The primary vehicle for this step is a networking technique, known by a variety of names: Evaluation Fairs; Conferences; Learning Communities; Network Meetings. With this activity grantees/sites are told at the evaluation planning stage that they will be brought together at the end of the year to share their evaluation findings. In consultation with the client organization and other key stakeholders, evaluators set the outline for the oral presentations and usually require grantees to submit a written report at the same time. The power of the event is incredible on a number of levels: providing a time for grantees/sites to celebrate accomplishments; allowing other grantees to hear about promising practices; and providing a forum that, over time, contributes to strengthened evaluation reporting. Evaluators then use the written reports and oral presentations to develop cross-site summaries of accomplishments.

This multi-site, Collaborative evaluation system has been used in a variety of programme settings both large and small, including early childhood; health education; substance abuse prevention; organizational development for non-profits; sustainable agriculture; international survey research; college access; school career awareness; literacy; and K-12 international presenter programmes (See Jay, O'Sullivan, and Costello, 2006, for a detailed description of how this worked in a statewide programme that promoted college access with 15 school districts).

Positive effects of Collaborative evaluation

Using a Collaborative evaluation approach can result in a number of positive outcomes which both directly and indirectly effect the evaluation, along with participating stakeholders and programmes. Foremost is the fact that Collaborative evaluation has been shown to improve data collection and reporting (O'Sullivan and D'Agostino, 2002). Further, the approach is cost effective in that it includes programme staff as part of the evaluation team, thereby reducing the number of external evaluators needed to collect data for the evaluation. For the grantee/site staff and client organizations, participation in a Collaborative evaluation causes them to better understand the evaluation process and, thereby, become better consumers of evaluation. The process also is empowering to them as they learn that they can actively manage what, previously, was daunting and unmanageable.

Conclusions

Collaborative evaluation strategies include involving stakeholders in a variety of ways:

- reviewing programme status, using site visits to develop relationships with stakeholders;

- developing and refining evaluation plans with individual stakeholders activities that allowed for common and unique outcomes;

- providing stakeholder specific evaluation training for staff members and community partners;

- implementing the evaluation by developing a performance measure tracking-system with substantial assistance from stakeholders;

- supplementing evaluation evidence with newly proposed data collection activities to verify findings; and

- sharing evaluation finding through the use of Evaluation Fairs.

A key lesson learned is that communication is crucial to effective on-going collaboration and the successful implementation of the evaluation plan. Systematic engagement of stakeholders in the process creates an environment where people are aware not only of what is going on, but also why it needs to happen.

Thus, evaluators must regularly make efforts to communicate with new stakeholders, who may join the evaluation efforts already in progress, to make sure they support the plans. With such a collaborative environment, when assistance or adaptations are needed, stakeholders are more likely to cooperate. They are also more likely to use evaluation findings, and that should result in programme improvement.

References

Alkin, M. C., Ed. (2004). *Evaluation roots.* Thousand Oaks, CA: Sage.

Alkin, M. C. (2004). *Context adapted utilization: A personal journey.* In M. C, Alkin (Ed.) Evaluation roots (pp. 293-303). Thousand Oaks, CA: Sage.

American Evaluation Association. (2004). *AEA Guiding Principles for Evaluators.* Retrieved January 7, 2009 from http://www.eval.org/Publications/GuidingPrinciplesPrintable.asp

Cousins, J. B. (2004). *Crossing the bridge: Toward understanding us through systematic inquiry.* In M. C. Alkin, Ed. Evaluation roots (pp. 293-318). Thousand Oaks, CA: Sage.

Cousins, J. B., and Earl. L. M. (1992). *The case for participatory evaluation*, Educational Evaluation and Policy Analysis 14(4): 397-418.

Cousins, J. B., and Earl. L. M., Eds. (1995). *Participatory evaluation in education: Studies of evaluation use and organizational learning.* London: Falmer.

Fetterman, D. M. (2001). *Foundation of empowerment evaluation.* Thousand Oaks, CA: Sage.

Fetterman, D. M. (2004). *Branching out or standing on a limb: Looking to our roots for insight.* In M. C. Alkin, Ed. Evaluation roots (pp. 304-318). Thousand Oaks, CA: Sage.

Fetterman, D. M. (2005). In D. M. Fetterman and A. Wandersman, Eds. *Empowerment evaluation principles in practice* (pp. 1-26). New York, NY: Guilford.

Fetterman, D. M., and Wandersman, A. (2005). *Empowerment evaluation principles in practice.* New York, NY: Guilford.

Fitzpatrick, J. L., Sanders, J. R. and Worthen, B.R. (2004). *Programme evaluation,* 3rd Edition. Boston: Pearson.

Jay, M, O'Sullivan, R. G., and Costello, E. P. (2006). *Using Collaborative Evaluation to Capture GEAR UP North Carolina's Programme Outcomes.* Paper presented at the Annual Meeting of the American Educational Research Association - San Francisco, CA, April 11, 2006.

King, J. A. (1998). *Making sense of participatory evaluation practice*, New Directions for Evaluation 80: 57-67.

King, J. A. (2004). *Tikkun Olam: The roots of participatory evaluation*. In M. C, Alkin, Ed. Evaluation roots (pp. 293-303). Thousand Oaks, CA: Sage.

O'Sullivan, R. G. (2004). *Practicing evaluation: A Collaborative approach*. Thousand Oaks, CA: Sage.

O'Sullivan, R. G. (2007). *EvAP Evaluation Institute Training Manual*. Chapel Hill, NC: Evaluation, Assessment and Policy Connections (EvAP), School of Education, University of North Carolina at Chapel Hill.

O'Sullivan, R. G., and D'Agostino, A. (2002). *Promoting evaluation through collaboration: Findings from community-based programmes for young children and their families*. Evaluation, *8*(3), 372-387.

O'Sullivan, R. G., Hudson, M. L., Jay, M., Anderson, A., Steiner, N., Muller, T. and Powers, E. (2006). *Toolkit for Evaluation of K-12 International Outreach Programmes*. Retrieved January 11, 2009 from http://cgi.unc.edu/research/k-12-toolkit/evaluating-k-12.html.

O'Sullivan, R. G., Skaga, J., and Chernow, M. (2005). *First 5 Los Angeles Child Care Initiative Evaluation Phase II Final Report*. Chapel Hill, NC: Evaluation, Assessment and Policy Connections, School of Education, University of North Carolina at Chapel Hill.

Patton, M. Q. (2004). *The roots of utilization-focused evaluation*. In M. C. Alkin, Ed. Evaluation roots (pp. 276-292). Thousand Oaks, CA: Sage.

Patton, M. Q. (2008). *Utilization-focueds evaluation (4th Edition.)*. Thousand Oaks, CA: Sage.

Rodruguez-Campos, L. (2005). *Collaborative evaluations: A step-by-step model for the evaluator*. Tamarac, FL: Llumina Press.

Stufflebeam, D. L., and Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco, CA: John Wiley & Sons.

Part 3
Good practices and lessons learned from country-led monitoring and evaluation systems

MONITORING AND EVALUATION IN SOUTH AFRICA. MANY PURPOSES, MULTIPLE SYSTEMS

by Indran Naidoo,
Deputy-Director General,
Monitoring and Evaluation,
Public Service Commission,
Republic of South Africa

Background

This paper provides a reflection on the work and experience of the writer over the past 15 years, in the monitoring and evaluation sector of the South African government. It draws on his experience as manager of the country`s first departmental monitoring and evaluation unit[1] at the Department of Land Affairs (DLA), and later, senior manager at the Public Service Commission (PSC) of South Africa. The PSC[2] is a body mandated to oversee the performance of the country`s Public Service, and is constitutionally obliged to "investigate, monitor and evaluate" the Public Service, "without fear, favour or prejudice". It does not do recruitment, as many of its counterpart do, but focuses on monitoring and evaluation. This paper sets out to assess how monitoring and evaluation has evolved in the Public Service, and to highlight the linkages between monitoring and evaluation and democracy. It describes his experience in managing the monitoring and evaluation unit at the DLA, and also considers the growth of monitoring and evaluation at the continental level. In terms of the latter, it focuses on some of the experiences around monitoring and evaluation encountered by the PSC, which is one of the key monitoring and evaluation players in the Public Service.

---

1   The writer set up the country`s first monitoring and evaluation Directorate in 1995, which was expected to support the implementation of an ambitious land reform programme.

2   Republic of South Africa. The Constitution, Chapter 10

Monitoring and evaluation in the era of the post-1994 euphoria

Democracy was achieved in South Africa in 1994, when the country shed apartheid and held its first democratic elections. The apartheid Public Service existed at the behest of the apartheid regime, and implemented policies which were essentially anti-democratic and which entrenched values and practices that created a discriminatory and unaccountable Public Service. Given that the engine of the democratic transformation was the very same Public Service that implemented apartheid policies, it was necessary for this inherited bureaucracy to be reconfigured and revised to operate in a manner that reflected the democratic era.

Since democracy, there has been acceptance of the need for monitoring and evaluation within government, and this has now formed the basis for engagement with international bodies with whom the country needs to work with. The purposes of monitoring and evaluation were several; to support democratic engagement by promoting transparency and accountability, by providing evidence for planning and, in general, supporting the developmental State. The concept of monitoring and evaluation as it is understood now in South Africa is different from how it was used previously, where it was largely to support the Apartheid State.

With democracy, the South African Government aligned itself with international imperatives for good governance, and committed itself to the achievement of the United Nations Millennium Development Goals and to address the issues identified in the *Country Assessment Report* of the African Union Peer Review Mechanism. To this end, the South African Cabinet approved a set of 76 human development indicators and the Presidency has since published, on an annual basis, three editions of *Development Indicators*. The indicators are clustered into ten broad themes, namely: economic growth and transformation; employment; poverty and inequality; household and community assets; health; education; social cohesion: safety and security; international relations, safety and security and good governance. These indicators provide evidence-based pointers on the impact of government programmes on the lives of all South Africans. Based on these indicators, the Presidency also publishes, on a periodical basis, reviews which summarise data on trends, in order to track progress towards the implementation of government programmes. Currently, two such reviews have been published, namely: *Towards a Ten Year Review,* published in

2003 (South Africa) and *Towards a Fifteen Year Review,* published in 2008. It is evident from these reviews, that the country has embraced monitoring and evaluation, and that it views monitoring and evaluation as a critical tool and strategy for improving governance which, in South Africa, is based on the premise of a developmental State that is pro-poor and interventionist.

The process of changing the Public Service was a difficult challenge, and involved a process of rationalising many spatially carved out and racially defined governance structures, which supported the divide and rule ethos of the apartheid government. In the period leading to democracy and shortly thereafter, major policy reforms took place, the result of which are contained in the Constitution of the country, passed in 1996. This set the basis for a constitutional democracy, and paved the way for the reform of the Public Service. Contained within the Constitution are 9 values and principles (CVPS) for public administration. The CVPs have been used by the Public Service Commission (PSC) as a framework for measuring the quality of governance, and is its main framework for conducting its monitoring and evaluation work.

An important element of South African democracy was the establishment of institutions of democracy, listed in Chapter 9 and 10 of the Constitution. Whilst each of the institutions may operate in terms of particular legislation and emphasis, they are joined by the common purpose of ensuring that there is no regression to the apartheid era; that the benefits and freedoms of democracy are retained; and, that institutions such as the Public Service operate in an accountable manner. These oversight bodies have helped to establish in South Africa, the elements necessary for monitoring and evaluation, such as providing an enabling environment for scrutiny and accountability of persons and institutions. The re-defined Public Service is expected to be citizen-centric and to support the transformation policies of government.

The multiparty democracy in the country has provided an enabling environment for the institutions of democracy to operate unfettered by political interference. Currently the work of these institutions is regularly reported on in the media. The media contributes in a large part to the performance information produced on government, which supports the discourse on public sector performance, thus promoting good governance.

The interest in issues of public sector performance is driven by a highly politicised citizenry, who view the freedom to express opinions about government performance as part of the democratic package of benefits, hitherto denied. Through an independent and vibrant media, and the accountability framework which is in place, an interest in the performance of government is sustained. The PSC has responded to the high demand on government for performance information that comes from citizens, the media, parliament and from within the government itself. The demand from the citizens relates to their personal circumstances, how they receive services and the quality thereof, thus tools like Hotlines tend to be popular and reflects public sentiment on performance issues. The PSC has scanned the environment and differentiated its products and services, and engages intensely with role-players and stakeholders to remain topical. The PSC demonstrates in many ways how streams of performance information, generated by oversight bodies, can impact on decision-making, galvanise public opinion, and raise levels of transparency in support of accountability. It also demonstrates the link between democracy and monitoring and evaluation ,both of which share one commonality in that they promote transparency. Monitoring and evaluation activity can thus be seen as enriching for democracy, as it helps to deepen democracy by bringing higher levels of accountability and transparency into the Public Service. If monitoring and evaluation is located strategically within the decision and opinion-making architecture of the government, it can become more effective.

As an example of how monitoring and evaluation has worked, the experiences of the writer whilst at the PSC helps to illustrate some of the contextual and managerial issues which had to be worked through in setting up a monitoring and evaluation system. Many of these are pertinent in other contexts.

Monitoring and evaluation in the Department of Land Affairs

Importance of the land question for socio-economic transformation

The first monitoring and evaluation unit within a South African government department was started in the National Department of Land Affairs (DLA). This was a department, reformed in 1994, with new policies and which had a critically important political

mandate to reverse the legacy of apartheid that expressed itself in the skewed racial/spatial divide, much of which persists today in terms of land ownership patterns. A plethora of colonial and apartheid legislation largely dispossessed the majority Black population, confined them to disconnected, self-governing states (Bantustans), reduced agricultural potential of Black farmers, and impeded socio-economic progress. The minority White population, on the other hand, benefitted from these policies, gained access to most of the productive and well developed land, and were supported through state intervention to advance economically.

Against this background, the African National Congress (ANC) put in place policies and announced a land reform programme to address these inequalities. The land question was important in the run-up to the first democratic elections, and people saw the return of land rights and land restitution, as indicators of a successful democratic transition. The same department that produced and implemented apartheid land policies, now had to implement land reform programmes, and this was a major challenge from a capacity point of view. The new configured department had to bring in new people with new ideas and creativity, and reverse decades of entrenched land holding patterns and power-relations that existed across the country. The first Minister for Land Affairs in the democratic government, Derek Hanekom, managed to pass policies that would in practice bring about the envisaged changes. Implementation proved to be much more difficult. The monitoring and evaluation unit of the DLA was expected to, amongst others, reflect performance both within and outside the department, and in a context of high expectations, performance results became the source of interest and debate.

Installation of the monitoring and evaluation Directorate in the DLA

The introduction of monitoring and evaluation into the DLA came at a time when the new government was seeking ways to ensure its success, pivotal to which was having high quality performance information at its disposal. Monitoring and evaluation tended to be brought in to support programmes that had a high social and political priority, and where there were high expectations for these to succeed. In light of this, the Monitoring and Evaluation Directorate was set up in 1995 in the DLA, and received initial technical support from the Danish International Development Agency (DANIDA). This took the form of periodic support of a technical advisor, and

more on-going support from Non-Government Organisations (NGO), namely, the Land and Agrarian Policy Centre (LAPC). The NGO sector for land was very active in the early period of the land reform debates, and brought critical insights into the policy formulation process. This sector supported mainly rural communities, who bore the most severe brunt of apartheid policies. The widespread and brutally implemented apartheid land policies resulted in many of the social and economic problems which persist today. The land reform programme offered much hope, as the return of land was also about a return of dignity. However, in the context of a negotiated political settlement, and a market-led land redistribution programme, actual and potential beneficiaries saw the process as taking too long. The restitution programme, involved thousands of claims against the State, but included millions of beneficiaries, and resolving these claims required intensive research, which was complex and time-consuming. In this context of high expectations from the land reform programme, there was an equally high demand from the monitoring and evaluation unit to produce the information which would ensure its success.

The first 3 years (1995-1998) of the newly established monitoring and evaluation unit was taken up in staffing, building capacity and setting up systems to produce credible information on the land reform programme. Parallel information systems were necessary given that there was no coherent land reform system within the institution at the time. The unit set up needed to establish its own geographic information system so as to indicate on a monthly basis, through its Land Reform Barometer, how many hectares had been transferred and how many beneficiaries had been brought into the programme. This seemingly simple task involved intense monitoring of land transfers and scrutiny of project documents. It benefited the process by producing easily accessible maps which used colours and graphs to show progress, and which in turn made it possible for the political leadership and management to identify performance trends. The Land Reform Barometer soon evolved and became used for accountability purposes, from all quarters, and helped to direct attention to performance. A problem with presenting such stark performance data is that it is a-contextual, and when used for accountability purposes, can cause problems.

The question of information management remains central to monitoring and evaluation, and the setting up of information systems was a priority for the unit which used its own staff to gather information. It should be noted that land reform information was politically

sensitive, so establishing the unit's credibility was important since it formed the basis for the policy, programme and project assessments which were undertaken. The centralisation of the information system within the monitoring and evaluation unit was important in order to ensure adherence to agreed upon definitions and standards, so that the analytical reports would not be challenged on the basis of poor information. It must be remembered that many were questioning the success of the land reform programme. The monitoring and evaluation function can be rendered ineffective, if it does not rely upon a credible data system, and this must be prioritised in the establishment of any monitoring and evaluation function.

Apart from producing the statistical and geographic information, at different scales, the monitoring and evaluation unit, also produced a series of assessments on land reform projects, and which contained critical beneficiary perspectives. For this purpose, expertise was secured to build the capacity of monitoring and evaluation staff around participatory methodologies. Fortunately the unit recruited highly skilled staff and the unit was able to converse in all 11 official languages of the country, which facilitated community engagement. The reports on the pilot projects indicated at that stage that whilst policies were in place, there were serious implementation challenges; a point which was reinforced by a subsequent, externally undertaken assessments. After 5 years of implementation, a key finding was that inter-governmental coordination was weak (meaning that land was delivered without services), and there were many issues about whether the planning processes met the aspirations of communities, especially after land transfer. There were many questions raised about the sustainability of projects, a point documented in several internal and external assessments of the land reform programme.

At an operational level, the monitoring and evaluation unit had to also contend with its relevance being challenged. The key question was who "owned" the monitoring and evaluation function. The policy-makers saw it as necessary to provide information on how policy was proceeding based on implementation evidence. The implementers saw it as a voice for their discontent with policy, and a way to express their frustrations at the coal-face of delivery. Some saw it as just a tool just to produce information, without necessarily recommendations, as this was the prerogative of management who were the ultimate decision-makers. Others viewed monitoring and evaluation as an independent voice, which needed to express opinions about progress or the lack thereof. As can be imagined,

the monitoring and evaluation unit was constantly challenged about who its master was, and what its purpose should be. The staff within the monitoring and evaluation unit, who were still building up their own experience and knowledge, wanted to see quick reactions to their reports, and were frustrated when information went into a decision-making hierarchy where it was difficult to discern the contribution of monitoring and evaluation in decision-making. The problem was compounded by mixed messages sent to the monitoring and evaluation unit about its purpose, which is a common problem faced by monitoring and evaluation units. In the absence of a precedent for monitoring and evaluation in the country, the unit at times had to constantly reappraise itself, and try to meet several, and at time conflicting imperatives. This is often an institutional challenge for monitoring and evaluation units, and requires clear leadership and support.

One of the examples of trying to respond to a diverse and somewhat powerful grouping came in 1998, when the regional heads of the DLA, demanded their own monitoring and evaluation capacity. The views on this were simmering from 1996, with much discontent expressed by regional heads about what the monitoring and evaluation function was doing, which they saw as being too critical of their implementation efforts. They argued that monitoring and evaluation should be hands-on, and monitoring and evaluation should be a part of the implementation process, rather than writing reports from afar based on periodic visits. The management of the monitoring and evaluation unit decided to secure such capacity, and provided regional monitoring and evaluation officers. However, when this exercise was assessed after 2 years, it had not produced the regional monitoring and evaluation capacity envisaged. The monitoring and evaluation officers were still viewed as undertaking "policing". The intervention was probably not very successful because of the lack of experience in doing monitoring and evaluation, in an institution that was also developing its own managerial capacity. At the national level, the monitoring and evaluation unit, whilst ostensibly independent, fell within a managerial hierarchy, which it had to negotiate. Because of its access and reporting lines to the apex of decision-making in the organisation, the monitoring and evaluation unit was perceived as policing, and therefore seen as another form of scrutiny in a context where managers were doing their best to deliver under trying circumstances, but receiving undue criticism from the monitoring and evaluation unit. This experience points to the importance of addressing the location question,

but recognising that purpose should inform location and strategy. Monitoring and evaluation cannot be all things to all people, and this must be clarified and reinforced, or it can easily become blamed for problems outside its responsibility. The question of where to locate the monitoring and evaluation function, and under whose command, is often an issue in most organisations. There is a need to balance issues of relevance and support, with being sufficiently detached to remain independent and credible.

There are several lessons which can be drawn from the experience of the monitoring and evaluation unit at the DLA which are instructive for those establishing this function. There remains the classic challenge that context, which as the DLA example illustrates, affects strategy and design. If there was credible and useful information, little effort would have to be invested in designing information systems. Another area relates to the relationship between what management sees as their prerogative in directing the monitoring and evaluation unit, and what the management of monitoring and evaluation sees as important to produce. This relates directly to the question of location and authority. There will always be the reality of the hierarchy, and monitoring and evaluation units need to be supported in their work through higher level statements and actions. It is not acceptable for monitoring and evaluation units to have themselves to explain their purpose to strong voices, many of whom have their own agendas, especially when monitoring and evaluatioN also has an accountability dimension. There would be instances where negative performance comes to the fore and causes embarrassment, which must be anticipated. monitoring and evaluation practitioners often feel vulnerable and unprotected, which may account for the high turnover of monitoring and evaluation staff in many institutions.

There is a need to determine why and what one can get from different interventions. An example is the initial quality of life system that was designed externally and which the monitoring and evaluation unit was expected to implement. The unit did not see the value in the system. It was viewed as not being relevant, too cumbersome to implement and not owned by the implementers. Following a consultative national workshop, the system was overhauled to make it more relevant and to allow for the production of more useful reports. The experience indicates that context should determine the tools, and when there is any sense that the value of such tools is dubious, they should not be used. It is highly disrespectful to take up people's time by asking them questions when they fail to see

the benefit and receive no return on their investment because it does not solve their problems. The problem with the initial system was that it was predominantly quantitative and it was implemented in a context where these issues were not viewed as priorities.

Finally, leading monitoring and evaluation requires negotiation skills. Monitoring and evaluation cannot be all things to all people, and must differentiate itself by stating what its products and services are upfront, and prioritise. If this does not happen there is the tendency to expect too much from monitoring and evaluation, and expect monitoring and evaluation to be a panacea, or it may result in monitoring and evaluation being blamed for not dealing decisively with poor performance. monitoring and evaluation is not a stand alone. It is part of the management architecture, and when it interfaces with decision-makers it needs to be sensitive to the obvious vested interests. A key skill is the ability to communicate, which is often forgotten. It is important to realise that monitoring and evaluation is a human endeavour, which affects the moods and aspirations of people. Thus it is important to tread carefully.

The discussion now moves on, below, to describe some of the evolutionary threads for monitoring and evaluation at the continental and country level.

Precedents for monitoring and evaluation in South Africa and the continent

Monitoring and evaluation in South Africa became more pronounced from around 2000, although there had been monitoring and evaluation activity since 1995, but this was not co-ordinated. This was understandable in that the post-1994 period was one of policy formulation, reconfiguration and setting up of departments, and getting the engine of government moving. Apart from the DLA monitoring and evaluation function, monitoring and evaluation was largely absent in government in a formal sense during the early years. The exposure to monitoring and evaluation was largely confined to the few people who attended conferences outside the country, as an monitoring and evaluation association did not exist, although an evaluation network (EVALNET) organised monitoring and evaluation events. On the African continent, however, there were certain important initiatives led by international bodies that sowed the seeds for monitoring and evaluation, and which helped to introduce the monitoring and evaluation practice. Some of these galvanising moments are referred to below.

In November 1998, a conference was held in Abidjan, which brought together teams from 12 African countries and 21 international development agencies. It presented an opportunity for fostering networking amongst practitioners on the continent. This was followed by the inaugural conference of the African Evaluation Association (AFrEA), in Nairobi in 1999. Attended by over 300 evaluators from 35 countries. This conference helped to embed the monitoring and evaluation discourse more firmly on the continent.

Within South Africa, the first regional event was hosted in Johannesburg in 2000. This was organised by the Development Bank of Southern Africa (DBSA), in collaboration with the African Development Bank (AfDB) and the World Bank, around the theme of *Monitoring and Evaluation Capacity Development in Africa*. The event was attended by experts and resource persons with delegates from 8 African countries. It sought, amongst other things, to address the linkages between monitoring and evaluation and the development challenges in Africa; provided country perspectives on how demand and monitoring and evaluation infrastructure was built; addressed evaluation capacity development through new methodologies; and, provided African sector experiences on monitoring and evaluation. It also addressed the issue of developing national evaluation associations and opportunities for international cooperation (AfDB, Summary Report on Seminar Proceedings, November 2000).

A further watershed event was held in 2004, when the PSC co-hosted the 3rd AFrEA Conference in Cape Town, South Africa, under the theme; *Evaluation Matters, Africa Matters, joining forces for development*. This event drew 450 participants from 61 countries, and was characterised by the large representation from high-level members of African governments. As with other events of this nature, it offered over 20 pre-conference training sessions, as part of the capacity building initiative. The PSC in 2005 supported a development evaluation seminar in Tunisia, which was attended by representatives from 21 African countries. The 3rd AFrEA Conference was followed by two other events, in Niamey in 2006 (AFrEA IV) and Cairo in 2009 (AFrEA V). The AFrEA V event was significant, in that the African Evaluation Association (AFREA) became formally constituted, with potential to translate into a more vibrant continental association operating beyond just the biennial events where it was most visible.

In 2004, the idea of a *South African Monitoring and Evaluation Association* (SAMEA) was presented, and whilst a community of practice had been operating until then, the need for formalising this gained support and SAMEA was launched in 2006. In 2007, the first SAMEA Conference was co-hosted by the PSC under the theme *Evaluation in Action*, focusing on utility, and in 2009 the 2nd Biennial Conference, also co-hosted by the PSC was held under the theme: *eVALUation*, focusing on the values of evaluation. Both these events were attended by over 400 people, including government officials and monitoring and evaluation people, NGOs, academics, consultants, practitioners, and politicians, all of whom engaged over the 4 day event in training, networking and discussing the approximately 50 papers presented on the conference themes. A richer understanding of what monitoring and evaluation can and cannot achieve was gained from these events. For the first time, abstracts and papers were produced which spoke to the practice and indicated that there was now a stronger engagement with the discipline. As part of its policy for engaging productively with stakeholders, the PSC has entered into a memorandum of understanding with SAMEA, to take forward its commitment to public participation.

During the period from 2002 to present, there have been Presidential statements about the importance of monitoring and evaluation, the most significant being an appointment of a Minster responsible for this function, based in the Presidency. The African National Congress (ANC) 2004 Election Manifesto also committed to fostering monitoring and evaluation. Currently systems are being developed which will allow for performance to be reported on a set of key performance indicators. Once these performance outcomes are reported on, the status of monitoring and evaluation could be enhanced further in the country. In a relatively short period there was increased exposure to monitoring and evaluation through the hosting of various events, which have been largely PSC driven. The participant profile at these events reflects a high diversity of people, organisations and institutions, which augurs well for the growth of the discipline.

The next section looks at how the PSC`s monitoring and evaluation systems have evolved, and what their outputs and possible outcomes have been.

The Public Service Commission as a central driver for monitoring and evaluation in the country

Setting up a new identity

The PSC entered the monitoring and evaluation arena from 1997, when a new Public Service Commission Act was passed, built on the Constitution, which gave the revised PSC new powers and functions. It broke from its original 1912 remit, set up and modelled on the old British Civil Service Commission Act, to oversee appointments and prevent favouritism in public servant appointments. Although ostensibly independent, the institution under colonialism and apartheid grew into one that supported the executive, and hence its legitimacy would have been under question in the democratic era. By focusing on investigating, monitoring and evaluating, the PSC helped to galvanise support for monitoring and evaluation in the country, which as illustrated makes it a nodal point for monitoring and evaluation.

In operationalising this new identity, the PSC had to respond to the imperatives of the developmental State, which sees the South African State and government needing to intervene in all aspects of South African society, to redress the legacy of apartheid and bring about transformation and development. Since 2000, the PSC has taken the lead in developing a monitoring and evaluation discourse in the country, which it did by partnering with associations such as AFREA and SAMEA. More importantly, it designed and implemented several monitoring and evaluation systems to give effect to its constitutional mandate around monitoring and evaluation, and these have introduced a very high level of transparency around government performance, which has come about due to its many assessments, and the strategies it has used that cut across the media and civil society where it places its results and work in the public domain for engagement.

Using the constitutional basis expeditiously

The PSC has been set up to be the custodian of "good governance". It uses as its monitoring and evaluation framework the 9 constitutional values and principles (CVP) for public administration. The CVP can be viewed as a comprehensive definition of good governance in South Africa. Contained within the CVP are universally accepted

normative concepts for good governance, such as transparency and accountability. Also included are topical issues such as the fight against corruption (CVP 1: ethics) and human resources and transformation issues as contained in CVP 8 and CVP 9 respectively. Central to the CVP is the question of the judicious management of financial resources, which is contained in CVP2: the efficient, effective and economical use of resources. The benchmarks for performance are explicit, and the PSC has unpacked each CVP into specific monitoring and evaluation systems/programmes/interventions, resulting in a range of monitoring and evaluation products and services targeted at different sectors of society, and which seek to drive the good governance ethos in the country.

The table below sets out the CVP components and illustrates some of the work done in these areas by the PSC.

Table 1: Constitutional value and principles, engagement in area and products (selected examples)

Constitutional value and principle (CVP)	*Engagement in area*	*Products*
1. Professional ethics	Helped form the National Anti-Corruption Forum (NACF) in South Africa (civil society, business and government), currently serve as Secretariat Communication across society of the expectations of public servants, so as to raise awareness and prompt reporting of corruption. Through monitoring and evaluation systems, report on the management of disciplinary processes within departments Management of the National Anti-Corruption Hotline (NACH), allowing public access to a toll-free, multilingual, anonymous facility to report corruption Prevention and detection of conflicts of interest of public servants Investigating complaints received Production of policy on issues of ethics	Secretariat of the NACH, Resolutions and reporting thereof Code of Conduct produced for public servants Recovery of ZAR100 million, and dismissal of 50 officials. Received over 100 000 calls, and referred over 6000 for investigation Management of financial disclosures of civil servants, and scrutiny thereof Investigative reports resulting in action taken on errant public servants

Constitutional value and principle (CVP)	Engagement in area	Products
2. Efficiency, effectiveness and economy of the public service	Track and report on the quality of financial management in departments, and aggregated results into oversight reports Reporting on financial misconduct	Reports on the management of financial misconduct
3. A development orientated public service	Assess the extent to which departments engage in public participation processes, to ensure a citizen centric public service is in place	Reports on the extent to which departments engage in public participation practices
4. provision of services which are impartial, fair, equitable and without bias	Assess service delivery from citizen perspectives, by conducting Batho Pele (people first) surveys, as well as announced and unannounced inspections	Reports on adherence to the 8 Batho Pele principles
5. People's needs must be responded to and the public must be encouraged to participate in policy making	Assessment of departmental-public interface processes, to assess the extent to which policies are relevant	Assessment through Public service monitoring and evaluation System Analysis of complaints received
6. Accountable public administration	Processes which call leadership to account monitored, e.g. Management of the evaluation of Heads of Department, as well as implementation of all other monitoring and evaluation processes Reporting on results of departments to political oversight committees	Overarching reports such as the State of the Public Service reports indicates the effectiveness of the accountability framework Implementation of compliance systems
7. Transparency to be fostered through the provision of timely, accurate and accessible information	Through media briefings, roundtables, hearings, reports, the PSC puts into the public domain performance assessments of departments	
8. Good human resource management and career development practices to maximise human potential must be cultivated	Assessment of HR practices in order to see whether it has the desired effect Policy advice Adjudicate grievances referred to it after internal departmental processes have been exhausted	Adjudication of grievances
9. A representative public service	Assessment of progress against targets	Reports on state of representivity in the Public Service

From the above, it can be noted that the PSC has translated and concretised the CVP into interventions and products, which have been aimed at different client bases. These include:

- Parliament, the legislature and its committees.

- Political and administrative heads of government departments.

- Civil society, through its engagement via conferences, seminars etc.

- The academic sector, by presenting papers, soliciting for critical comment on reports (e.g. the State of the Public Service roundtables).

- Organised labour, by co-hosting the Public Sector Co-ordinating Bargaining Chamber (PSCBC) conference.

- The monitoring and evaluation community, by co-hosting the AFrEA 2004 event, subsequent support to AFrEA, and the co-hosting of the SAMEA 2007 & 2009 biennial conferences.

- Continental associations set up to promote good governance. It helped form the Association of African Public Service Commissions (AAPSCOMS), a body meant to support good governance in Africa.

- The media, through regular press briefing on its reports and findings.

This indicates an appreciation of the importance of building strategic alliances, and engaging with a cross-section of society. It has been used as a means to establish its own identity and credibility, and by being respected across all sectors of a diverse society, it is able to capitalise on the unique perspectives each sector/body brings, and thus provide more incisive assessments and advice. The PSC sees engagement, not detachment, as being a way to influence decision-making and public policy, and has built a track record to attest to the value of such a decision.

Critical areas of work

It is easy for any oversight body to become unfocused. The PSC has now reached a point where it has regular products, which are followed by engagement strategies to take findings forward. The following are examples of the regular, routine products of the PSC and these serve as a basis for producing performance trends.

- The annual *State of the Public Service* (SOPS) reports. Thus far 9 editions have been produced, with the last 5 being thematically written. These use the CVP as chapters, and within each produce an interpretation of the CVP as it relates to a theme; reflects on performance trends over previous periods; and, captures recent research and findings, from all related work on the subject. These provide policy directives going forward. Each of the chapters/CVP helps to emphasise why the area of performance is critical for the Developmental State and for reinforcing good governance. The report may be considered a meta-evaluation, and is the most widely cited and debated assessment of the State in the country. See www.psc.gov.za

- Individual departments are using the *Transversal Public Service Monitoring and Evaluation System* (TPSMES), which was designed using the CVP as a framework. It produces both quantitative and qualitative assessments of departmental performance, based on the application of a set of standards and indicators; interviews; document analysis; and, responses to interim reports. The assessment draws on other assessments, such as that of the Auditor-General, and in this way reinforces the work of the institutions of democracy, especially those dealing with the Public Service. The implementation process seeks to engage management at the inception and reporting stage. The reports end in performance scores given for each CVP and for the department as a whole. A set of recommendations is also provided, the implementation of which is tracked and publically reported on. The system uses performance bands (a five point scale from weak to excellent), and the performance of departments for each cycle is made public. Thus far 131 departments have been assessed using this system, and 6 consolidated reports produced.

- A set of reports are also produced on the management of financial misconduct, which is used to indicate the extent of financial corruption on how this has been attended to. Similarly, trends reports are produced on the management of grievances, the submission of financial disclosures etc.

- Fact sheets on compliance have proved to be a quick way to indicate poor performance, and this has worked in highlighting non-compliance in the areas of the filing of performance agreements and the conducting of the evaluations of heads of department. If performance management is not done properly at this level, it has implications for the rest of the Public Service.

- The feedback rate to the National Anti-Corruption Hotline (NACH) is a good indicator of the capacity of departments to investigate cases referred to it. In this regard the 2009 *State of the Public Service* report notes a decline in feedback, which is a concern in that follow-through on reported cases is critical for demonstrating the resolve to address corruption.

Supporting transparency and accountability

Whilst issues of transparency and accountability may be taken for granted in mature democracies, these remain pertinent issues in South Africa, where there was much distrust about the intention of the State. Now the Public Service needs to be citizen-centred and to subject itself to public scrutiny. In such a context, the implementation of monitoring and evaluation remains a challenge, for reasons similar to the ones mentioned in the DLA example described earlier in this paper. There is often agreement at the intent level about the importance of monitoring and evaluation, and a commitment to co-operate, but when the monitoring and evaluation process begins one can expect a level of resistance. Fortunately, the legal authority of the PSC has reduced some of this, as it has the power to summons, which it has used to ensure compliance. However, implementing monitoring and evaluation will always be a challenge as placing performance results in the public domain, which the PSC does, would obviously bring tensions to the fore.

The PSC has recognised that monitoring and evaluation has a strong accountability thrust, which the developmental State needs in order to instil new values, both in terms of how its institutions operate, and how its public servants behave, termed Batho Pele (people first). In light of this the accountability thrust has been expressed in large part through implementing systems which are compliance biased, using the assumption that the policies in place are inherently good and valuable for the transformation and development of the citizens of the country. Until compliance improves the accountability thrust of monitoring and evaluation must remain. The PSC's *State of the Public Service* reports indicate clearly that compliance remains a problem, which implies that monitoring and evaluation for accountability needs to remain.

This paper has spoken about administrative accountability, tested using the TPS monitoring and evaluations. In terms of the avowed intention of being a citizen-centric public service, the PSC has put much effort into developing monitoring and evaluation system to

assess how services are received. In the final analysis, the true test of whether the 15 years of transformation has impacted at the citizen level will be whether there has been improved service delivery. Various assessments indicate that services have improved, but the pattern is uneven across the country and varies, depending on which services one is considering. Through the development of citizen forum toolkits (which establish what the drivers of citizen satisfaction are); conducting assessments; conducting Batho Pele surveys (to see if all that has been promised is delivered); conducting announced and unannounced inspections; and, engaging with the NGO sector and other stakeholders, the PSC has produced reports which indicate what the quality of services are in the Public Service. Evidence from PSC assessments indicates that departments perform poorly when it comes to interfacing with citizens, and whilst during the post-1994 period citizens were at the centre of policy formulation, this has not been sustained. Service delivery protests, whilst not a perfect indicator of the quality of services received (given that many of these have been orchestrated), in some instances indicate that citizens are not happy with the services which they receive, and resort to such means to make their discontent known.

Conclusion

This paper has painted a picture of what monitoring and evaluation means for the Public Service, and has shown links to democracy, transparency and accountability. It has reflected on the importance of galvanising moments, namely conferences on the subject of monitoring and evaluation held under pertinent themes, and it has attempted to provide a candid view of the challenges in actual implementation of monitoring and evaluation. It has shown that monitoring and evaluation has advanced in the country, at least at the level of managers in the Public Service, by leading the practice through its practical work, and thus supporting a discourse on the subject of good governance. The journey is far from over, it has only just begun. What remains evident is that monitoring and evaluation, whether led from oversight institutions with authority, or from individual units, cannot afford to become complacent. monitoring and evaluation needs to constantly scan, engage, refine, communicate, debate; all in order to remain relevant in a world where there is much competition in the provision of opinions and evaluations. Finally, monitoring and evaluation is not a panacea, it is but a stream of information driven by a sound intent and supported by a credible

methodology. Whether it gets considered, depends on how well the potential recipients are prepared for results, and whether it can be communicated and taken through to make a difference, or, in other words, to demonstrate utility.

References

AfDB, African Development Bank, (2000). *Summary Report on Seminar Proceedings: Regional Workshop and Seminar on Evaluation Capacity Development for Africa*. Johannesburg, South Africa 25-29 September 2000 Tunisia Johannesburg.

Mashupye, H.M. (2002). *The Changing Role of the South African Public Service Commission During the Period 1990-2000: An Evaluation* An Unpublished Master of Public Administration Dissertation, University of South Africa: Pretoria.

South African Monitoring and Evaluation Association, (2008). *SAMEA Annual Report*, (2006/2007) SAMEA: Parktown.

Republic of South Africa. (2004). *Transversal Public Service MONITORING AND EVALUATION System* Public Service Commission: Pretoria

Republic of South Africa. (2003). *Towards a Ten Year Review.* The Presidency: Pretoria.

Republic of South Africa. (1996). *Constitution of the Republic of South Africa Act, Act No. 108 of 1996*, Republic of South Africa

Republic of South Africa. (2007). *Government Wide Monitoring and Evaluation Policy Framework* . The Presidency: Pretoria.

Republic of South Africa. (2008). *Development Indicators* . The Presidency: Pretoria.

Republic of South Africa. (2008). *2001-2008 State of the Public Service Report*, www.psc.gov.za

Republic of South Africa. (2008). *Towards a Fifteen Year Review Synthesis Report*. The Presidency: Pretoria.

Republic of South Africa. (2009). Public Service Commission of South Africa, *State of the Public Service*: Public Service Commission of South Africa, Pretoria

Republic of South Africa. *(*2009). *Development Indicators*. The Presidency. Pretoria

Republic of South Africa. (2009). *Improving Government Performance: Our Approach.* The Presidency: Pretoria.

Republic of South Africa. (2009). *Internal Report on the 2nd Biennial South African Monitoring and Evaluation (SAMEA) Conference*, 17-21 August 2009. Public Service Commission.

COLOMBIA'S NATIONAL MONITORING AND EVALUATION SYSTEM, SINERGIA[1]

by Manuel Fernando Castro,
Senior evaluation officer, The Work Bank,
and former Director (2002-2006), SINERGIA

by Diego Dorado, Director, SINERGIA

by Ana Carolina Ramirez, Advisor, SINERGIA

Introduction

Colombia's national monitoring and evaluation system (SINERGIA) is one of Latin America's more visible efforts for institutionalizing monitoring and evaluation at the government level. Established in 1994, following the mandates of the National Constitution, the system has endured numerous periods of progress, stagnation, and setbacks, due to institutional, political, and fiscal challenges faced by the country over time. Nonetheless, SINERGIA has attained a significant level of consolidation and customization and is often referred to as a good practice system by multilateral organizations, donor agencies and academic analyses.[2]

The monitoring and evaluation system adopted by Colombia has placed a strong focus on strengthening the public management techniques which orient the public administration dynamic towards performance based management. This has implied a series of reforms in the normative and institutional framework, such as the *Law of the National Development Plan* (Law 152 of 1994), which establishes the key elements to take into account in the process of public policy monitoring and evaluation. The Law also defines the strategic guidelines and priorities of the Plan as the reference point for this process. To lead this process, the Constitution and the normative framework appointed the National Planning Department as the entity in charge of the technical and operational duties regarding the construction and consolidation of the system of monitoring and evaluation.

---

1 This document draws on a recent World Bank paper written by one of the authors that discusses all these issues in more detail. See World Bank ECD Working Paper Series No. 18

2 See World Bank-IDB 2006; Guerrero 1999; Ospina 2003; Zaltsman 2006; Mackay 2007; or Villarreal 2007.

Compared with experiences in other countries, SINERGIA's institutionalization process is remarkable for the combination of high-ranking formal mechanisms (e.g. constitutional mandate and laws covering the whole of the public administration) with the development of informal practices in key areas of the public sector (e.g. monitoring and evaluation activity in planning offices, programme units, and managerial controls from the President's Office). It is also notable for an increasing utilization of results information – particularly by the President's Office and some key agencies – and by the strength it has added to formal accountability and transparency of national programmes and sector policies.

SINERGIA is, however, not fully institutionalized and yet it faces enormous challenges. Most of them relate to the acknowledgement of monitoring and evaluation as key instruments inside the public administration, as well as the assessment of the progress that the national administration has achieved in terms of the modernization of its processes in a more efficient scheme. Some of the system's institutional arrangements have also begun to show the need to create a policy framework which defines the incentive structure that will help to overcome the disadvantages related to a limited evaluation market, limited funding for evaluation, and the limited quality of information and targets definition.

This paper provides an overview of the Colombian monitoring and evaluation system's most salient features. It highlights success factors and obstacles encountered during implementation, and identifies a number of lessons that might be of interest to evaluation practitioners and officials implementing monitoring and evaluation systems in developing countries.

Institutional base and operating arrangements

Since the early 1990's Colombia has achieved significant advances in terms of the design and implementation of an institutional framework oriented towards performance-based management. In fact, the initiative for the construction of the system comes from the Constitution itself, making it not only a mandatory issue, but a strategic instrument for Colombia's public administration. However, in order to construct a normative framework which facilitates not only the design, but also the use of the performance-based management instruments, such as monitoring and evaluation, the design

and implementation of the system has required a series of structural reforms.

The most significant reform is related to the *Law of the National Development Plan* (Law 152 of 1994), which establishes the key elements to be taken into account in the process of monitoring and evaluation of public policy. The Law also requires that this process should take as reference the strategic guidelines and priorities of the Plan. To lead this process, the Constitution and the normative framework, appointed the National Planning Department as the entity in charge of the technical and operational duties regarding the construction and consolidation of SINERGIA.

SINERGIA's main objective is to influence policy processes by providing timely and quality performance information to relevant decision-makers. It therefore seeks to ensure the use of results information, as a management tool as well as strengthening internal and external accountability. Accomplishment of these objectives has made coordination a paramount requirement, between different technical and political levels of government (e.g. the President's Office, ministries and agencies). Notwithstanding the ups and downs of the relationship between stakeholders, the objectives of SINERGIA have managed to persist over time, improving its scope and focusing primarily on the consolidation of the culture of performance-based management at the national level.

In recent years, SINERGIA has been developing an institutional framework oriented to the achievement of this objective. The institutional structure acquired for the system is based on the idea of involving the relevant actors of the policy-making process in the implementation of a monitoring and evaluation working plan. Figure 1 shows the institutional framework that is operating nowadays.

Figure 1: SINERGIA's Institutional Framework
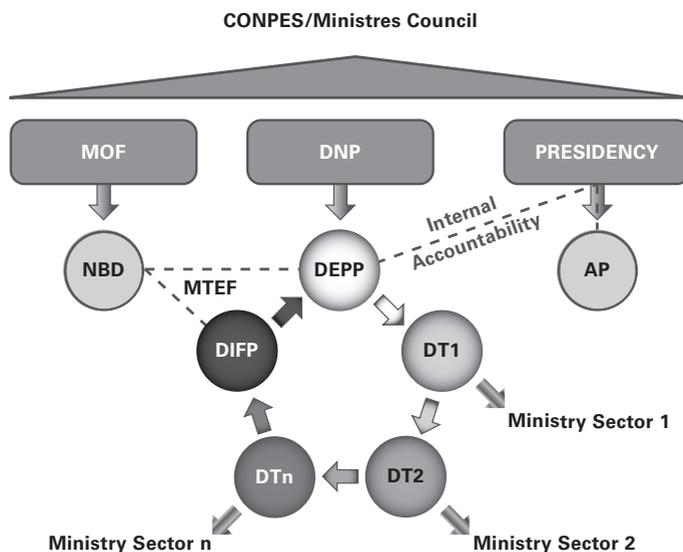
SINERGIA has three main actors which are: the Minister of Finance (MOF), National Planning Department (DNP) and the Presidency. Each actor plays a key role in the achievement of the system's objective and for the coordination between them a coordinator is appointed in one particular office of each agency.

The system's technical arm is the National Planning Department. Within the Department, the Office of Evaluation of Public Policy (DEPP), which is one of 10 technical offices (DTn) of the Department, holds the technical secretariat of SINERGIA. This Office is in charge of leading the process of monitoring the *National Development Plan* and the design and execution of the evaluation's agenda. Also within the National Planning Department is the Office of Public Investment and Finances (DIFP), which is in charge of leading the process of formulation and implementation of the investments projects, at the national level. This office has a specific role in the execution of the *National Development Plan*, because a significant part of the goals set out in the plan will be delivered by the implementation of investment projects. This makes the allocation of the capital budget a decision process that must be supported by performance information. Both the Office of Evaluation of Public Policy and the Office of Public Investment and Finances are headed by a technical director, reporting directly to the National

Planning Department's General and Deputy Directors who have the status of Minister and Vice-Minister, respectively.

The Office of the President's Advisor (AP)[3] plays a key role in the institutionalization of monitoring of the *National Development Plan* targets achievement, because the Advisor to the President has the power to confront the minister's regarding their performance in fulfilling the set targets. Depending on the progress reported by each ministry in relation to the targets, the Advisor identifies the particular cases in which the intervention of the President himself would help the ministry to get back on track towards fulfilling the target. This type of intervention is known as management control and is one of the instruments used by the administration to implement internal accountability.

The Advisor to the President also participates in the definition of the evaluation agenda by identifying the political priorities which need to be evaluated in order to guarantee achievement of the *National Development Plan* goals.

The participation of the Ministry of Finance is through the National Budget Office (NBO) and focused primarily on the formulation of the Medium Term Expenditure Framework. With the participation of the Office of Public Investment and Finances, the National Budget Office leads the process of formulation of the framework. This framework is a 4 year rolling plan that includes current and capital expenditures at the national level. The Office of Evaluation of Public Policy plays an advisor role in the budgeting process, focused on supplying performance information to be taken into account in the allocation of resources.

The challenge for this budgeting process is to guarantee that the evaluations produce timely and quality information to enrich the decision-making process for the allocation of resources. Currently, the three actors inside the budgeting process are working in the development of a protocol and a management evaluation methodology, so that in the future the Medium Term Expenditure Framework will become a strategic planning tool.

To ensure the required coordination, the system has an Inter-Sectoral Committee for Evaluation and Management for Results. This committee consists of representatives from the national planning and budget authorities, the President's Office, and line ministries and

---

3   Since 2002, this office is in charge of the Agenda of the Ministers Council.

agencies of key programmes requiring evaluation. The Committee reviews and approves an annual evaluation agenda which establishes evaluation priorities and the methods to be applied. So as to ensure utilization of evaluation findings in policy-making, as much as harmonization among planning, budgeting, implementation and evaluation processes, the Committee also oversees ongoing evaluation processes and identifies the eventual course of action to be proposed to the government.

The participation in the Committee by line ministries and agencies is meant to ensure use and appropriation of evaluation at the sector policy and budget formulation levels.[4] Sub-national governments, Congress, control agencies, and civil society organizations are given an explicit, but not binding, role as external users of the system's performance information, which once produced is made publicly available through a variety of means (e.g. website, virtual reports, printed reports to Congress, media releases, seminars, etc.).

SINERGIA's components and monitoring and evaluation tools

SINERGIA's components and monitoring and evaluation tools have sought to provide solutions for different performance information problems, although, over time, not all of them have been successful, thus requiring redesign and improvements. The underpinning starting point is an explicit differentiation, both conceptually and operationally, between monitoring, evaluation, and accountability activities. SINERGIA's organizational arrangements are defined by this differentiation which also establishes a single performance management language among the different agencies and stakeholders, all with a view to facilitating progress toward institutionalization.

Each of these components has its own set of methodological tools, but with strictly complementary mechanisms. These components are discussed below.

Results-based monitoring

SINERGIA's monitoring component attempts to provide systematic performance information on the *National Development Plan*. Since 2002, it has set out base lines, performance indicators and targets

---

4   See CONPES 3294 of 2004.

for each of the programmes set down in the Plan . This process facilitates on-going reference to the level of achievement of the goals set by the executive.

- *Government Targets Monitoring System (SISMEG)*[5]

To undertake the monitoring task, SINERGIA developed a monitoring tool called "Government Targets Monitoring System" (Sistema de Seguimiento a Metas de Gobierno –SISMEG). This tool is basically a technological web-based platform customized to the scope and technical needs of SINERGIA. As such, performance information captured and administered by SISMEG is structured around the higher-level strategic objectives of the *National Development Plan.* SISMEG seeks to respond, primarily, to the President's information needs for exerting managerial control over the executive and for external accountability towards the general public and Congress.

Given the scope and purpose of SISMEG, all monitoring information is captured directly at the source, that is to say, at the ministry and agency programme units. These units are made publicly responsible for the information which they provide. A manager – whose name is made public through the website – is assigned to each goal (usually to a group of goals) in order to strengthen accountability for the information reported.

Based on the records, SINERGIA's technical secretariat (the Office of Evaluation of Public Policy) controls and validates information quality. For that purpose, a team sector specialists within the Office is responsible for maintaining continuous dialogue, with agencies and goal managers, on the quality of performance data loaded into the system. The results-monitoring component thus operates as an integrated, inter-institutional systemic function, which includes quality parameters, rules for updating data, and coordination mechanisms for producing, administering, reporting and using performance information.

SISMEG monitoring database is currently a powerful source of government data. Programme trends between different administrations can be determined, whilst maintaining an up-to-date inventory of indicators. Statistics from SISMEG are used by the President's Office in press releases and in reports to Congress;

---

5   This system was developed with support of UNDP, through the use of the SIGOB system. However, at this moment, even though in Colombia people continue to call the system SIGOB, the database, protocols and architecture were develop according local requirements.

in development plan evaluations and monthly reports prepared by the National Planning Department; in sector ministers' meetings and reports; and, in some Ministry of Finance budget documents submitted to Congress, although the later is less common. Despite all of these improvements, the availability, timeliness and quality of information still remain a challenge.

- *Mechanisms for setting goals and indicators*

Contrary to general perception, it is not easy to draw up goals and indicators. A technique and a method are required, as well as clearly defined uses for the information. In SINERGIA's early years, low relevance for policy makers (too much operational focus) and a disproportionate number of indicators, resulted in bad quality (little relevance) and low usage of performance data. Lack of explicit mechanisms to define goals in National Development Plan documents was also a major constraint.

Currently, SINERGIA's goals, targets and indicators are set with every ministry and national agency during the *National Development Plan* formulation process, and the approval of each minister, as well as of the President at the Council of Ministers, is required. Once approved, goals and indicators are widely disseminated in printed reports prepared by SINERGIA and the Office of Evaluation of Public Policy so as to keep a record of the government's commitment.[6] Since 2006, an annex to the Plan with goals, indicators and responsible units is also included.

In general, SINERGIA's definition of indicators, targets, goals and base lines seeks to ensure three methodological criteria:

(i)   strategic focus on *National Development Plan* objectives relevant to the President, to citizens, and decision-makers;

(ii)   technical and political validation with agencies, with responsibility for each of them assigned to an official; and

(iii)   communication of government priorities and benchmarks so as to enhance coordination and harmonization while strengthening agencies' and programmes' focus on results. Box 1 provides an overview of the process, identifying the five main steps.

---

6   See "The Goals of the National Development Plan 2002-2006: A Commitment of the Uribe Administration." National Planning Department. August, 2003.

Box 1. Setting performance indicators and using monitoring information

Step 1. Definition and selection of indicators

- A proposal is drawn up by staff of the Office of Evaluation of Public Policy and by sector/agency planning and budgeting offices.
- The proposal is submitted to ministers and agency directors.

Step 2. Discussion and validation

- Discussion and agreements are undertaken at individual meetings with ministers and directors.
- Relevance of the respective goals and indicators is assessed, taking into consideration:
    (a) funds allocated in previous years, and actual performance (baseline);
    (b) current budget allocations;
    (c) technical and institutional capacities for reaching the goals;
    (d) agencies' control over result variables; and
    (e) potential restrictions to performance.
- Agreed *National Development Plan* indicators and goals are presented to the National Council for Political Economy and Society (CONPES)/Council of Ministers for approval.

Step 3. Disclosure:

- Publication of *National Development Plan* goals and Indicators booklet.
- Submission to Congress, universities, and the media for reassuring public commitment.
- Indicators loaded onto website for the monitoring system of government's goals, with names of managers responsible for reporting.

Step 4. Monitoring

- Goal managers record progress information, directly pointing out the factors determining performance.
- Public explanation provided for shortfalls in performance (town hall meetings and ministry councils).
- Staff of the Office of Evaluation of Public Policy monitor indicators for identifying obstacles, interacting daily with staff responsible for the programmes.
- The President holds periodic individual management control meetings with ministers and managers.

Step 5. Reporting

- The Office of Evaluation of Public Policy draws up printed reports, selective electronic reports, and the President's Annual Report to Congress.

- *Management control meetings with the President*

One of the most powerful mechanisms for developing and consolidating the monitoring and evaluation system has been the periodic progress review meetings, held since 2002 by the President with his ministers. These meetings are also attended by deputy ministers, who are sometimes supported by their technical managers. Staff from the Office of Evaluation of Public Policy participate in all meetings and provide technical assistance.

The meetings are executive management style, which is reflective of the current President's personal results-oriented style. There is therefore some risk that the current arrangement will not continue when a new administration comes into office. However, the value of this practice has already extended beyond the central government since many sub-national authorities are already using similar mechanisms in their own administrations.[7]

Government Targets Monitoring System facing a new National Development Plan

The experience of SINERGIA in leading the monitoring of national development plans at the national level extends for almost 3 presidential periods, using SISMEG in the last 2 years of monitoring. From this experience many 'lessons learned' have been identified. These include: the importance of consultation during the design of the indicators; the routines established to inform on progress towards targets; the naming of the staff responsible for the fulfillment of the targets; and, a positive and constructive approach to analyzing the performance results.

Currently, SINERGIA is facing an unprecedented challenge in the democratic history of the country, which is the assessment of a *National Development Plan*. Thanks to the presidential re-election, there has been continuity of the public policies established in the *Law of the National Plan*. It will also face the formulation of a new national development plan, which will be accompanied by new monitoring tools designed to overcome the obstacles noted in the past and to reinforce the strength of SISMEG.

---

7   Specific cases of this are the cities of Pasto and Medellín, but many other towns and provinces have developed their own versions of this mechanism.

Links between results and funding

The allocation of resources scheme used in Colombia is based on a model of counterweights lead by the Ministry of Finances and the National Planning Department. Both agencies start with the financial forecast in the medium term fiscal framework; they proceed to do an indicative allocation using the medium term expenditure framework, later to be included in the general national budget.

Connected to this scheme, the National Planning Department has designed a process for analysis and prioritization of resources, which is based on the interaction between the DTs and their corresponding sectors. To achieve these, technical offices interact with the Office of Evaluation of Public Policy, in order to use the information produced by monitoring and evaluation as a variable that will help them to better prioritization their investment. The result of this exercise is endorsed by the Office of Public investment and Finances and the National Budget Office, with the technical leverage of the Office of Evaluation of Public Policy.

Once this process has been implemented it requires some adjustments for using evaluations results in the allocation of resources. For this purpose, the participation of the Office of Evaluation of Public Policy in the allocation process must be seen to guarantee more effective evaluations.

The evaluations inside the public policy cycle

The development of monitoring and evaluation inside SINERGIA starts with an understanding of the public policy cycle, and merges this with the value concepts associated with State interventions to provide public benefit.

The cycle, understood as the process by which all the public policies are built up and executed, is especially relevant for the development of SINERGIA during two phases, which are monitoring and evaluation. The monitoring focuses on the assessment of the progress towards meeting the targets set for each policy. The evaluation enables the identification of causalities and good practices which will help improve the current policies as well as the policies of the future.

Figure 2: Public Policy Cycle
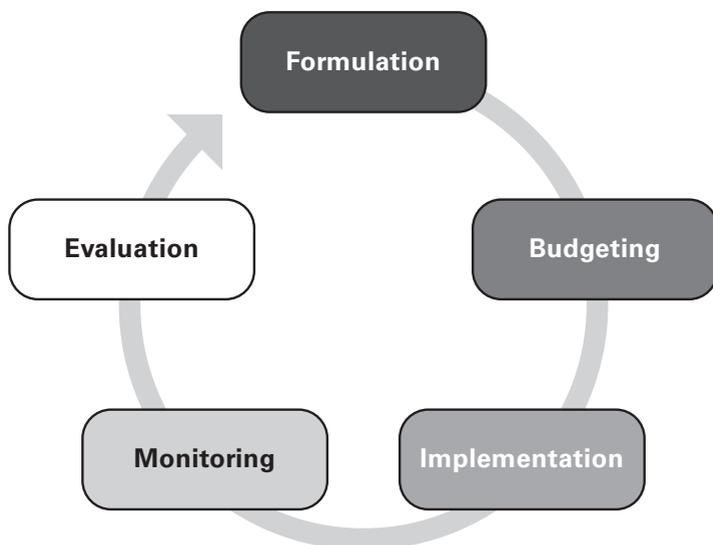
The 'chain value analysis' helps to identify the links through which the public policy will generate results. It separates the steps which involve a greater public manager role, and it identifies where the resulting benefits will make the intervention most relevant. This enables the scope of SINERGIA's evaluations to be determined according to the links to be analyzed and the hypothesis to be proven, all of which in turn will be the result of a well understood public policy process.

Figure 3: Types of evaluations
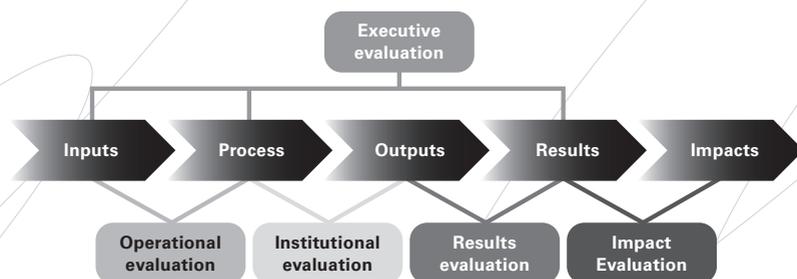
SINERGIA has now identified five types of evaluation, with different scope and target units of analysis. Some of these types of evaluations are similar in terms of the methodological techniques used in the analysis, but the differentiation is relevant, because this affects the formulation of recommendations.

Below is a brief description of each of the five types of evaluation:

1) **Executive evaluation**. This type of evaluation studies the programme from its design stage and analyzes the linkage between the design and the implementation process in order to assess the achievement of the goals and fulfillment of the outputs which are supposed to be delivered to the target population. The main source of information for this type of evaluation is administrative records and programme documentation.

2) **Operational evaluation**. This type of evaluation does a deep analysis of the programme's macro and micro-process. The idea is to use the findings on the projects as inputs to guarantee that the outputs are being produced in the most efficient way.

3) **Institutional evaluation**. This type of evaluation studies the institutional framework behind the programme and analyzes the structure of incentives and the organizational structure created for a programme in order to operationalize it.

4) **Results evaluation**. This type of evaluation focuses on the analysis of the improvement in the programme in terms of welfare of the population beneficiaries from the programme. It also studies the effects that the delivery of outputs has on the population.

5) **Impact evaluations**. This is the most rigorous type of evaluation, because it is based primarily on the construction of experiments to determine the comparative effect of the programme intervention on an individual within the programme compared to an individual not receiving the programme's intervention.

SINERGIA'S evaluation agenda

With the introduction of new methods, the number of evaluations increased significantly over time. However, the total number can still be considered low compared to the total number of national programmes (as determined by the programmatic classification of the investment budget); the total number of investment projects; and especially, in relation to the actual use of the findings for policy-making or budgeting purposes.[8]

As for other national systems[9], increasing the number of evaluations has been a gradual process for SINERGIA. The supply of evaluators at the country level was, and is still, limited and could not be expanded overnight. In addition, a larger evaluation agenda also implied additional organizational capacities which SINERGIA simply did not have and which could only build-up over time.

To overcome these limitations SINERGIA adopted a contracting out schema that facilitated the expansion of evaluation activity and, in addition, helped to ensure evaluations were conducted impartially so that their findings and recommendations could be independent and credible. External evaluators – from prestigious national and international universities and firms – were selected, based on their expertise and proposed evaluation methodologies. The estimated cost of the proposed evaluation was of secondary importance in the selection process.

If the planning agency was already distant from implementation activities (because it was not involved with the execution of programmes), by adopting this approach SINERGIA instituted a de facto firewall that further separated the programme implementation functions from the contracting and supervision of evaluation activities. Through this approach SINERGIA-Office of Evaluation of Public Policy came to play an intermediary technical role between programmes and evaluators, to relieve the programmes of the operational burden that evaluation on a permanent basis implies.

---

8   Indicators to measure the extent of evaluation activity in Colombia have mainly included the percentage of the National Investment Budget and the number of evaluated programmes. Like all performance measures of this kind, these indicators present methodological problems associated with the amount of earmarked resources which are not necessarily subject to evaluation through conventional programme evaluation methods. Limitations in the programmatic classification of the budget could also prevent determining the precise number of total programmes.

9   For instance, Chile's Management Control System, run by the MoF, and one of the most advanced systems in LAC performs about 14 evaluations a year.

This approach guaranteed that programmes would not exert any distorting influence on the evaluative judgements. To make this mechanism operational, SINERGIA established a team of highly qualified professionals with the exclusive responsibility for technical and administrative interaction with evaluators. This mechanism was well received by agencies, since programme officials, usually overly burdened with implementation activities, had been able to give evaluation the priority and attention it requires.

Developing evaluation capacities is another core mission of SINERGIA. To undertake this task, apart from delivering conventional training courses and seminars with academic and multilateral organizations, mechanisms were established for rapid transferring of skills to national evaluation professionals. Contracting arrangements involving joint ventures between highly qualified international experts and national research centres, firms, or universities were given higher scores in procurement processes. Knowledge and technology-transfer channels were thus automatically set-up, and in a few years Colombian firms were already carrying out top-quality evaluations on their own. The cost of conducting complex evaluations also fell, since local professionals are typically less expensive than international ones. This, in turn, made it possible to perform more evaluations.

The linkage between evaluation and other monitoring efforts

As mentioned in the overview, SINERGIA is a system dedicated to the monitoring and evaluation of public policy. Even though there are marked differences between monitoring and evaluation, it is important to understand that both of them complement each other.

For example, the information which results from monitoring is very important for the construction of the evaluation agenda. This is because the use of indicators to oversee the management of the *National Developing Plan* goals will allow the analyst to identify any stage in the delivery chain that is not working. With this in mind the scope of the evaluation can be oriented towards understanding what is happening in implementing that stage and its affect on the final results.

Colombia has, in the last decade, significantly improved the supply of information systems available for monitoring of public programmes and investment projects. Nowadays there are three information systems oriented towards follow-up of programmes and

projects, so as to guarantee that their execution corresponds with the established work plan and that the set goals will be achieved. These systems are:

i)   SIGOB: the monitoring system of government's goals;

ii) SUIFP: the unified public investment system; and

iii) SISCONPES: the monitoring systems of political guidance.

The use of evaluations

As indicated throughout this document, the evaluations are designed and executed with the aim of giving feedback on the public policy cycle. With this in mind, the idea is that the evaluations focus on the understanding of the causalities of outputs, results and impacts, so that the public policy implementation can become a more efficient process. Therefore the stakeholders who will use the evaluation results are, in addition to National Planning Department, the policy-makers, the programmes' directors and the Ministry of Finance. Of these, the National Planning Department is the agency in charge of the long-run strategic planning in Colombia, and of the formulation of the public investment budget.

However, it is important to note that the stakeholders' interaction is led by the technical secretary of SINERGIA which is the Office of Evaluation of Public Policy within the National Planning Department. For this purpose they have created three committees as follows:

i)   the Committee of Public Policy Evaluation in Social Protection;

ii) the Committee of Public Policy Evaluation in Competitiveness and Business Development and;

iii) the Committee of Public Policy Evaluation in Modernization of the State.

These three committees are formed in order to guarantee that the agenda of evaluation, and the analysis of the evaluation results, are commensurate with the macroeconomic reality of the country, the allocation of public resource and the strategic information needs of each sector. In this way it is possible to ensure that the evaluation projects are relevant and timely, and that the type of evaluation design selected is the best possible methodological choice in order to obtain the information requested by the policy-maker.

Additionally, it is worth mentioning that as SINERGIA is inside the government, the evaluations are contracted out to external consultancy firms, think-tanks or universities, in order to ensure the independence of results and policy recommendations. However, the design of terms of reference for tenders is the responsibility of the Office of Evaluation of Public Policy team with the supervision of the above mentioned committees.

Since the evaluation projects are executed by third parties it is important to establish review protocols to ensure good quality information and data input to the evaluation. For this purpose, SINERGIA has established a correction process, led by the corresponding committee with assistance from a high profile professional, expert in the application of evaluation methodologies. This correction process helps to ensure that the project outputs are useful and coherent with the aim of closing the public policy cycle.

The indicators and variables used for the analysis of the information input to the evaluation undergo a rigorous review protocol. This review protocol is oriented towards guaranteeing that the instruments used to asses a programme are comparable with the existing data, have no information bias and are representative of what one wants to know.

In complement with the above, SINERGIA has constructed a data base of peer reviewers, which are mainly academics recognized for their expertise in the study of the sectors' behaviors and of the impact of State intervention on long term trends. These reviewers study the evaluation project report and develop a technical note containing their impressions of the study. The content of the note is shared with the members of the corresponding committee and with the team undertaking the project.

Until now, this type of exercise has been very important, because the review is in itself a means of verifying the quality of the analysis and the information produced by the evaluation project.

Finally, the results of the evaluations are presented in seminars and discussion events, in order to spread the policy recommendations formulated from the results of the evaluation, and to inform on how the programme is adjusting the implementation processes, based on the feedback received with the evaluation.

Reporting and accountability

As a supplement to the monitoring and evaluation functions, reporting and accountability is the third component of SINERGIA. With a view to institutionalizing dissemination channels and accountability methods, this component seeks to ensure that the general public can access performance information and that publicity becomes a high power incentive for data providers to ensure information accuracy. The four instruments detailed below are central to this reporting and accountability process.

Annual report by the President to Congress

For many years the annual President's report to Congress was a mere formality in Colombia. The reports were no more than a compilation of unrelated documents submitted by the different ministries to the President's Office before being sent on to Congress in July. Following government guidelines on public sector modernization and transparency, re-designed annual reports to Congress became one of SINERGIA's central products. Key features of the redesigned report include: [10]

- linkage between election proposals, government goals, and actual achievements;

- sequential assessment including a review of recent past and current state of issues, prospects and specific targets;

- review of both achievements and setbacks;

- reader-friendly presentation, simple language and graphic illustration of sequences;

- a short document accompanied by an executive summary; and

- consistent information and widespread dissemination.

The media often disseminates sections of the report, especially administrative flaws, but opinion journalists have made more balanced references to it. The President's communication office disseminates its content widely. The user-friendly format helps to convey the messages effectively to officials and the general public. However, there are indications, based on external analyses contracted out by SINERGIA, that utilization of the report by Congress and the media is still low. This aspect seems to be strongly

---

10   The reports started in 2003 and are available electronically at: http://www.dnp.gov.co/PortalWeb/PND/InformealCongresodelaRep%C3%BAblica/tabid/210/Default.aspx.

connected to the fact that SINERGIA is run by the government and, therefore, to the perception that it cannot be entirely trusted. This, in fact, is considered to be a major weakness of the report.[11]

Semi-annual and quarterly reports

Semi-annual and quarterly electronic reports facilitate SINERGIA's dissemination of performance information. These reports make evaluation information (not monitoring information alone) available more quickly so that feedback from universities, think-thanks, practitioners, etc., can be gathered. The frequency of these reports has changed over time to meet the assimilation and production capacities of different stakeholders, and also in response to the availability of relevant information. In addition, since 2008, the Office of Evaluation of Public Policy has prepared monthly reports for the President's Office, which rank ministries and agencies according to their accomplishment of the government's four-year goals.

Currently, SINERGIA's reporting reflects growing efficiency in processing, administering, and consolidating performance information. Such efficiency gains are due largely to the monitoring system of government's goals, which currently provides online information as soon as administrative or statistical records are produced at source.

Town Hall meetings and Accountability Councils of Ministers

Two additional tools are the current government's weekly Town Hall meetings and the annual Accountability Councils of Ministers. The Town Hall meetings are a public control and accountability mechanism, which promote interaction among national government, regional authorities, and the general public. Subjects of interest to each region are discussed at the meetings, and accountability is checked in terms of progress made on national programme targets in the regions.

The Accountability Councils of Ministers are held each year also under the direction of the President. At these sessions, each minister presents his results to the country in a televised programme and answers questions from the public. The televised councils are arranged by the President's Office, and the SINERGIA Office of Evaluation of Public Policy helps to ensure that results information is accurate and internally consistent.

---

11   See Castillo, C. "Communication Strategy for Improving SINERGIA's Dissemination of Results and Accountability towards the General Public," diagnostic document. December 2007.

Alliance with Civil Society Organizations

SINERGIA has also forged alliances with civil society organizations aiming to stimulate demand for the system's performance information and to create additional mechanisms for public and independent scrutiny of the performance information it produces.

Based on the experience of the successful Bogotá, Cómo Vamos? programme[12] – an accountability initiative promoted by the private sector in the country's capital – SINERGIA endorsed and facilitated the development of the Colombia Líder (Colombia Leader) initiative. This initiative is promoted and funded by the business community, think tanks, and the media, and has been underway since 2007.[13] Currently, Colombia Líder assesses the results of national and sub national poverty reduction programmes and to that end, among other sources of information, it is one of the main external users of SINERGIA's performance data.

Current developments and next steps

Colombia has achieved considerable success since the early 1990s in the creation and strengthening of the national results-based management and evaluation system, SINERGIA. Much of that success has been achieved during President Alvaro Uribe's two term administrations, and reflects the deep commitment of the current government to social accountability, good government and state modernization.

Even so, considerable work remains to be done. A number of actions are currently being implemented to achieve greater cost-effectiveness from SINERGIA's monitoring and evaluation information and other data from existing information tools[14]. Primary amongst the challenges identified for the system is the movement towards better

---

12 *Bogotá Como Vamos?* (Bogotá, How Are We Doing?) monitors changes in people's quality of life, emphasizing compliance by the district administration to the National Development Plan. The project is an alliance between *El Tiempo* (the main national newspaper), the Corona Foundation, and the Chamber of Commerce. The monitoring is done in terms of access to quality goods and services, and takes public perceptions into account. See www.bogotacomovamos.org.

13 *Colombia Lider* (Colombia Leads) is a private initiative for monitoring and evaluating state results and government policies. It involves Semana Publications, the Colombian Banks Association (Asobancaria), the Antonio Restrepo Barco Foundation, the think tank Fedesarrollo, the United Nations Development Programme, the Friedrich Ebert Stiftung (FESCOL) Foundation and the RCN radio and television channels. See www.colombialider.org.

14 Mainly, SIGOB, SUIFP, SIIF, SISCONPES and a number of sector systems.

informed budget and decision making processes using the information produced through monitoring and evaluation.

In terms of monitoring, the big challenge is to institutionalize the culture of performance-based management inside the government agencies, so that the quality of the information they report to the system is improved as it becomes a management tool. In the same way, the formulation of the indicators and goals which will be included in the next *National Development Plan* must be the result of a rigorous exercise in identifying the delivery chain behind each objective and goal set.

Achieving this approach will permit the government targets monitoring system to produce alerts on target fulfilment and in this way enrich internal accountability with timely information to allow for a more efficient policy decision-making process.

As mentioned in this document, the advances that SINERGIA has registered in terms of evaluation are remarkable. Nonetheless, there are three main challenges to overcome:

(i)   development of the evaluation market;

(ii)   creation of an incentive structure that will permit a better allocation of resources for evaluation; and

(iii) development of a processes oriented to institutionalize internal and external accountability.

On the first point, the development of the evaluation market is a priority because the agenda for evaluation is becoming bigger and more strategic. This means that the competitive firms in the market today will probably not have the capacity to absorb all the evaluation projects contemplated in the agenda. Therefore, it is necessary to create a structure of market incentives to enlarge the market competitiveness and to permit it to embody top line mechanisms for public policy evaluation.

Regarding the allocation of resources for evaluation, the government agencies today face a trade-off in terms of how to distribute the resource assigned to them. This has made them prioritize programmes which will help them to achieve their goals, giving less importance to the evaluations which would help them understand the causality behind the achievement of the goal. Until now the evaluations have mostly been financed by the multilateral banks and international cooperation, and little has been done with the public budget. This is not sustainable in the long run.

Finally, as mentioned above, President Uribe's administration has achieved a lot through ad hoc internal accountability instruments, and there are many elements which can be drawn upon in order to embed routine use of monitoring and evaluation in support of performance based management.

Lessons learned and conclusion

A number of lessons for other countries can be drawn from SINERGIA's experience. However, specific circumstances drawing primarily from political, institutional, and cultural contexts of each country should not be overlooked. The lessons include:

- Institutionalizing a government monitoring and evaluation system is much more than a technical endeavour. From the outset, it needs to be considered as a long term public sector reform that, to be successful, will need to affect public officials' behaviours.

- As a major reform, it also needs to be in line with broader political objectives (e.g. democratization, governance, accountability, etc.), and to make such concepts operational in public management routines.

- Given the long-term character that monitoring and evaluation reforms entail, regulations can provide a firm basis for ensuring sustainability during government changes.

- The principal assets of such a system are a technically sound design and robust methodologies for drawing up performance indicators, goals, and baselines. Information quality, and hence credibility, relies on these aspects. Achieving such assets requires skills and ability to combine local and international knowledge.

- Indicators should be focused on variables which are relevant to the head of government (President, Governor, etc) and decision-makers, as much as to citizens. Understanding the delivery chain is a prerequisite for success in that regard.

- Information systems should be differentiated from institutional monitoring and evaluation systems themselves. The former should be adapted to the operational needs of the institution if adapting the institutions to the tools is to be avoided.

- A coordinating body with strong technical capabilities and influence is a necessary condition, but is not enough by itself to ensure institutionalization. Developing a system requires management,

a sense of opportunity, and strategic vision, not to mention support from entities with sufficient rank and political power.

- If continuity is to be achieved in implementing a government-wide system, evaluation policy guidelines with clear objectives and well-defined incentives are necessary. Also required is constant dissemination of such guidelines, within and outside government.

- Strong planning and budgeting frameworks, with different authorities co-existing side-by-side can hinder institutionalization. Organizational arrangements with clear rules of operation are a necessary requirement to overcome such obstacles.

- Development of evaluation capacities is a must. Intervention of monitoring and evaluation authorities is needed to increase the supply of qualified evaluators, either directly (through training programmes) or indirectly (through contracting incentives).

- Demands for monitoring and evaluation information are the driving force behind a system but they do not occur on their own. Demand should be enhanced by a supply of usable data and by ensuring that it reaches key users. Performance information is an imperfect market; public intervention is therefore needed.

- The credibility of a system is built around the quality of its monitoring and evaluation products, and not necessarily on its formal independence. The fact that managerial feedback requires active participation of government agencies should not be overlooked.

- Re-valuing the importance of accountability to Congress based on monitoring and evaluation work requires the use of simple formats and language as much as strategies for communication and dissemination.

- Monitoring and evaluation systems implementation needs a balanced and rather opportunistic pace. While there is a need to reduce the costs of progressing too quickly (e.g. inability to assimilate) it is also necessary to minimize the cost of implementing too slowly (e.g. loss of interest).

For the past 15 years, SINERGIA has been the main reference point for Colombia's reform to a performance-based administration. The system has endured many changes to regulations and official policy by various administrations, and many of its tools have become increasingly accepted practice in the public sector, especially in the central government. Most effort has been aimed at introducing

monitoring and evaluation practices into the planning and budgeting processes, especially in central government. However, the results-based planning model has superseded the budgeting model, owing mainly to the architecture of the central administration and the fact that the system is institutionally based within the National Planning Department.

Compared with similar experiences in other countries, SINERGIA's institutionalization process is notable for the way in which it has combined high-ranking, wide-ranging formal mechanisms (e.g. a constitutional mandate and laws covering the whole of the public administration) with the development of informal practices in key areas of the public sector (e.g. monitoring and evaluation culture in ministries and entities). Considering the lessons from international experience, Colombian officials have followed a progressive and timely implementation strategy closely aligned with broader political and institutional reforms.

The analysis presented in this paper has shown some of the steps taken by Colombia to implement a national monitoring and evaluation system since 1991, highlighting some of the achievements and obstacles encountered as well as next steps envisaged by its current administration. SINERGIA can be categorized as a whole-of-government monitoring and evaluation system. Its design, implementation, and reform have been based on best international practice, with significant innovation and self-development incorporated as it evolves. Significant effort has been made in the country and implementation of the system can be considered a success in many ways. However, much more work will need to be done, and more creativity will be required, to fully institutionalize monitoring and evaluation at all levels of government.

References

ANIF (National Association of Financial Institutions). (2007). *Evaluación del Plan de Desarrollo 2002–2006: Resultados y Credibilidad.* Comentario Economico. Octubre. Bogota, DC.

Castro, M.F. (2009). *Insider Insights: Building a Results-Based Management and Evaluation System in Colombia.* ECD Working Paper No. 18, IEG, World Bank, Washington DC.

Consejo Nacional de Política Económica y Social. (CONPES). (2004). *Renovación de la Administración Pública: Reforma del Sistema Nacional de Evaluación,* Documento 3294, Bogotá, DC.

Curristine, T. (2006). *Performance Information in the Budget Process: A Comparison of the OECD Countries*. In Fiscal Transparency. Proceedings of a SINERGIA/World Bank/Inter-American Development Bank and International Monetary Fund conference. Bogotá, DC.

Departamento Nacional de Planeación. (2003). *Informe del Presidente de la Republica.* Presidencia de la Republica. Alvaro Uribe Velez al Congreso. Bogotá, DC.

Departamento Nacional de Planeación. (2006) *Presupuesto de Inversión por Resultados.* Presidencia de la Republica. 2004, 2005, y 2006. SINERGIA. Bogotá DC.

International Monetary Fund. (2005). *Colombia Public Investment and Fiscal Policy*. FAD. April. Washington, DC.

Mackay, K. (2007). *A Diagnosis of Colombia's National M&E System – SINERGIA.* ECD Working Paper No. 17, Independent Evaluation Group, World Bank, Washington DC.

World Bank. (1997). *Colombia: Paving the Way for a Results-Oriented Public Sector*. Washington, DC.

THE NATIONAL MONITORING AND EVALUATION SYSTEM IN SRI LANKA. EXPERIENCES, GOOD PRACTICES, CHALLENGES AND THE WAY FORWARD

by Velayuthan Sivagnanasothy,
Director General, Department of Foreign Aid and
Budget Monitoring, Ministry of Plan Implementation, Sri Lanka.

by Vaidehi Anushyanthan,
Assistant Director (Planning), Department of Foreign Aid
and Budget Monitoring, Ministry of Plan Implementation, Sri Lanka.

Background

Governments are challenged to respond to the urgency of citizens' needs and to be more accountable to them. All over the world, the public is putting government under growing pressure to show that they are providing good results and value for money. In the past 3 to 4 decades significant effort and budgetary resources have been deployed to accelerate development; reduce poverty; ensure equality; and, to improve social living standards and the quality of life of the people. With fast tracking of public investment programmes, the country has been able to maintain its average annual economic growth rate at 5%. However, the country experienced challenges in translating this economic growth momentum into poverty reduction. The level of poverty remained, respectively, at 26% and 22% in the past two decades. This situation led to the need to strengthen the planning, monitoring and evaluation system of the government to focus on the delivery of outcomes and impacts beyond the traditional output focus. The concept of 'managing for development results' has therefore been adopted gradually since mid 1990s. Managing for development results (MfDR) is about public sector management and the concept strongly emphasizes that shared vision, clear goals and measurement of results will lead to a better future. MfDR is a change management process that emphasizes a shift in focus from inputs, activities and outputs to outcomes and impacts. It promotes the concept of "accountability for results".

The Government of Sri Lanka (GOSL) fully recognizes the growing international consensus that MfDR[1] is an essential aspect of good governance to improve development efficiency and effectiveness; transparency; accountability; and, informed decision-making. In the recent past, globally, monitoring and evaluation expanded and diversified in many contexts and with many uses such as: decision-making; organizational learning; knowledge base; programme improvement; policy development; impact/outcome assessment; improved service delivery; accountability; performance audit; empowerment; and, even transformation. Ambitious government systems, to meet multiple stakeholders' needs, tend to achieve most of these desired uses. A good monitoring and evaluation system should go beyond institutional boundaries to cover national, sectoral, programme and project level in order to ensure results orientation within government.

The GOSL as a signatory to the *Paris Declaration on Aid Effectiveness* (March 2005), is committed to government-wide institutionalizing of MfDR. The MfDR process typically involves several phases such as: articulating and agreeing on objectives; selecting key performance indicators (KPIs); setting targets; monitoring performance; and, analyzing and reporting on results against targets. It also facilitates management to take timely corrective actions. Different countries will approach MfDR in different ways since it is not a "one size fits for all" model. Ultimately it should lead to sustainable improvement in the country's development outcomes. The GOSL is committed to promote the application of MfDR principles at various levels namely: the national; sectoral; agency; and, project level, and at various stages namely: planning; implementation; and, post-implementation stages. At the planning stage the results oriented country owned National Development Plan and sectoral plans are being aligned with the medium term expenditure framework and the use of a performance budget instead of a line item budget has been practiced increasingly. Line ministries are required to justify their budgets with well defined output/ outcome indicators. Today what counts is not so much how many clinics have been built, but whether citizens health has improved; not how many schools have been constructed, but how many girls and boys are better educated.

---

1   The 2004 Marrakech Roundtable on MfDR noted that the following principles as central: (1) focusing the dialogue on results at all phases of development process; (2) aligning programming, monitoring and evaluation with expected results; (3) keeping measurement and reporting simple; (4) managing for, not by, results; and (5) using results information for learning and decision making as well as for reporting and accountability.

Sectoral plans use outcome based key performance indicators in setting targets and directions. In order to institutionalize performance budgeting systems, the Department of National Budget and Treasury has revised its budget circulars to focus on 'results'. The Government's three year medium term expenditure framework incorporates outcome based key performance indicators to justify public expenditure. This helps to allocate resources rationally in a results oriented manner. Similarly national audits are gradually moving toward 'performance audits' or 'value for money audits' with more attention on economy, efficiency and effectiveness.

Managing for development results: a whole of government approach

The institutionalization of managing for development results (MfDR) is seen as a major shift in focus in the monitoring and evaluation arrangements in Sri Lanka. MfDR includes a management cycle of: direction setting; planning; implementation; and, delivering and reviewing results. The reviews then feedback into the cycle to improve future planning and so make ongoing improvement. The following diagram represents the localized version of the MfDR process in Sri Lanka.

Figure 1: "Managing for development results" cycle

Direction Setting
What do we intend to achieve over next 3-5 years?
Identify vision, mission, thrust area, goals and KPIs

Planning
What is the best way to achieve this?
Identify strategies and actions

Implementation
Are we implementing as planned and managing risks effectively?
Deliver interventions and Monitor Progress

Review
What impact have our interventions had?
Assess efficiency, effectiveness, impact and sustainability

Continuous Improvement

In a Sri Lankan context, the organizations who are managing for results must:

- Have a clear vision of what they want to achieve.

- Plan their work keeping in mind the clear vision and mission.

- Deliver what they planned in a manner consistent with public service ethics, values and standards whilst meeting standards such as timeliness, quality, quantity and within cost.

- Track their progress by monitoring; measuring and evaluating; learn from success and failure; and, make continuous improvement.

The MfDR approach adopted in Sri Lanka is more of a "whole-of-government" approach, covering national, sectoral, institutional and project level. A results focus is being built into the national development strategy, sectoral plans, ministries and projects.

Figure 2: Results focus in "whole-of-government" approach

351
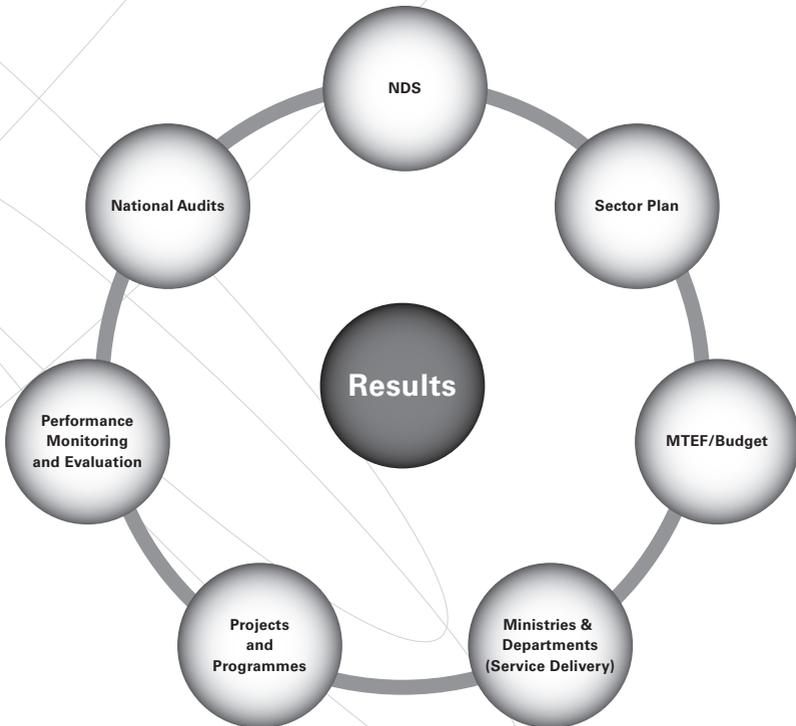
Enabling environment

In Sri Lanka the government's monitoring and evaluation of projects, programmes and development initiative is not a new phenomenon. Efforts to improve plan and project implementation have been a feature of development efforts since the early 1970s. One of the special features in Sri Lanka is the functioning of a separate Ministry of Plan Implementation charged to serve "as a national focal point for monitoring and evaluation of all government development projects and programmes, to ensure achievement of results and development effectiveness". The Ministry of Plan Implementation, is headed by a very senior cabinet minister and is mandated with the responsibility for monitoring and evaluation of all government policies, programmes, projects and institutions. The Department of Foreign Aid and Budget Monitoring is the key functional arm of the Ministry of Plan Implementation and with skills and technical know how to undertake the monitoring and evaluation functions. The Ministry is technically supported by the Department of Foreign Aid and Budget Monitoring. This dedicated institutional arrangement places monitoring and evaluation more strongly within the institutional setting of the Government of Sri Lanka.

One of the key features in Sri Lanka is the top level commitment for monitoring and evaluation with a focus on MfDR. In the 1990s, with the technical support of the Asian Development Bank (ADB), the post-evaluation system was strengthened in the Ministry of Plan Implementation . As a result, a number of, projects and programme post-evaluations, were conducted by the Ministry. In the late 1990s, the UNDP provided technical support in a large way to strengthen the Results Based Monitoring and Evaluation System in Sri Lanka. This enabled the government officials at the national and sub-national level to understand and recognize the importance of results focused monitoring. Also many positive factors had contributed to positive enabling environment, such as: political will; overarching policy; coordination of information collection; flow of information from line ministries and projects to the Ministry of Plan Implementation's Department of Foreign Aid and Budget Monitoring; strengthening of the electronic Information Management System in the Ministry of Plan Implementation's National Operations Room; and, demand for information for decision making. However, concerns such as the capacity of government agencies; the large number of ministries; and, the resultant coordination issues, are some of the challenges that need to be addressed.

The Ministry of Plan Implementation has introduced MfDR in government to track the line ministries, and their programmes, development results. 'Logical framework analysis' and 'results frameworks' are being used increasingly within government in planning and in monitoring and evaluation arrangements. All these developments clearly indicate that the concept of 'managing for results' set-out in the *Paris Declaration* is being institutionalized and is gradually moving forward in Sri Lanka.

The Ministry of Plan Implementation and its Department of Foreign Aid and Budget Monitoring, championed the institutionalization of MfDR within government in the belief that the process has potential to ensure results by shifting the focus of public sector management, from inputs and activities, to outcomes and results. Many initiatives have been moving forward on a gradual basis since the early 1990s. These include: institutionalization of post-evaluations; process evaluations and impact evaluation; mainstreaming of project performance management system; results based management; performance budgeting and performance auditing; localization of Millennium Development Goals (MDGs); and, results reporting through country-based monitoring and evaluation systems. There exists a strong link between the Ministry of Plan Implementation, the Ministry of Finance and Planning, and the President's Office, and they work together very closely on monitoring and evaluation strategy. The line ministries, agencies, project management units, Provincial Councils, and District/Divisional Secretariats work very closely with the Ministry of Plan Implementation and provide progress reports, on a periodic basis, in specified formats. His Excellency the President chairs the progress review meetings which are held on a regular basis to assess: progress of development projects; programmes; sectors; ministries; and, institutions. This high level forum serves as a guiding and trouble-shooting forum with top level political commitment.

The Cabinet of Ministers made a formal decision and directed the Ministry of Plan Implementation to monitor all development projects over Rs 50 million, executed by the line ministries, and to submit quarterly progress reports to the Cabinet of Ministers. The Ministry of Plan Implementation has developed an on-line, real time web-based electronic monitoring and reporting system to collect, analyze and report progress of all the development projects over Rs 50 million. Progress Review meetings are held by the Ministry with project management units and officials of line ministries on projects which are behind schedule. These meetings address the issues to ensure efficient and effective implementation.

The MfDR principles are being increasingly applied in the national monitoring and evaluation systems and related areas. Some of the major achievements and lessons are discussed below.

Government-wide performance measurement system and score cards

The GOSL has introduced MfDR, with the technical support of the UNDP, to track performance of ministries and institutions using output and outcome indicators. This new management style holds ministries/ departments and public officials accountable for results rather than 'efforts'. At a macro perspective, ministries and departments receive budget appropriation from Parliament to carry-out a specific mandate. This mandate has to be translated into detailed management expectations. The results framework of the ministry or department sets out the breath, depth and meaning of the management expectations. By measuring performance against the management expectations set out in the results framework, the institution will be in a better position to make an objective assessment of the results achieved. MfDR translates the vision and mission of public sector agencies into the practicalities of better management of the organization, at every level. MfDR otherwise know as 'results based management' involves defining the expected results; monitoring and evaluation of progress towards the achievement of expected results; integrating lessons into management decisions; and, reporting on performance. MfDR, when in operation, ensures the "line-of-sight", which implies that everyone in the organization understands the strategic vision and mission irrespective of their level and position in the organization.

In Sri Lanka a comprehensive performance measurement system was piloted in 2006/7 within 4 key line ministries (education, health, agriculture, highways). The selection of pilots was initially based on the interest shown in the reforms and the commitment of the leadership of the ministries to the change management initiatives. A range of activities such as awareness programmes; advocacy and sensitization to policy makers; and, training programmes, were conducted with the technical support of the UNDP and ADB. Currently, MfDR has been expanded and is operational in 35 line ministries. An MfDR core group has been established to drive this initiative, with the leadership of the Secretary of the Ministry of Plan Implementation. The experiences of the North American results based management models operated in Oregon (Oregon

benchmarks), Minnesota (Minnesota milestones), Virginia States, and the Canadian models were taken into consideration in developing a localized model. The performance tracking system in Sri Lanka will be expanded to all ministries.

The Ministry of Plan Implementation has established a web-based MfDR platform in the National Operations Room (the information arm of the Ministry of Plan Implementation), to maintain an agency results framework and score card/report card for each line ministry. The centerpiece of the government's MfDR is the "agency results framework" and "agency score card". The agency results framework sets out the mission of the Ministry and its core business (thrust areas, goals, key performance indicators) with baselines and medium-term targets. The customized score cards/report cards will appear on the computer screen to enable the assessment of the performance of the line ministries. The score card employs a red-yellow-green grading system to track performance of government institutions. This 'dashboard' serves as an 'early warning' signal to alert the Ministry when there is slippage in the achievement of targeted outcomes expressed by the key performance indicators.

Figure 3: Snapshot of the agency results framework for the Ministry of Health

AGENCY RESULTS FRAMEWORK

**Thrust area 1 – Curative and preventive health services Strategic Objective –
Provision of comprehensive Health service delivery and health actions**

No	Goal	Key Performance Indicators	Base line 2006	Targets				Dimension
				2007	2008	2009	2010	
1	Reduction of infant mortality	Infant mortality rate	11.2/1000LB (2002)	11.0	10.8	10.6	10.4	Effectiveness
2	Reduction of under five mortality	Under five mortality rate	4.4/1000 Under five population 1997 (AHB 2003)	4.0	3.8	3.6	3.4	-do-
3	Reduction of maternal mortality	Maternal mortality rate	38/100,000 LB (2004) (FHB)	36	35	33	31	-do-
4	Reduction of prevalence of underweight (malnutrition) children under	% of underweight children (Wt/Age) under five years	29.4% (2000) DHS	25%	24.5%	24%	23%	-do-
5	Reduction of incidence of low birth weight babies	% low birth weight babies	17.6 (2005) AHB	16.5	16	15.5	15.5	-do-

A snapshot of the agency results framework for the Ministry of Health is presented above. Under the thrust area of curative and preventive health services, you may see the goals and key performance indicators such as: infant mortality rate; under-five mortality rate; maternal mortality rate; percentage of under weight children; etc. The medium term key performance indicator target in the agency results framework enables the assessment of performance of the Ministry of Health in the relevant thrust area.

The framework is not carved in stone, it will continue to evolve as ministries/departments strengthen their management practices. The MfDR is to be used as a means to improve ministry/departmental results reporting to the Cabinet and to the Parliament. The score card raises a "red flag" when something is wrong.

Figure 4: Agency performance measurement framework
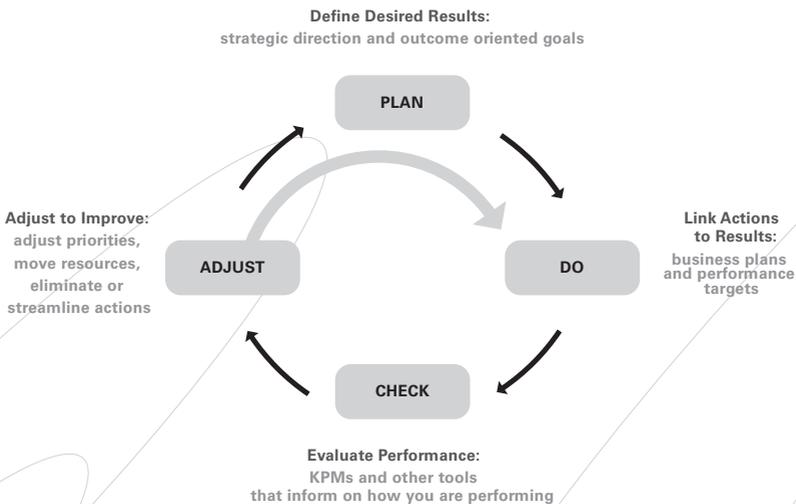
The Sri Lanka MfDR model follows the performance measurement framework of North American models and is further articulated through a plan–do-check-adjust cycle, as depicted in the agency performance measurement framework (fig 4). At the planning stage, the desired results are defined. Performance expectations are defined to capture key dimensions of performance which are of interest to important stakeholders. Agreement of expectations implies a shared vision of what is to be accomplished.

At the "do" stage, the strategy/actions of the ministries are linked to the desired results.

At the "check" stage, performance is evaluated using the key performance measures

Finally resources are adjusted and streamlined with the strategies and action to achieve desired results.

As in many developing countries, availability of trained monitoring and evaluation personnel is a key constraint in Sri Lanka. Also, incentives need to be in place to reward success. MfDR when fully operational, will help government to strengthen performance accountability and improve the continuous learning culture.

Readiness Assessment

With the support of the Asian Development Bank (ADB), a readiness assessment on MfDR was conducted in key ministries and agencies to assess the capacity gaps. Based on the readiness assessment findings, a capacity development plan has been developed to prioritize capacity building initiatives in the selected institutions. The spiral diagram below is a representative sample which depicts the preliminary results of the readiness assessment in 9 core areas.

The capacity development plan helps to address capacity gaps at the institutional level to enable the implementation of MfDR.

Figure 5: Overall analysis of Readiness Assessment for a sample institution

1. The Operating Environment
2. Clarity of Results, Mandate and Organization purpose
3. Adequacy of Resources
4. Strategies and Policies
5. Structure
6. System and Processes
7. Organization Culture and Values
8. Vision and Leadership
9. Monitoring Mechanisms
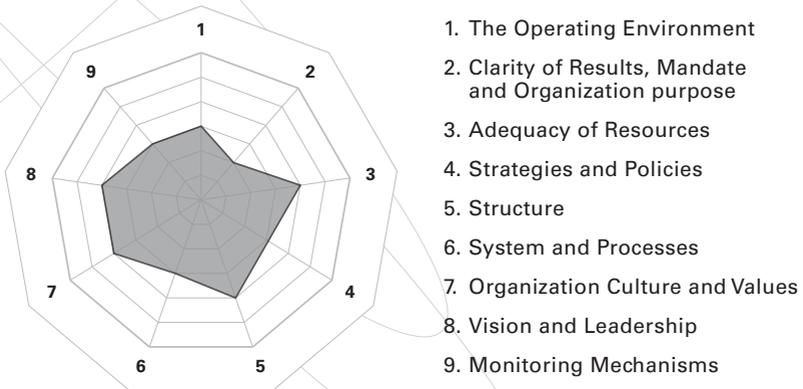
Based on the readiness assessment, the capacity development plan included initiatives such as strengthening the operating environment, clarifying the organizational results, results alignment to systems and processes, and upgrading MIS for improved data collection, analysis and reporting.

Implementation of managing for development results in government

The Ministry of Plan Implementation placed MfDR in the policy arena by initiating a note to the Cabinet on MfDR, by the Hon. Minister of Plan Implementation, outlining the importance and the steps necessary to institutionalize MfDR on a government wide scale. The responsibilities for the institutionalization process were assigned to the Ministry of Plan Implementation. The technical back-stopping of the Department of Foreign Aid and Budget Monitoring was complemented by the advisory technical support of the UNDP and ADB.

The institutionalization of MfDR in Sri Lanka involved sequential steps and processes which are described below.

The launching of the process involved many key sequential initiatives which include: the establishment of a core group to drive the initiative; endorsement of government policy on MfDR; formulation of a government-wide strategy; formulation of a "change agent" concept, to establish ownership and leadership in line ministries and agencies, to take the initiative forward, supported by advocacy and capacity development to sustain the initiative.

The approach adopted by Ministry of Plan Implementation in launching the institutionalization of the MfDR is discussed below.

Championing the Initiative

The Ministry of Plan Implementation, as an apex ministry responsible for monitoring and evaluation, championed the institutionalization of MfDR government-wide. A core group was headed by the Secretary, Ministry of Plan Implementation and comprised of representatives from the Ministry of Finance and Planning; Department of National Planning; Department of National Budget; the Auditor General's Department; the President Office; the Prime Minister's Office; Department of Census and Statistics; and, selected Secretaries of the line ministries. The core group was to provide

guidance, direction and leadership to the initiative of mainstreaming the MfDR in government. It also functioned as a think tank to formulate strategy and the action plan. The concept of a core group helped to establish government-wide commitment and ownership to the process. The core group formulated the strategy and action plan to institutionalize MfDR with technical inputs from the Ministry of Plan Implementation's Department of Foreign Aid and Budget Monitoring.

Managing for development results as a government policy

The Cabinet of Ministers approved the MfDR initiative and empowered the Ministry of Plan Implementation, as a lead organization, to lead this initiative government-wide. The top-level political support, which comes through the endorsement by the Cabinet of Ministers, indicates the political will and policy commitment to take this initiative forward. The Ministry of Plan Implementation initiated a planned programme of advocacy and sensitization at the policy level. Dr. Ray Rist, Advisor to the World Bank, was invited to conduct an initial orientation and sensitization programme which helped to share international experience and to raise within government, the importance of results based management/MfDR at the policy level. A number of programmes were conducted at policy level to sensitize the policy makers and senior government officials on the importance of the MfDR initiative.

Managing for development results – strategy and action plan

The strategy provided phased approaches including a pilot phase involving 4 line ministries and the Ministry of Plan Implementation. Based on the lessons of the pilot phase, the MfDR institutionalization was expanded to cover 35 line ministries. A comprehensive strategy and plan of action was developed which included: capacity development; staff training; advocacy; strengthening information systems; methodology; and, reporting arrangements.

Change agent concept – leadership at all levels of government

The Ministry of Plan Implementation provided the technical support for introducing the MfDR methodology and approach. Each Ministry was encouraged to establish a steering committee of senior

officers to take forward this change management initiative by functioning as 'change agents'. The change agents functioned as drivers of change and helped to facilitate the MfDR initiative at the ministry Level. The change agent concept clearly shows the importance of leadership at all levels of government to take this effort forward.

Adoption of a process approach – consensus building

It is important to note that in each case, the agency results framework of each ministry was developed by that ministry through a participatory process with the Ministry of Plan Implementation technical guidance. Each ministry assumed ownership and the process approach helped to create this ownership. Workshops were held in the line ministries and involved all key players. These consultations gave a sound basis for ensuring wider ownership. They also helped to tap the expertise, skills and experience on the one hand, and the trust, support and network on the other.

Buy-in

The process approach, and the involvement of senior government officials and other relevant stakeholders, created wider ownership. This contributed as a positive factor for buy-in, so ensuring sustainability of the initiative.

The Ministry level MfDR training workshops involved active participation of the top, middle and lower level staff of the institutions. This participation enabled the operationalization of the concept of 'line of sight'. Moreover, the other stakeholders were also consulted in this process.

As the agency results framework of the line ministry is to be closely integrating into the national budget to make it a 'performance budget', the initiative has built in factors of sustainability. Also, the link with the Auditor General, to measure performance through 'performance audit' arrangements, further strengthens the MfDR initiative as a national process. Expected results are mutually defined and agreed upon through a consensus building process involving all major stakeholders. This enhances the stakeholder's sense of ownership.

Linking resources to results – budget call 2010

In order to strengthen the result based budgeting in the public sector, the MfDR initiative taken by the Ministry of Plan Implementation has been linked to the annual national budgeting

process starting from 2010. The budget call requires all secretaries of line ministries to submit an agency result framework, for their ministries, to the Department of National Budget, along with the 2010 budget estimates. The key performance indicators should be identified at output as well as outcome levels, by the respective line ministries, to justify the annual budgetary provisions requested. The agency results frameworks for budgets were developed by the line ministries in close consultation with the Ministry of Plan Implementation and Department of National Budget. The MfDR initiatives in the budgeting process facilitated the linking of resources to results.

Managing for development results as an administrative reform

The National Administrative Reform Council is a special public agency created under presidential directive as per the constitutional provision vested under H.E the President. It formulates and directs the implementation of administrative reforms in the Government of Sri Lanka. The Secretary to the President chairs this high level Administrative Reforms Council. The MfDR initiative of the Ministry of Plan Implementation is recognized as an important administrative reform programme by the National Administrative Reform Council. This Council holds regular meetings to promote the MfDR initiative of the Ministry of Plan Implementation. The Council has established a network of management reform cells, in all government institutions, to make administrative reforms a dynamic and participatory exercise. Hence, the progress of MfDR is monitored through these cells and reported to Council.

Country-level community of practice to facilitate peer to peer dialogues

The core officials of the Ministries, to which MfDR was introduced, formed into a 'community of practice' and shared their experiences and expertise through a face-to-face and peer dialogue. This, in a way, helped to establish connection among the 'islands of best practice' and helped the learning from successes and challenges. Moreover a quarterly news letter on 'results focus' helped to compliment and supplement the face to face community of practice dialogue.

Capacity building/readiness assessment

Capacity building is fundamental to institutionalize MfDR in government. The readiness assessment tool developed by the Asian Development Bank (ADB) has been used to identify capacity gaps in line ministries and to formulate the capacity development plan. The UNDP and ADB support helped to formulate the capacity development plan as a complementary activity to institutionalize MfDR.

In the implementation of the capacity development plan, much emphasis was given to awareness creation, staff capacity building and on the job training on MfDR methodologies, tools, techniques and practices.

Strategy to action

The MfDR approach encourages Ministries and Departments to understand the 'results chain' and establish logical linkages between planned inputs, expected activities/outputs and envisaged outcomes based on the 'theory of change'. The Ministry of Plan Implementation of the Government of Sri Lanka took the following key steps to operationalize MfDR in Line Ministries.

Articulating and agreeing on objectives. Identifying clear and measurable objectives (results) is aided by a results chain/ logic model. This basically involves preparation of a vision statement, mission statement, thrust areas and goals. The Government's 10 year National Development Framework; sector plans; localized Millennium Development Goals; and, Mandate of the Ministries are considered as a fundamental basis to articulate and agree on line ministry objectives.

Selecting indicators. Output and outcome based key performance indicators were identified to measure progress towards each objective. The key performance indicators for each expectation in the framework is meant to convey the breadth and meaning of the expectation.

Setting targets. The main aspect of the process is the setting of explicit medium-term targets for each indicator, to provide medium term directions, and to judge performance.

Monitoring performance. This involves developing a performance monitoring system to collect data regularly on actual results.

Analyzing and reporting. The process also involves reviewing, analyzing and reporting actual results against the targets.

Integrating evaluation. Evaluation is an integral part of the process to provide in-depth analysis of the "Why" question and to provide complementary performance information not readily available from performance monitoring system.

Performance reporting and feedback. Performance information is used for internal management accountability, learning and decision making process and for performance reporting to stakeholders.

The first three 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 steps combined are essential to an effective results based management system.

The National Operations Room is the information platform of the Ministry of Plan Implementation to support evidence based decisions. The MfDR web-platform was established in the National Operations Room to facilitate and track results and performance of line ministries.

In the past, the progress monitoring of ministries was focused heavily on financial progress (budget utilization), and physical progress (activity monitoring), with very little emphasis on achievement of outcomes, results and policy objectives. However, implementation issues and inter-agency coordination issues were addressed through the high-level monitoring meetings held by the President. The Ministry of Plan Implementation, on a quarterly basis, submits to the Cabinet of Ministers the progress on capital budget of all line ministries and development projects over Rs 50 million. In the recent past, the emphasis was more on 'synthesis type reporting' rather than the traditional way of producing a large volume of reports that resulted in 'information overload'. The institutionalization initiative of MfDR is a major shift in focus in the monitoring and evaluation system in the government.

The MfDR initiative is to be complemented with performance agreements to ensure ministries and institutions manage priorities by establish strong links between agency plans and employee performance. While the Ministry of Plan Implementation is introducing MfDR to ministries and departments, the Department of Public Enterprises of the Ministry of Finance and Planning has introduced MfDR as part of its corporate plan initiative to state owned enterprises and statutory bodies. Advocacy and awareness-building efforts on MfDR are in progress. Spending money was no longer sufficient evidence for results. Ministries and agencies are asked to demonstrate how their programmes and strategies were effective in producing the intended results. As stated by Osborne and Gaebler (1992) in Reinventing Government "a perfectly executed process is a waste of time and money if it fails to achieve the outcomes desired". Performance Management redirects our efforts away from "busyness" towards "effectiveness". Being busy is not the same as producing results.

Electronic project monitoring system: a distinctive feature in Sri Lanka

One of the noteworthy and significant aspects is that a home-grown, user-friendly, national, web-based electronic on-line project monitoring system (ePMS) has been established in the Department of Foreign Aid and Budget Monitoring of the Ministry of the Ministry of Plan Implementation, to track the implementation progress (financial/physical), and the results of all development projects and programmes. The national ePMS is a pro-active, web based project monitoring system that helps to track whether development projects and programmes are implemented according to the agreed work plan in a results oriented manner. The system provides access to project information on a donor-wise, sector-wise and ministry-wise basis.

The system uses early warning (traffic lights) signals which enable 'problems projects' to be separated from others. The system helps to identify bottlenecks, delays, issues and constraints in the implementation of projects and any additional needs of the executing agency. It includes: results monitoring using logical framework analysis; monitoring compliance of loan covenants; tracking cash flow and reimbursable foreign aid claims; procurement progress; and, major issues with pictorial proof. Flash reports on problem projects help in troubleshooting exercise and are also submitted to the Cabinet of Ministers on a quarterly basis to fulfill the reporting

requirements. The ePMS also captures feedback from beneficiaries and citizens. Notable features of the system are:

- Keeps track of the development project progress electronically, covering financial and physical progress.

- Harmonizes results reporting through the use of results frameworks. National ePMS is a home-grown, country owned, country driven system that provides the public sector with paperless monitoring capability.

- Facilitates pictorial proof to demonstrate progress and results at ground level.

- Data capture at source, reducing transaction costs.

- Provides alerts and early warnings on shortfalls and gaps, including time and cost-overruns. Provides red alerts on problem projects and projects behind schedule.

- Escalates to the higher authorities the problems, issues and constraints faced in the implementation of development projects and facilitates trouble shooting of problem projects and projects behind schedule.

A recently completed Donor evaluation mission rated the ePMS as a success story in terms of its comprehensive coverage, periodical updating, and use of information for troubleshooting. However, the current low level utilization of the system by sector ministries indicates an unexploited opportunity. This is being addressed. Also, it is necessary to enhance 'data capture at source'. The Ministry of Finance and Planning is using the system as part of its efforts to strengthen project management capacities. The Ministry of Plan Implementation uses the system for trouble shooting and for better overall coordination to help to address execution problems.

The National ePMS was based on the Malaysian National Operations Room model used by Dr.Mahathir Mohamed the then Prime Minister of Malaysia to track implementation progress of all development programmes. The Sri Lankan ePMS is a home grown model with a user friendly facility to track development project progress. It is not intended to be used as a "policing function" but as a system that helps to identify projects behind schedule or 'sick' projects. The system depicts on-schedule, behind schedule and sick projects using a traffic light colour code by sector and ministry classification of projects, and facilitates troubleshooting. Currently there are over 120 large and mega projects in the National ePMS.

The National ePMs captures many key elements that are fundamental and essential to track the progress of development projects. The key elements are:

- *Project profile:* the profile provides the basic information of the project; Its name; source of funding; implementing arrangements; objectives; purpose and outputs; location; cost; timelines; and, officer responsible and other basic information.

- *Financial progress:* tracks the cumulative financial progress as well as the monthly financial progress against total funds available and the annual budgetary targets. It also monitors actual disbursements against targeted disbursements on a project wise basis.

- *Activity monitoring:* the project outputs are broken down into components, sub-components, activities and sub-activities using 'work breakdown structures'. All activities have planned timelines and targets and progress is monitored against such timelines.

- *Logical framework analysis:* the results framework of the projects shows the 'programme theory' with a results chain and a measurement system, including key performance indicators and risks. The system helps to track results and major risks.

- *Monitoring compliance of loan covenants:* this helps to track the compliance of loan covenants of all donor funded projects.

- *Procurement Monitoring*: project management mainly involves managing contracts. Hence, procurement management is an important aspect. Procurement monitoring helps to keep track of the procurement progress which is a core area in all development projects.

- *Monitoring major issues:* the system helps to keep track of the major issues affecting the implementation of projects and the action taken on such issues. This facilitates troubleshooting.

The major functions of ePMS are as follows:

- It serves as the executive information for central agencies such as President's Office, Ministry of Plan Implementation, Ministry of Finance and Planning and other line ministries.

- Provides information to project managers to manage projects effectively. Hence, the ePMS supports managerial functions.

- Access to information by citizens helps to create transparency and accountability. Hence ePMS helps to support the governance functions.

366

Figure 6: Electronic Project Monitoring System

Data Captured At Source
Project Office feeding Information to ePMS
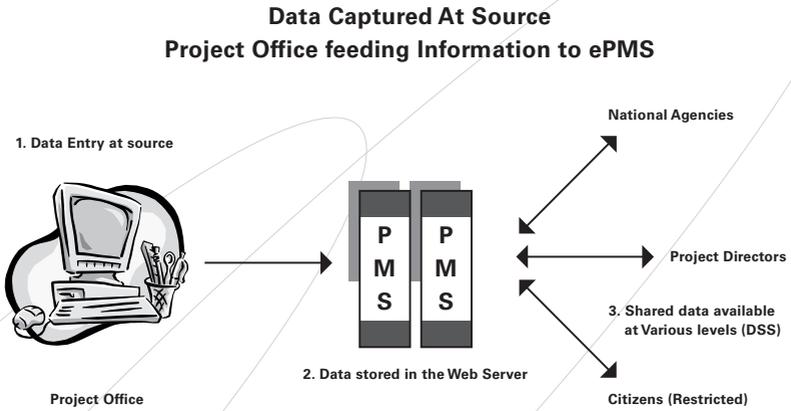
The above diagram depicts the flow of capturing data from its source to its end use. The project information is captured at source from the project office, stored in the web servers and finally, shared with potential users at various levels. Hence the system serves as a 'decision support system' and 'executive Information system'.

The ePMS has special report generating facilities. It includes:

• *Flash report:* higher level flash report to senior policy makers on overall financial progress, physical progress, status of projects, major issues and action taken in the implementation of projects.

• *Summary report:* the summary reports on aid utilization and budget progress can be accessed ministry-wide, sector-wide, donor-wide and Project-wide to various monitoring authorities.

• *Tailor made reports:* tailor made reports can be generated using the crystal report facility.

The national ePMS has many salient features. It helps to create an on-line information anywhere, anytime and enables Information be projected on to a large screen and used for discussions at meetings.

The National Web based Electronic Project Monitoring System can be accessed through http://www.fabm.gov.lk/

National evaluation arrangements.

On-going, ex-post and impact evaluation. The Department of Foreign Aid and Budget Monitoring of the Ministry of Plan Implementation undertakes on-going, ex-post and impact evaluation of selected mega projects and disseminates evaluation findings to concerned stakeholders. The evaluation function of the Ministry of Plan Implementation was strengthened through technical assistance support of the Asian Development Bank. The OECD/DAC evaluation criteria such as: the relevance of the strategy; efficiency of implementation; effectiveness; impact; and, sustainability are given due consideration in undertaking the evaluation. The OECD/DAC Evaluation Quality Standards are also considered to maintain the quality of evaluations. On invitation by donors, the Department of Foreign Aid and Budget Monitoring of the Ministry of Plan Implementation participates in joint evaluations with donors which helps to create national ownership and build local capacity in evaluation. The Department also undertakes diagnostic rapid assessment through field visits to problem projects and submits flash reports to the Secretary of the Ministry of Plan Implementation to facilitate troubleshooting of projects behind schedule. Given the human resource constraint in the Ministry, it was decided to outsource the evaluation of priority mega projects. The Ministry, in close consultation with relevant line ministry, decides on areas to be covered in the terms of reference for evaluation. Also, in designing the terms of reference, the Ministry identifies the information needs of the President, Cabinet and other stakeholders. The findings of such evaluations are disseminated to the line ministries and project offices for necessary follow-up action.

A key problem in many countries has been the inability to access the evaluation information available on various development projects and programme, either already completed or currently being implemented by the Government. Such evaluation lessons and findings are important and useful to improve the quality of new projects and programme, especially as it helps to avoid past mistakes and to build on best practices in the formulation and designs of new projects.

Evaluation information system. Having recognized the importance of a systematic use of evaluation and feedback arrangements, the Ministry of Plan Implementation has taken action to establish a web-based post-evaluation information system to ensure effective dissemination of evaluation findings, lessons learnt and the

synthesis of such findings. Such evaluation information will provide sector-wise synthesis to ensure more effective feedback and assist in integrating evaluation findings into the planning, budgeting and policy making process. Also, the public availability of the evaluation reports through the evaluation information system, is expected to improve public accountability and transparency.

Evaluation answers the questions of "what works, what does not work and why, and in what context does it work". The responses are important for planning and programming and they contribute to development effectiveness. An evaluation information system enables development practitioners to access evaluation information anywhere, anytime. It thus empowers them to make evidence based decisions in development. The establishment of the system is considered as a critical milestone in the MfDR initiative in Sri Lanka as it is expected not only to improve aid effectiveness but also to promote a learning culture. The support towards the development of evaluation information system was extended by the UNDP.

Evaluation of the implementation of the Paris Declaration on Aid Effectiveness. The Ministry of Plan Implementation and the Department of Foreign Aid and Budget Monitoring undertook an evaluation of the implementation of the *Paris Declaration on Aid Effectiveness*. The GOSL strongly believes that the five principles of *Paris Declaration on Aid Effectiveness* (national ownership, alignment, harmonization, managing for development results and mutual accountability), are fundamental to improve aid effectiveness and development effectiveness.

Hence, with UNDP's support, the Ministry of Plan Implementation undertook an independent evaluation to assess the implementation progress and results of *Paris Declaration* commitments.

An evaluation 'reference group' was formed to advise and a 'management group' was set up to coordinated the evaluation.

This evaluation identified the impediments for the implementation of *Paris Declaration* and helped the GOSL to take suitable action to ensure effective implementation of *Paris Declaration Principles*. The evaluation findings have also fed into the global evaluation process, enabling the OECD/DAC Development Evaluation Network to undertake a synthesis of the *Paris Declaration* evaluation in 8 countries and by 11 donors. These findings were useful in the formulation of the *Accra Agenda for Action* at the High level forum on Aid Effectiveness held in September 2008 in Ghana.

The Accra High Level Forum endorsed the following as the fundamental action required to accelerating aid effectiveness: the need for aid predictability; use of country systems by donors; untying of aid; and, mutually agreed conditionalities instead of imposed conditionalities. This implies that increasing aid effectiveness requires radical changes in the behaviour of donors and development partners. Implementation of the *Paris Declaration Principles* and of the *Accra Agenda for Action* are fundamental to the change management process and improved development effectiveness.

Sri Lanka Evaluation Association. The Sri Lanka Evaluation Association SLEVA, as a civil society organization, plays a catalyst role in advocacy; awareness creation; training and helping in developing standards; ethics; methodologies; and, best practices to improve evaluation culture. Its membership comprises academia; researchers; private sector consultants; government officials; and, NGO representatives interested in evaluation. SLEVA works closely with the Ministry of Plan Implementation in building monitoring and evaluation capacity and culture. SLEVA works with Ministry in areas such as evaluation training, sharing evaluation best practices, supporting the organization of evaluation forums on evaluation topics, international and national conferences on evaluation and, promoting communities of evaluation practitioners in the country.

Strategies to strengthen the national evaluation system

The following strategies have been identified to strengthen the national monitoring and evaluation system of the Ministry of Plan Implementation.

Policy commitment and support

- Advocacy and sensitizing, at political and policy level, on the importance of results-based monitoring and evaluation, to ensure acceptance and placement of monitoring and evaluation in key decision-making centers of the government in order to create local demand for monitoring and evaluation.

- Ensure monitoring and evaluation institutions are linked to the planning, budgeting, resource allocation and policy functions of the government and that MfDR concepts are integrated in all areas of the development cycle.

Legal and budgetary support

- Develop a legal foundation to make monitoring and evaluation and MfDR mandatory. Use law, decree, cabinet decision or other high level pronouncement to legitimize MfDR concepts and results-based monitoring and evaluation systems.

- Provide sufficient financial allocation for strengthening MfDR and monitoring and evaluation in the line ministries.

- Ensure that there is a right balance between "monitoring" and "evaluation." Preferably, separate evaluation from monitoring to ensure balanced resource allocation for evaluation.

Sound institutional arrangement

- Strengthen institutional arrangements to place monitoring and evaluation and MfDR in a strategic context.

- Establish links between evaluation exercise and performance audit exercise by encouraging partnerships between evaluation institutions and performance audit institutions (i.e. Auditor General's Department) with regard to accountability oriented evaluations.

Standards, ethics and guidelines (quality of evaluations)

- Develop evaluation standards, guidelines and ethics to ensure good quality evaluations. Ensure scoping sessions are conducted to clarify the evaluative questions and to ensure needs of the potential users are taken into consideration and that the timing of evaluation is appropriate.

- Encourage the national evaluation association to actively promote an evaluation culture and MfDR concepts.

- Develop standards and criteria for good evaluation in collaboration with civil society such as SLEVA and undertake meta evaluations to ensure quality evaluations.

Strengthen evaluation guidelines and systems

- Localized guidelines are being strengthened for systematic evaluations.

Strengthen methodologies and practices

- Make evaluation a specific process within the development policy and project cycle. Expand evaluation to cover projects, programmes, sectors, policies and institutions. Encourage

synthesis of project evaluations to provide sector-wide learning. Promote cost-effective rapid assessment methods under time, budget and resource constraints. Ensure consistent, localized evaluation methodology and terminology.

- Re-examine the approaches and tools for evaluating the multiple dimensions of development. Encourage the use of diverse or multiple methods, as well as participatory methods for lessons learning (utilization) oriented evaluations.

- Encourage more joint evaluations instead of donor-driven evaluations.

Evaluation capacity development

- Strengthen the professional evaluation capacity within the government through continuous staff training.

- Promote in-country evaluation faculty development programmes in Sri Lanka Institute of Development Administration (the government arm for training), and in other universities at graduate and post-graduate level. The Postgraduate Institute of Management of the University of Sri Jayewardenepura has introduced a Master of Public Administration Programme which includes MfDR, project monitoring and development evaluation in the curriculum.

- Strengthen the documentation centre on evaluations, promote exchange of experiences and access to best practices and sharing of databases.

Strengthen the feedback arrangements

- Improve disseminations of evaluation reports through in-house workshops/seminars, customized reports, evaluation summary reports, press briefings, and a post-evaluation information system.

- Establish strong feedback arrangements among evaluation; planning; decision-making; policy formulation; project appraisal; programme management; budgeting; and, resource allocation functions.

- Ensure action is taken on the recommendations in evaluation reports. Wider dissemination of evaluation information should preferably include the Parliamentary Public Accounts Committee, Parliament Library and media. User friendly evaluation synthesis or summary reports should be widely circulated.

- Stimulate the evaluation issues in the country's development dialogue and sector programme assistance. Monitoring and evaluation units must have active involvement in the planning of new programmes.

- Incorporate evaluation lessons into the new project concept documents or project submission formats so that past mistakes are not repeated. Revise project submission formats to incorporate evaluations lessons from past projects.

The ultimate success of evaluation depends on how well the planners and decisions makers utilize the valuable monitoring and evaluation findings and lessons to improve future programme, projects, policies and institutions.

Success factors

Sri Lanka's MfDR and monitoring and evaluation systems and practices have been internationally recognized as best practice approaches worthy of scaling up. Senior Government officials (from Yemen, India, Afghanistan, Uganda, Asia Pacific Community of Practice on Managing for Development Results and International Programme for Development Evaluation Training), visited Sri Lanka to study the emerging good practices of MfDR and monitoring and evaluation initiatives.

The *OECD Sourcebook* (2008, Third Edition) on emerging good practice in MfDR highlights the Sri Lankan case study on MfDR. The Asia Pacific Community of Practice on MfDR has identified the following factors for successful institutionalization of MfDR in Sri Lanka.

- Strong support and buy-in from the top and strong and sustained leadership for results oriented reforms has been instrumental in taking this change process forward. Government policy on MfDR endorsed by the Cabinet of Ministers and champions at national level, and change agents at different levels of government, help to drive the initiative.

- The MfDR approach adopted in Sri Lanka is a whole-of-government approach covering national, sectoral, institutional and project level. It is not just viewed as a technical tool but as a comprehensive way of thinking to achieve outcomes and impacts.

- A cascading approach which combines a mix of top-down and bottom-up approaches, and a sequenced approach beginning with pilots and then expanded and mainstreamed in government. This overall approach provides a more pragmatic strategy.

- The MfDR reforms complement the existing initiatives, strategies, and general reform agenda and are part of the country systems. Hence, the "perfect" should not become the enemy of the "good" in the MfDR area as most developed countries are still struggling to achieve satisfactory solution.

Issues and challenges

In many countries including Sri Lanka, the wider dissemination of monitoring and evaluation findings continues to remain a problem. Monitoring and evaluation institutions and the planning institutions seem to function in isolation and do not have an effective formalized feedback arrangement to integrate lessons into the planning and design of new projects. These institutional gaps defeat the very purpose of monitoring and evaluation. Therefore it is necessary to establish strong links between monitoring and evaluation on the one hand, and on the other hand, policy formulation, reforms, planning, budgeting and resource allocation functions. The GOSL has identified the need to establish strong feedback mechanism. Feedback is the weakest link in the project cycle. If one takes the project cycle as a results chain, "A chain being only as strong as its weakest link". This issue is now being seriously addressed.

In many cases, donors and partner countries still continue to be disbursement oriented. Moreover, donors tend to be using their own donor systems rather then country systems to maintain visibility and attribution. The lack of demand for MfDR; shortage of professionals; multiple results frameworks; too many indicators; lack of aid predictability; and, weak statistical capacity have been identified as constraints in many developing countries. Also, it is necessary to keep in mind that MfDR, as a change management process, takes time.

While recognizing the demand side of the equation, for creating local demand for evaluation with utilization focus, the supply side of the equation which includes skills, procedures, methodology, data systems, manuals etc., has also to be addressed. The need to focus on national evaluation capacity development is equally important. However, making monitoring and evaluation information available does not necessarily mean effective utilization. It is hard to justify the existence of an monitoring and evaluation system which is not

utilized effectively. However, the GOSL was able to address some of these issues with the technical assistance support of UNDP and ADB.

Attribution is a challenge for measuring performance in ministries and departments. In selecting the key performance indicators care should be given to the attribution issues. Unlike the private sector, the public sector does not have a single 'bottom line'. Many areas of government activities are characterized by multiple stakeholders and divergent interest. Target setting itself is a challenging process. It is difficult to get consensus on goals/values and difficult to agree on key performance indicators. Specifying and agreeing on expected results is not easy. The results chain is not always as logical as expected. Indicators are missing for some results areas. Targets and baselines are not given. Setting achievable targets is not possible in the absence of a baseline. The greatest problem associated with performance management is unrealistic expectations. Information overload is also a problem. Policy-makers were unable to absorb the vast amount of information provided to them. Outputs are manageable to institutions, but outcomes are the results of collaborative efforts of other institutions. Agencies can more easily be held accountable for output targets but not necessarily for outcome targets.

MfDR systems in some countries did not fully achieve their expectation as there is a "stick" but no "carrot". Also, the performance level slipped partly due to lack of resources and unrealistic expectations such as, for example, "no child left behind". To ensure successful operation of results based management systems there should be 'incentives' for achievements and some form of 'penalty' (punishment) for slippage. Also, the weak link between "agency performance" and "individual performance" is a concern. It is also necessary to understand who needs what information, for what purpose, and when. It should not be overly supply-driven information. Moreover, in some cases, MfDR creates fear of being held accountable for performance when cooperation and assistance from outside the organization are needed for success.

It is necessary to look at the balance between learning and accountability. While independent evaluation is important for ensuring objectivity, too much emphasis on the accountability focused, donor-driven, independent evaluation function can be a potential constraint for lessons learning and feedback. Hence, the importance of lessons learning and ownership has been recognized in the GOSL evaluation system.

For too long, many countries have not adequately responded to the criticisms that ex-post evaluations are done late and viewed as 'post-mortem' exercises which do not contribute much to strategic decision making. It is necessary to recognize the importance of lessons learning and performance accountability. Increasingly, concurrent evaluations are encouraged for mid course corrections. Also, the GOSL encourages donors to undertake more joint evaluations to ensure national ownership, lessons learning and capacity building.

Despite the enormous methodological and technical challenges, it is recognized that institutionalization of evaluation is the way forward to ensure results orientation in development work. Moreover, development policy and aid tend to shift, from projects and programmes, to sector-wide approach and as such monitoring and evaluation approaches need to cover policies, sectors and thematic areas on a country wide basis. Policy evaluations, sector evaluations and thematic evaluations are becoming equally important and the GOSL has given much emphasis to such evaluations.

There has been a general tendency to monitor rather than to evaluate. It is necessary to give equal importance to evaluations by finding a right balance between the two activities. GOSL is mindful of these aspects and monitoring and evaluation is viewed from a wider country based context.

Conclusion

Institutionalization of MfDR has been instrumental in creating a shift in focus in monitoring and evaluation systems and practices by expanding the focus of "results" in development.

The Government's ambitious institutionalization of MfDR in Sri Lanka has laid the foundation for emergence of the thriving results-focused monitoring and evaluation community in the Government of Sri Lanka. Although there is a long way to go before MfDR will be completely assimilated into the public sector culture, the initiatives have achieved significant progress towards a government wide shift to results based monitoring and evaluation practices.

References

Government of Sri Lanka. *Sri Lanka National Development Strategy*

Government of Sri Lanka, *Sri Lanka National Web-based Electronic Project Monitoring System* http://www.fabm.gov.lk

OECD. (1991). The *DAC Principles for the Evaluation of Development Assistance*, OECD.

OECD. (2004). *The Second Roundtable on Managing for Development Results,* Marrakech 2004.

OECD. *(2005). The Paris Declaration on Aid Effectiveness* High Level Forum on Aid Effectiveness. Paris 2005.

OECD. (2006). *Evaluation Quality Standards.* OECD/DAC, *at* www.oecd.org/dataoecd/30/62/36596604.pdf

OECD. (2008). *Accra Agenda for Action.* High Level Forum on Aid Effectiveness. Accra. 2008.

OECD and World Bank. (2008). *Managing for Development Results Sourcebook*. OECD/DAC and World Bank, Third edition, 2008.

Osborne D., and Gaebler T. (1992) *Reinventing Government.* Addison-Wesley Publ. Co.

THE COUNTRY-LED SYSTEM TO MONITOR AND EVALUATE PUBLIC POLICIES IN MALI. GOOD PRACTICES AND LESSONS LEARNED

by Sekouba Diarra,
National Coordinator of the PSRP
Ministry of Economy and Finances, Mali

by Robert Ndamobissi,
Chief of Planning, Monitoring and Evaluation
UNICEF Mali

Introduction

The implementation of the 2005 *Paris Declaration* that established the 5 principles aid effectiveness: national ownership/leadership; alignment; harmonization; the results-based management; and the co-responsibility of Government and donors is very developed in Mali, as is the follow-up of the 2008 *Accra Commitment for Aid Effectiveness*. The country-led monitoring and evaluation is so well built around the main public policy, the second economic growth and poverty reduction strategy for the period 2007-2011 adopted by government and all donors. This strategy is the broad develop-mental framework for the country and is complimented by various sectoral strategies and investment plans like the health compact document (1999-2009 with extension to 2011) and the education plan of 1999-2011.

The monitoring and evaluation system and its governance

The monitoring and evaluation system in Mali is institutionalized by a legal decision of Government (Decree of the President of the Republic) around the *Poverty Reduction Strategy Paper* (PRSP) with the intent of ensuring regular measurement of the effectiveness of expected results of PSRP, and to show the challenges, strengths and weakness for decision-making and national budgeting process. A consensual *evaluation framework of the PSRP* has been adopted by Government, donors and civil society. It contains 60 selected multi-

sectoral strategic indicators of impacts and outcomes, disaggregated by gender. The time frame for the main evaluations and for the review to be completed is formally established by the Government. The Ministry of Economy and Finances is leading on the implementation of the evaluation framework, working closely with the other Ministries responsible for data collection and publication of their sectoral indicators. The top priority demand for country-led evaluation is based on the completion of the annual review of the PSRP and the annual review of the sectoral development plans which are mandatory and conditional for the allocation of global and sectoral budget support by the World Bank, other bilateral donors and the European Union.

The monitoring of the implementation of the 12 indicators of the *Paris Declaration* is also well advanced in Mali under the leadership of the Ministry of Economy and Finances. The review of financial contributions from Donors and assessment of the extent to which they respect the 5 principles is undertaken through a participatory approach. The national report is validated by all stakeholders. A national Committee of Aid Harmonization is established by the Government and led by the Deputy Ministry of Economy and Finances who leads on this issue.

The third main framework of the monitoring and evaluation system in Mali is the existence of the *National Strategy of Statistics* called (Director Scheme of Statistics) for the period of 2006-2012 adopted by law through the National Assembly. It contains the main objectives and expected strategic results, operational outcomes and outputs which will strengthen the whole multi-sectoral statistical information system in Mali. This national Strategy is led by a national committee called the Superior Council of Statistics (a political and legal institution) leaded by the Ministry of Economy and Finances which is the strategic decision-making body. The Superior Council is supported by a Technical Committee of Statistics.

The fourth institutional arrangement for governance of monitoring and evaluation in Mali is the establishment of a General Office of National Control to ensure relevance, coherence and good governance of programmes and projects. Called the Bureau du Verificateur General, it is based on good practice for inspection in Canada. The Bureau is an independent and very powerful institution linked directly to the Head of the State. This institution is a good example of upstream transparency and accountability for good economic governance. The annual report produced by the national Comptroller is shared with the media, donors and Parliament. A press

379

conference is held around the critical issues. Donors formally request a government response to the questions and to critical programme and project management issues of revealed by the report.

The Government of Mali has accepted the completion in 2009 of the Peer Evaluation of the Socio-economic and Democratic Governance practices in Mali that has been done by the NEPAD Experts. The findings, conclusions and recommendations have been presented to the President and all government bodies and were disseminated to all public. This is a good practice of self awareness of strengths and weakness of Governance in Mali.

The decentralization of the governance of the monitoring and evaluation system is well institutionalized in Mali through the creation of 11 Sectoral Directorates of Planning and Statistics (CPS) and the establishment of the Regional Directorates of Planning and Statistics in all 8 regions sub-divided into the sub-region.

The creation of the 11 CPS is linked to the 12 strategic interventions for the 4 strategic priorities of the PSRP. Each CPS is led by a national director. The CPS is responsible for the design, development and monitoring of the sectoral plans and the design and supervision of the most important evaluations within the monitoring and evaluation framework for sectoral plans, programmes and projects, in partnership with donors. An annual action plan containing the main activities for data collection, studies and evaluations is planned in line with budget and sources of funding. The allocation of the national budget takes into account the planned activities of the 11 CPS. The annual national action plan for statistics is the integration of the CPS's 11 action plans.

The monitoring and evaluation system contributes significantly to the improvement of policies and programmes: evidenced-based and the results-oriented approaches to planning, budgeting and reporting become an owned practice of transparency and governance. The national budgeting process of the year t+1 (where 't' is the current year) started with the completion of the annual sectoral review of the year t-1 and the orientation of the social budget allocation for the year t+1 is based on the key findings and recommendations from the annual review of the PSRP implementation of year t-1 that is completed at June. The discussions and adoption of the national budget allocation by the Parliament is based on the report of the PSRP annual review. The accountability of managers of programmes and projects is being improved due to the inspection action of the Comptroller General.

Good practices within the monitoring and evaluation system in Mali

Participatory planning processes involving civil society, children, donors, and decentralized regions provide a common consensus on the vision and strategic results and outcomes. The participatory process of the annual review of PSRP and sector-wide action plans involving all Government counterparts, civil society, decentralized services/authorities and donors.

Political commitment of Government comes from the endorsement of the conclusions and recommendations of the annual review of the PSRP adopted during the Council of Ministers and published in the Official Journal and through the media. This process leads to a high level of accountability. There is also an established civil society organization for monitoring and evaluation of the PRSP in Mali.

Strong partnership and coordination

Aid coordination has been strengthened in Mali with the good practice of regular upstream monitoring and evaluation of key strategic priorities based on evidence. There is a quarterly meeting of Government and all donors to discuss the strategic issues and perspectives around the national development framework. The minutes are adopted jointly and shared with all stakeholders. There are also thematic groups of donors jointly led by deputy ministers and donors and good coordination of the national strategy for development of statistics.

A pluri-annual operational action plan for statistics has been established for three years and contains all the sectors' planned studies, survey and data publications. An annual action plan has also been established. Since, 2009 an annual review of the national action plan for statistics has been completed (the technical and political review) led by the CSCRP and involving all Ministries, NGOs and donors and the decentralized statistics services.

Mali has a good example of existing sectoral thematic donor groups which ensure the coordination of technical and financial assistance for implementation of the national strategy for development of statistics, and for strengthening national capacity on data collection and the publication and completion of annual action plans for statistics. This offers a good opportunity for the Multiple Indicators Cluster Survey (MICS) and *Devinfo* database to be adopted as the national tools for monitoring the PSRP and sectoral plans in Mali.

The existence of the UN inter-agencies group on monitoring and evaluation, led by UNICEF, with an annual work plan represents good practice for inter-agency partnership to strengthen the monitoring and evaluation system in Mali. A good practice example is the realization in 2009, of the first joint annual review of *United Nations Development Agreement Framework* (UNDAF). This replaces the annual review of each Country Programme Action Plan and is a good example of harmonization. UNICEF did not itself undertake the annual review, but provided strong technical leadership on the design and the writing of the result-oriented report, by applying the standard of quality assurance for evaluation (the formative evaluation).

The ownership and national leadership of the *Devinfo* database, linked to the PSRP evaluation framework, is an example of good practice for the sustainability of, and building partnership for, *Devinfo* in Mali. The Ministry of Economy and Finances is leading on advocating to all bilateral and multi-lateral donors, the importance of a *Devinfo* database in Mali: when one donor comes with a proposal to support the development of the database, the PSRP coordination unit and the NIS it and instead invite all the donors together to contribute to the same *Devinfo* database.

A strong partnership was established around *Devinfo* in Mali, through the signing in 2008 of a memorandum of understanding involving the Government, the World Bank, UNDP and UNICEF. The aim of this partnership is the scaling-up of utilization of the national socio-economic database using *Devinfo* technology for the monitoring of PSRP and Millennium Development Goals (MDGs) in Mali.

To facilitate knowledge management, a national website for statistics is functioning well and is regularly updated. A website for the national *Devinfo* database has been established (Malikunnafoni.com) and linked to the INSTAT web site. There is a regular publication of a sectoral statistics year book for education and health. A yearly national bulletin of statistics on children and women is published by the Ministry of Woman, Child and Family and was officially launched by the Government for upstream advocacy on children issues.

UNICEF and UN key specific contribution to the monitoring and evaluation system in Mali

UNICEF Mali and the UN agencies provided key strategic contribution to the monitoring and evaluation system in Mali through the strengthening of national capacity in management and utilization of *Devinfo* for the monitoring and evaluation of the public policies, MDG and UNDAF in Mali. A total of 17 ministries, 8 regions (a total of 120 focal points and users) and 15 UN agencies have been enabled to use the *Devinfo* database containing 525 indicators. Sub-sectoral databases have been established for those ministries and the 8 regions and contribute to the national socio-economic database.

The supply-side of evaluation and results-based management was strengthened by supporting the development of the formal national civil society association for evaluation named the Association of Promotion of Evaluation in Mali (APEM). In addition, technical support was provided for the study the capacities needs for evaluation within Mali and for the preparation of a national symposium on evaluation to be held in August 2010. The aim is to reinforce the culture of systematic evaluation and the development of a national plan for evaluation capacity building, to be used to institutionalize national evaluation policy and for the establishment of a strong partnership around the standardization of evaluation in Mali.

Challenges and opportunities

The main challenges for the monitoring and evaluation system in Mali are:

i) The updating of impact and outcome indicators for monitoring the progress of MDG and PSRP. Due to the high cost of household survey, the funding is not guaranteed by Government.

ii) The alignment of the annual review or specific country programme from each donor (bilateral and UN agencies) with the existing national system of monitoring and evaluation (national and sectoral annual review). For example, how could the EU or UNFPA or UNICEF abandon their individual annual review of programme and adopt Mali's existing sectoral review.

iii) The lack of human capital (less than 5 high level statisticians and demographers are available at INSTAT).

The main opportunities are: the establishment of a common strategy for assistance to development in Mali signed between Government and all bilateral and multi-lateral donors around the PSRP; the UN partnership on monitoring and evaluation leaded by UNICEF; the existence of the civil society monitoring and evaluation association; the perspective of the implementation of a ONE UN (delivering as one body) in Mali that will start in 2013 as adopted jointly by Government and all UN agencies; and, the establishment of the health international partnership (IH+) in Mali which result in the signing of a performance contract around the health sector-wide action plan in Mali including one common framework for health monitoring and evaluation. The existence of the donors thematic group on statistics is a good opportunity to reinforce the synergy and complementarities for the implementation of the national strategy for statistics and the PSRP evaluation framework in Mali. The implementation of the *Paris21* programme, the progress of "effectiveness of decentralization" with existing regional development plan and local development plans is a good opportunity for strengthening the decentralized practice of the monitoring and evaluation system in Mali.

Conclusion and lessons learned

The existence of a national development framework (PSRP; sector-wide action plans and the decentralized development plan) with well designed key priorities and expected results, is the fundamental step for the monitoring and evaluation system. Political commitment in favor of the monitoring and evaluation of public policy is the biggest lessons learned for the institutionalization of the monitoring and evaluation system and its sustainability. National leadership is the key condition for the development of results based management and good governance for results. The existence of a strong partnership and coordination mechanism is the main condition for implementation. The key conditions for a strong institutionalized monitoring and evaluation system is the existence of a nation strategy for statistics; the pluri-annual and annual operational action plans, which are regularly assessed; and, the decentralization of the statistics system.

POLICIES, INSTITUTIONS AND PERSONALITIES. LESSONS FROM UGANDA'S EXPERIENCE IN MONITORING AND EVALUATION

by David Rider Smith,
Advisor, Monitoring, Evaluation and Statistics
Office of the Prime Minister and Uganda Bureau of Statistics

by Joe Kanengere Nuwamanya,
Director, Policy Coordination, Monitoring & Evaluation,
Office of the Prime Minister,

by Rosetti Nabbumba Nayenga,
Deputy Head, Budget Monitoring and Accountability Unit
Ministry of Finance, Planning and Economic Development

Introduction

Over the past two decades, considerable efforts have been made to establish a strong and robust basis for assessing public spending, and its effects on the development of the country and its citizens[1]. To better understand the linkages, patterns and constraints to growth and change, substantial resources have been spent on: establishing good time-series and qualitative data on key socio-economic indicators; public accounts; regular monitoring of public policy interventions; and, on policy relevant research, analysis and evaluation.

In order to understand where this investment in monitoring, research and evaluation has had the biggest impact on public policy and accountability, it is necessary to examine the relationships between policy, institutions and individuals in the public sphere. Evidence suggests that the linkage between measurement and policy change has been productive only when the environment in each of these spheres is conducive and interfaced.

---

1   These assessment instruments are described in annex 1

385

Poverty monitoring and its impact on public policy

The Government of Uganda started to produce poverty monitoring data in 1992, through the *Uganda National Household Survey* reports, and has since updated this information every two to three years. This data, however, did not play a vital role in assessing public policy until the launch of the *Poverty Eradication Action Plan* in 1997. This plan was the first poverty-reduction focused national plan in sub-Saharan Africa. It was a nationally designed and driven plan although the preparation of such a plan[2] was a precondition for debt relief under the *Highly Indebted Poor Countries* initiative. The *Poverty Eradication Action Plan* provided the platform for major public policy initiatives including: Universal Primary Education; Primary Health Care; the Plan for Modernization of Agriculture; land reforms; HIV/AIDS programmes; and, the ten year Road Sector Plan. Heavy investments channeled into these and many other initiatives made it necessary for the Government to assess closely the usefulness of these interventions in terms of transforming society welfare. Hence, the policy environment and actors were open to the production of reports that would, in effect, illustrate how far the national policy was being effective.

In 1999, the Government of Uganda designed a poverty monitoring strategy that guided the production of biannual poverty status reports, and associated monitoring reports and publications. The poverty status reports were high quality analytical pieces which drew upon quantitative and qualitative household and facilities survey data to determine the patterns and changes in rural and urban poverty. The analysis and findings in these reports were used to influence the focus of medium-term planning and budgeting in the country.

The institutions responsible for generating, analyzing and reporting poverty data were critical to this process, not only for their effectiveness as stand-alone institutions, but also for the inter-relationships between these agencies. The poverty status reports were reliant on good, regular statistical data production, and the work of the Uganda Bureau of Statistics was critical in this. The Bureau, with substantial financing from the World Bank, initiated a process of producing regular nationwide household surveys on household income and poverty (*Uganda National Household Survey*), on health

2   Which later became Uganda's Poverty Reduction Strategy Paper (PRSP)

status (*Uganda Demographic Health Survey*), on population trends (*Uganda Population and Housing Census*), and then later, on other economic and public policy issues (*Uganda Business Inquiries, National Service Delivery Survey, Census Atlas*).

Within the Ministry of Finance, Planning and Economic Development, a Poverty Monitoring and Analysis Unit was established, to monitor, analyse and report on data generated on poverty and related issues, including the preparation of poverty status reports (with financing from UK). This work was supported by the Government-sponsored Economic Policy Research Centre. The "Uganda Participatory Poverty Assessment Process" was also established in the Ministry of Finance, Planning and Economic Development, to provide qualitative data on key social economic indicators and the impacts of selected Government policies on the citizenry. The qualitative data was an integral part of the information used to prepare PSRs.

Each of these establishments: the statistics bureau; the poverty unit; the participatory process and the research centre, had highly qualified, committed and motivated individuals in key positions. The ability to produce high quality monitoring reports in a timely manner for political and administrative consumption reflected not only their individual abilities, but their willingness to work together to deliver demand driven monitoring reports. Within the Ministry of Finance, Planning and Economic Development, a strong working relationship between the Poverty Monitoring and Analysis Unit and the top management of the Ministry (composed of Ministers, Permanent Secretary and other senior ranking officials) meant that findings from these products made their way into policy and allocative decisions, and in turn, attracted increased official development assistance. Many of the issues raised in the poverty status reports influenced decision-making at both the Cabinet and the Parliamentary level, and helped in focusing expenditures in areas that were most meaningful for poverty reduction.

The window of opportunity and practice in the production of poverty status reports in Uganda reflected the priority and progress made in poverty reduction from the President down, and the relationships and abilities of the institutions and individuals involved. This took place throughout the first and the second *Poverty Eradication Action Plan* (1997-2000; 2000-2003) but, declined with the third Plan (2004-07). This decline happened as the Plan began to be less influential in public policy-making when, in the face of evidence

of malpractice and corruption in the use of the state's resources, debate on development within the country and amongst international stakeholders shifted towards economic growth and enhancing the accountability of the state.

The shift to budget and performance monitoring for accountability

Whilst the evidence from household surveys and poverty status reports began to reveal that the overall poverty headcount was reducing in the country (from 56% in 1992 to 31% in 2006), it was also being noted that growth and development (and indeed poverty itself[3]) was becoming increasingly imbalanced.

As the public purse expanded, based both on a strong and stable economic growth rate, relatively low inflation, and a considerable rise in official development assistance, so too did concerns on the application and accountability for public spending. Efforts to strengthen public financial management included the recognized need to understand how public resources were being applied centrally, and locally under the decentralized system of government and public programme implementation, introduced in 1997.

The emphasis on monitoring shifted away from periodic analysis of poverty trends and causes, and towards the monitoring of budget spending. During the latter half of the last decade, the Ministry of Finance, Planning and Economic Development introduced a series of reforms to enable ministries, departments and agencies and local governments to plan and budget annually according to clear budget lines, and against the provision of products and services. Systems have been introduced requiring all to report quarterly on spending and progress towards stated output (product and service) targets, as the basis for future financial releases.

This massive reorganization and growth in the administration of Government has placed increasing attention on the generation and use of administrative data and statistics, and on the monitoring and oversight mechanisms in place to capture and report on performance information.

---

3 Notwithstanding the decline in poverty and positive growth trends, inequality in Uganda remained high. Income inequality, as measured by the Gini coefficient, increased from 0.35 in 1999/2000 to 0.43 in 2002/03 before declining to 0.41 in 2005/06.

The political interest and pressure to monitor spending and results has increased since the re-introduction of multi-party politics in Uganda in 2006, and the growing attention of the domestic media and international community to unequal growth, and the incidence of corruption in the use of public resources. The President and other senior policy-makers have taken cognizance of these issues, and have placed increasing demands on the public service to improve its stewardship of resources and ensure effective development.

The impact of this increased demand on public institutions is still unfolding. Efforts to improve oversight in key service delivery institutions (through regular implementation and budget monitoring), and through inspection of service delivery, have increased, though not in a uniform or consistent manner. The former Poverty Monitoring and Analysis Unit in the Ministry of Finance, Planning and Economic Development has been transformed into a "Budget Monitoring and Accountability Unit" to track expenditure and outputs against budgets and planned activities in a sample of frontline institutions, through direct field monitoring and reporting. Monitoring is focused on the outputs agreed and signed up in the performance contracts between the Permanent Secretary of Ministry and the implementing agencies. Efforts to reorganize the overall inspection function of the Government are underway. A Presidential directive to initiate public fora to hold local government to account (so called public 'barazas') has been initiated by the Office of the Prime Minister, and the Uganda Bureau of Statistics is seeking to expand its mandate to improve the quality of administrative statistics.

Simultaneously, the Office of the Prime Minister has reinvigorated a national coordination mechanism providing a conduit for evidence, debate and decision-making up from the sector working groups through a series of committees to Cabinet. The formulation of the *National Integrated Monitoring and Evaluation Strategy* provided the framework to enhance coordination in the generation and use of data and information, and a platform for the establishment of working groups to address particular constraints in overall public sector monitoring and evaluation, in order to improve capacity and practice. One of the outputs of this strategy has been the production of bi-annual government performance reports on the output and outcome performance of all sectors. These reports provide the Cabinet with comprehensive Government results information for the first time, and are being used to scrutinize institutions.

In this new environment, accountability has become the central concern, with the consequent de-emphasis on generating information for the purposes of understanding causes and generating policy lessons. The considerable differences in practices across Government in the monitoring and inspection of public policies reflect the balance of priorities and capacities within certain sectors, and amongst the international community who continue to invest in certain sectors over others (notably the social sectors and justice, law and order institutions).

The effectiveness of the increased monitoring of public spending has yet to be born out. Regular, information on the performance of some sectors of Government has had some impact on public policy and on the enforcement of accountabilities, but it has also revealed the widespread misuse of funds. The political class and legislative arm have still to make best use of this information in holding culprits to account, and to reallocate resources to more efficient areas.

Using development assistance as a lever to strengthen results management

Monitoring is beginning to have an impact on performance within a sub-set of sectors which have been targeted for measurement by the international donors who are providing general budget support. Initiated with the Poverty Reduction Support Credit of the World Bank, the identification of performance measures, and the achievement of targets which trigger financial releases, have been expanded as the number of agencies providing budget support has grown.

Efforts to harmonize budget support operations have results in a joint operation, with ten international partners providing approximately 350 million dollars per annum[4]. The pressure on accounting for these resources has led to the formulation of a *Joint Assessment Framework* (JAF) to measure progress and actions in key front line service delivery sectors, in public and financial management, and in key underlying policy areas.

This *Joint Assessment Framework* is derived largely from Government's own performance measures in certain sectors. The focus on this set of measures, by influential budget support donors

---

4   Budget support in Financial Year 2008/09 totalled USD 382 million, equating to approximately 15% of the national budget.

(including World Bank, African Development Bank, European Commission and European bilaterals), is starting to have a considerable knock-on effect in terms of the frequency and quality of implementation monitoring, and on mechanisms for oversight. As donor disbursements are tied to the achievement of agreed upon targets in the framework, pressure to deliver, monitor and report on delivery has increased. In turn, the focus has also turned to the quality of indicators and data collection as the evidence base of performance.

Whilst there have been fears that the emphasis on the *Joint Assessment Framework* as a sub-set of Government's performance measures will become perverse (by focusing public resources and attention on the issues in the framework at the behest of other critical areas of Government performance), the impact of this monitoring on a few sectors is beginning to be recognized. The pressure on regular reporting, participating in regular review meetings, and following-up on commitments has come through the personal commitment of key individuals within Government. While pressure continues to be applied by international donors (premised on the need to account for spending to their own domestic constituencies), this is fully aligned with the positions of the key personalities in the Government, and thus messages and the direction of change are being reinforced.

Whilst budget monitoring and expenditure tracking has continued to strengthen across Government, so too these efforts to measure performance more generally in key sectors may spill over into wider concerted efforts across Government, tied to the introduction of performance contracts for senior civil servants and other related measures.

Monitoring myopia? Efforts to strengthen the analytical and the evaluative

The concerted efforts to strengthen monitoring have come at a cost. The practice of public sector evaluation has never been institutionalized in the country, but was reasonably well serviced in the late 1990s and early 2000s through the poverty status reports and other analytical tools and products. This has led to a deficit in the analysis of results and constraints, and in the identification of policy lessons and choices. Monitoring itself does not answer these questions or address these issues.

Over the past two years, the Office of the Prime Minister has sought to address this concern about the deficit of analysis, through the conduct of key evaluations. These include an evaluation of the successes and failures of the *Poverty Eradication Action Plan* over the decade of implementation, and another evaluation assessing the implementation of the *Paris Declaration on Aid Effectiveness* in Uganda. While somewhat supply driven, the evaluation of the PEAP has found an audience amongst policy-makers, with the findings and recommendations discussed twice by Cabinet, and has in turn influenced the shape of the *Poverty Eradication Action Plan's* successor, the *Five-year National Development Plan*.

Subsequent efforts to strengthen evaluation practice include the development of a national policy on monitoring and evaluation which defines roles, requirements and practices to embed into the public service. Based on the policy, the Office of the Prime Minister is leading on: establishment of a rolling national evaluation agenda; an evaluation fund to strategically finance major public policy evaluations; and, a series of measures to improve standards, capacity, dissemination and follow-up.

The success of balancing the monitoring drive to incorporate evaluations is dependant on the effectiveness of the Government's policy coordination efforts. The re-established national coordination mechanism to share policy and implementation issues (including the findings of major reviews and evaluations), through a series of technical inter-sector committees up to Cabinet, is bearing early fruit. The progress being made in this area is in turn largely due to the efforts of key stakeholders in the relevant Ministries.

Conclusion

The experience of Uganda over the past two decades has illustrated that the establishment and effectiveness of monitoring and evaluation practices are most acute when policies, institutions and individual actors come together. The composition and balance of these factors shift over time. Analysis and evaluative work were pre-eminent in the late 1990s/ early 2000s, with a more recent shift towards monitoring and expenditure tracking. Both domestic and international institutions and actors have, in different ways and at different times, played major roles in the practice of monitoring and evaluation. The clear need is to ensure that both monitoring and evaluation become embedded in public service practice in order to ensure accountability for resources, and to generate knowledge for policy refinement.

References

Government of Uganda. (1997). *Poverty Eradication Action Plan*
(and revisions of 2000/01, 2003/04)

Government of Uganda and Joint Budget Support Development Partners. (2009).
Joint Assessment Framework. October, 2009.

Ministry of Finance, Planning and Economic Development. *Poverty Status Reports*
(various)

Office of the Prime Minister (2006). *National Integrated Monitoring and
Evaluation Strategy*

Uganda Bureau of Statistics. *Uganda National Household Survey* (various)

Annex 1. Typology of monitoring, review and evaluations arrangements in Uganda

User/ Supplier	Monitoring	Review	Evaluation	Remarks
Government of Uganda				
Ministry of Finance, Planning and Economic Development (MOFPED)	Quarterly financial reports Quarterly budget monitoring reports (initiated in 2009) Bi-annual budget execution reports	Half year and annual budget execution report	Poverty Status Reports (discontinued in 2008)	Incremental budget reform towards output budgeting and expenditure reporting. Poverty Monitoring and Analysis Unit transformed into Budget Monitoring and Accountability Unit, with commensurate shift from two-yearly evaluative PSRs to quarterly budget monitoring reports based on direct assessment of spending at site of facilities services

User/ Supplier	Monitoring	Review	Evaluation	Remarks
Government of Uganda				
Office of Prime Minister (OPM)	Annual sub-county barazas (initiated)	Half year and annual Government performance reviews	Periodic evaluations of major public policies and programmes	Barazas are public fora where community monitors present reports on Government performance to State officials, for public debate and follow-up. OPM chair and secretariat of national monitoring and evaluation Technical Working Group that brings together monitoring and evaluation officers from across Govt, NGOs, academia and donors to strengthen monitoring and evaluation in MDAs. OPM reports to Cabinet.
Uganda Bureau of Statistics	Economic and Social Statistics Community Information System (CIS)	Census and Household Survey Reports		Production of regular (monthly/ quarterly) statistics in consumer and producer prices, trade, finance, education, health, labour, etc). Conduct of annual socio-economic panel survey; four-yearly demographic health, household budget and service delivery surveys). Establishment of household monitoring on key socio-economic data (CIS)

User/Supplier	Monitoring	Review	Evaluation	Remarks
Government of Uganda				
Sector Working Groups	Quarterly monitoring of sector performance. Reports submitted to OPM	Annual sector performance reports and sectors reviews (in 30-40% of sectors). Summary of sector performance in Budget Framework Papers alongside budgets for coming year	Occasional evaluations of sector performance issues	Service delivery sectors have the most comprehensive development and investment plans, and related monitoring and evaluation frameworks and data collection systems. Over half of all sectors do not have long-term plans, or monitoring and evaluation systems. Data from the latter sectors is aggregated from respective MDAs. OPM, UBOS and donors seeking to strengthen and broaden sector indicators and monitoring systems
Ministries, Departments and Agencies (MDAs)	Quarterly monitoring of budgets and output performance. Reports submitted to MOFPED	Annual performance reports (some MDAs). Performance assessment included in Ministwerial Policy Statements alongside plans for coming year.	Occasional evaluations of MDA performance issues	Service delivery MDAs has the most comprehensive monitoring and evaluation systems (e.g. education, health, water) financed largely from donors.

User/ Supplier	Monitoring	Review	Evaluation	Remarks
Government of Uganda				
Local Governments (LGs)	Quarterly monitoring of budgets and output performance. Reports submitted to MOFPED and line MDAs	Annual review of LG performance (compliance) conducted by Ministry of Local Government		Local Government mandate and capacity in monitoring limited. Majority of data gathered at LG level is submitted to line MDAs for analysis (e.g. health facility, schools data). Resource flows from the centre are predominantly conditional, and thus relatively little scope for performance monitoring to have an impact on local level decision-making
Parliament				
Office of the Auditor General			Value for Money (VFM) audits	Increase planned in response to parliamentary and donor demand
Public Accounts Committee (PAC)		Annual accounts of the GoU		PAC plans to review VFM audits starting in 2009
Parliamentary Committees		Annual budget submission documents		

User/ Supplier	Monitoring	Review	Evaluation	Remarks
Donors				
Donors	Provision of financing to support Government monitoring systems in certain sectors/ MDAs, in particular those covered by the budget support Joint Assessment Framework (JAF)	Annual (joint) sector reviews Annual joint review of Budget Support through the indicators in the JAF. Occasional public expenditure reviews	Financing of occasional policy evaluations Periodic evaluations of donor financed programmes and projects.	Considerable investment in strengthening Government monitoring and conduct of surveys that address service delivery, but not coordinated. Investment in reviews and evaluations of donor-financed programmes and projects

JOINT EVALUATION OF THE ROLE AND CONTRIBUTION OF THE UN SYSTEM IN SOUTH AFRICA: LESSONS LEARNED

by Oscar A. Garcia[1],
Senior Evaluation Adviser, UNDP Evaluation Office

by Angela Bester[2],
Director, Deloitte & Touche, South Africa

Introduction

On 12 June 2009 the final report of the *Joint evaluation of the role and contribution of the United Nations System in the Republic of South Africa* was presented to the Government of South Africa and to the Acting Resident Coordinator of the United Nations in South Africa. The evaluation was unique:

- It was initiated by the Government of South Africa seeking an evidence-base for a policy dialogue in order to enhance its partnership with the UN system.

- It evaluated the UN system in South Africa as a whole, rather than on an agency by agency basis.

- Its scope extended beyond UN development assistance to South Africa, to include the UN's role in supporting South Africa's efforts on the African Continent and in the global arena.

- It was conducted by a team independent of the UN and of the Government of South Africa.

- It was funded and managed jointly by the Government of South Africa and the United Nations Evaluation Group.

The purpose of this paper is to share the lessons learned from this evaluation so that partner countries and donors maximise the benefits they can derive from joint evaluations. The lessons learned

---

[1]   Oscar Garcia served as a member of the Joint Evaluation Management Group responsible for oversight of the Joint South Africa-UN evaluation.

[2]   Angela Bester served as evaluation specialist responsible for the development theme of the Joint South Africa-UN evaluation. This paper is written in her personal capacity.

can inform principles and good practice for joint evaluations. As the paper is about the lessons learned, it focuses on the process of the evaluation and does not discuss the evaluation's findings and recommendations.

The emergence of joint evaluations

Joint evaluations are gaining increasing currency although the concept has been on the agenda for the past decade. The definition of joint evaluation stipulated by the OECD Development Assistance Committee (DAC) as "An evaluation to which different donors agencies and/or partners participate" is very broad. In reality, there are different types of joint evaluations, varying according to the degree of participation and ownership by the partner countries.

The concept has been used to apply to cases where multiple donors collaborate to conduct a single evaluation of their programmes. The "jointness" of the evaluation is confined to the donors, who lead all aspects of the evaluation, even though there is a degree of consultation with the partner country. A primary motivation behind such joint evaluations is to reduce the burden on partner countries that results from multiple donors conducting individual evaluations. These joint evaluations are probably more accurately described as 'multi-donor evaluations'.

The push for greater participation and ownership of evaluation is, in part, a response to the *Paris Declaration on Aid Effectiveness* (2005). This declaration commits donor countries to improve the effectiveness of development cooperation through the following principles:

• National ownership of the development agenda.

• Harmonisation of donor efforts.

• Alignment of donor efforts with national priorities.

• Mutual accountability for development results.

Building on the *Paris Declaration*, is the *Accra Agenda for Action* (2008) that, amongst other things, commits donors to use the monitoring and evaluation systems of partner countries, to the greatest extent possible. A case in point is the UK Department for International Development (DFID). DFID launched its new evaluation policy in June 2009, committing the organisation to supporting developing countries to play a more equal role in DFID evalua-

tions, and whenever possible, to allow country partners to lead the evaluation.[3]

The push also comes from middle income countries, such as South Africa, who are less dependent upon aid than low income countries. These middle income countries wish to assert their role in joint evaluations even if their capacity is modest by comparison with that of donors. The growth of national and regional evaluation associations and networks can also be seen as a contributing factor to the extent that they support and strengthen national evaluation capacities. [4]

The perceived benefits of joint evaluations are that they:

- *Promote ownership of the evaluation results by the partner countries.* This increases the probability that the evaluation results will be used by them.

- *Stimulate or support the development of national evaluation capacity.* Partner countries can learn from their participation in the evaluation and may also be motivated to take initiative to build their evaluation capacity.

- *Enhance the relevance and effectiveness of aid* by providing evidence for better alignment of aid to the needs and priorities of the partner country.

- *Promote the principle of mutual accountability* for development results.

Joint evaluations with a substantial degree of partner country participation are more accurately termed 'Joint partner-donor evaluations', such as those conducted by DANIDA. In these evaluations the partner country is involved in the governance and management of the evaluations through a designated institution within the partner country. An example of this type of joint evaluation is that of the *Joint Evaluation of the Ghana-Denmark Development Cooperation, 1990-2006* undertaken with the Ghana National Development Planning Commission. Although there is participation of the partner

---

3   Department for International Development (June 2009), *Building Evidence to Reduce Poverty: The UK's policy on evaluation for international development.*

4   See Khayri Ba Tall, O., The role of national, regional and international evaluation organizations in strengthening country-led monitoring and evaluation systems, in Segone M et al (2009), UNICEF, World Bank, IOCE, IDEAS, MICS and Devinfo, *Country-led Monitoring and Evaluation Systems: better evidence, better policies, better evaluation results*, pp 119-134.

country in joint partner-donor evaluations, in practice they tend to be donor-led. This may be by design on the part of the donor agency, or as a result of limited evaluation capacity and resources within partner countries. A study of DANIDA's joint evaluations found that in 8 of the 9 cases evaluated, DANIDA played the lead role on substantive issues while the partner institution played a facilitating role. Issues such as evaluation capacity constraints in partner countries, inappropriate management structures, time constraints and weak communication were identified as factors contributing to the asymmetrical partnership in these evaluations.[5]

There are few examples of partner-led or country-led joint evaluations. Country-led evaluation means that the partner country owns and leads the evaluation agenda. This includes determining what policies or programmes will be evaluated, what the key evaluation questions should be, the evaluation methodology, and how findings will be communicated and used.[6] Examples include the Netherlands-China evaluation of the Development and Environment Related Export Transactions programme; the UNICEF-Bosnia and Herzegovina evaluation of child-focused policies in the social sector; and, the impact evaluations conducted by the World Bank. Figure 1 attempts to illustrate the different categories of evaluation.

Figure 1: Categories of evaluation

Number of donors or development partners ↑	**Multiple donors or development partners** **Evaluation led by donor** *(Multi-donor or multi-agency evaluation)*	**Multiple donors or development partners** **Evaluation led by Country Partner** *(Country-led Joint Evaluation)*
	Single donor or development partner **Evaluation led by donor** *("Traditional" evaluation)*	**Single donor or development partner** **Evaluation led by Country Partner** *(Country-led Joint Evaluation)*

Country partner engagement in evaluation process →

---

5   See Jensen, R.I. and Eriksson, J. (May 2009). *Evaluation study on experiences with conducting evaluations jointly with partner countries*, DANIDA, Ministry of Foreign Affairs of Denmark.

6   Segone, M. (2009). *Enhancing evidence-based policy-making through country-led monitoring and evaluation systems*, in Segone M et al. (2009), UNICEF, World Bank, IOCE, IDEAS, MICS and Devinfo, *Country-led Monitoring and Evaluation Systems: better evidence, better policies, better evaluation results*, pp 17-31.

The joint South Africa-UN evaluation that is the subject of this paper is probably one of the most ambitious country-led joint evaluations. It should be borne in mind that the relationship between South Africa and the UN system should not be conceptualised as a donor-recipient relationship. It is more appropriate to describe the relationship as a *partnership* aimed at delivering public goods in South Africa, the African Continent and globally. South Africa is classified by the World Bank as an upper middle income country and Official Development Assistance (ODA) to South Africa constitutes about 1% of the government's revenue. South Africa is a financial contributor to the UN system through assessed contributions. Furthermore, South Africa makes financial contributions to the UN Country Team and the UN funds, programmes and specialized agencies, and provides funding for numerous programmes and projects in the country implemented in cooperation with UN agencies. South Africa also provides financial and technical assistance to other countries on the African Continent, on a range of matters including security and public sector reform.

Given the context and the nature of the relationship between South Africa and the UN system, the joint evaluation approach may not be replicable in other contexts. The lessons learned however, can begin to inform principles and practice with regard to joint evaluations.

Joint evaluation of the role and contribution of the UN system in South Africa: an overview

The evaluation was initiated by the government of South Africa. The purpose of the evaluation was "...to enhance the role and contribution of the UN system in support of South Africa's policies and strategies for a better South Africa, a better Africa and a better world."[7] In the triennial comprehensive policy review of its operational activities, the UN General Assembly encouraged governments of member states to initiate and conduct joint and country-led evaluations to comprehensively assess the contribution of the UN development system to national development plans and strategies. These plans and strategies include efforts to achieve

---

[7] United Nations and the Government of the Republic of South Africa (June 2009). *Joint Evaluation of the Role and Contribution of the United Nations System in the Republic of South Africa*, p. ix.

the *Millennium Development Goals* (MDGs), and other international agreed development goals[8]. The government of South Africa took the lead in responding to this call.

Discussions between the government and the United Nations Evaluation Group (UNEG)[9] began in June 2006, and only by November 2007 did they reach a decision on how to proceed with the evaluation. A scoping mission took place in February 2008, and the inception phase began in late August 2008. The initial scope of the evaluation had a development focus because the evaluation was initially intended to inform the preparation of a new United Nations Development Agreement Framework, and the Common Country Action Plan for the period 2007-2010.

During the scoping phase the parties agreed to extend the scope of the evaluation beyond the UN's traditional development assistance to South Africa. As the evaluation was intended to inform a policy dialogue between South Africa and the UN, it was felt that the evaluation should be pitched at the level of the strategic policy priorities of the South African government, and at the activities of the UN system supporting these strategic policy priorities.

The final scope of the evaluation covered:

- *Operational activities for development in South Africa* (specifically those covered by the UN Development Assistance Framework and incorporated in the Common Country Action Plan).

- *Humanitarian assistance*, including disaster risk mitigation and disaster preparedness in South Africa and in Africa; participation of the Republic of South Africa in humanitarian response in Africa.

- *Environment*, notably South Africa's contribution to the follow-up of the Conventions on Climate Change, Biodiversity and Desertification.

- *Peace and Security*, including the UN's contribution to areas such as crime and drug control in South Africa and in the region as well as South Africa's contribution to peace-keeping and peace-building.

---

8   United Nations General Assembly A/62/253 August 2007 p.24.

9   UNEG is a professional network that brings together the units responsible for evaluation in the UN system including the specialized agencies, funds, programmes and affiliated organisations. UNEG has 43 such members.

The evaluation period was from 1994 up to 2008, thus covering the entire post-apartheid period. The intention behind the selection of this long time period was to show how the UN system responded to the transition and changing needs of South Africa. The scope of the evaluation covered all UN agencies and funds operating in South Africa. Also included in the scope were the operations of regional offices of UN agencies and funds based in South Africa. It is important to note that although many agencies were covered, this was not an evaluation of each individual agency. The intention was to consider how the UN system collectively contributed to the strategic policy priorities of the Government. The evaluation also elicited the views of UN officials and their South African counterparts based in New York, Geneva and Rome. The evaluation was ambitious in scope, and it was the first time that both the UN and the South African government had embarked on an evaluation of this nature.

In order to underscore the joint nature of this venture, the parties established a Joint Evaluation Management Group (JMEG) responsible for the governance and management of the evaluation. The JEMG comprised representatives from the evaluation offices of the UN including the Department of Economic and Social Affairs, UNDP, UNICEF and the Food and Agriculture Organization (FAO) and senior officials from the central government departments. The Department of Foreign Affairs (now called the Department of International Relations and Cooperation) played the lead role on the South African side and also provided secretariat support to the JEMG. The two parties shared the funding of the evaluation equally, thus reinforcing the joint approach to the evaluation.

An independent team appointed by the JEMG conducted the evaluation. A former Auditor-General of South Africa led the team and was supported by two deputy leaders and four thematic specialists. Five of the team members were South Africans who were not only experts in their respective fields, but also understood the political context and the complex institutional setting in which the UN system carried out its activities in South Africa. The institutional arrangements for the evaluation are shown in Figure 2.

Figure 2: Institutional arrangements for joint evaluation
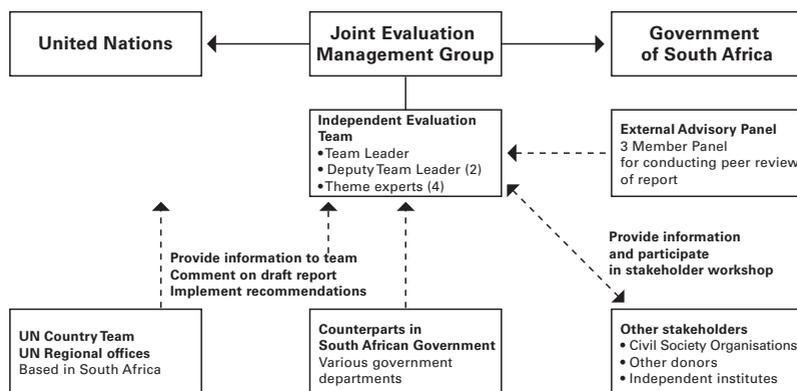
The evaluation used a combination of interviews and desk top studies for data collection. Towards the end of the data collection and analysis phase, the evaluation team convened a stakeholder workshop to share the emerging findings and very draft recommendations. The stakeholder workshop included government, the UN, other donors operating in South Africa, and civil society organisation. An independent external advisory panel reviewed the draft report of the evaluation team. Both the UN and the South African government were given the report to interrogate and to submit written comments, and noting any factual inaccuracies, omissions or errors of interpretation.

Box 1: Quick guide to the South Africa – United Nations joint evaluation	
Evaluation purpose	To enhance the role and contribution of the UN system in support of South Africa's policies and strategies for a better South Africa, a better Africa and a better world
Evaluation criteria	• Effectiveness • Relevance

Thematic areas	• Development • Peace and Security • Humanitarian • Environment • Cross-cutting issues: HIV&AIDS, Gender equality, Human Rights, Partnerships, Capacity development
Funding	• Costs shared equally between the UN and the Government of South Africa
Time frames	• Initial discussions between Government of South Africa and United Nations Evaluation Group (June 2006 to February 2007) • Scoping mission (February 2008) • Inception (August to September 2008) • Main phase (September 2008 to March 2009) • Release of final report (June 2009)
Evaluation Team	• Appointed by Joint Evaluation Management Group. • Team leader (South African), Deputy Team Leaders (South African and International), Team members (3 South African and 1 international)
JEMG Membership	• **South Africa**: Department of Foreign Affairs, Presidency, National Treasury, Statistics South Africa, Department of Public Service & Administration, Office of the Public Service Commission, Department of Provincial Affairs and Local Government. • **UNEG**: Food and Agriculture Organisation, UN-Department of Economic and Social Affairs, UN-Office of Internal Oversight Services, UN Environmental Programme, UNDP, UNICEF, World Food Programme
External advisory panel	• 1 international member and 2 South African members • Selected by the Joint Evaluation Management Group • Provided comments on the draft evaluation reports

Stakeholder workshop	• Organised and chaired by the evaluation team
	• Participants included UN agencies in South Africa and those with regional headquarters in South Africa, government ministries, civil society organisations, donors operating in South Africa
	• Purpose was to share initial findings and test draft recommendations

The purpose of this paper is not to report on the findings and recommendations of the evaluation. It is however important to note that some of the findings were critical of the UN system and of the South African government. The evaluation produced 12 recommendations, including the development of a new comprehensive cooperation framework and the appointment of a new UN Resident Coordinator endowed with executive powers for the whole system, including its budget. Keeping the number of recommendations small was deliberate on the part of the evaluation team. It wanted the UN and the Government of South Africa to focus its efforts on those strategic levers which propel the partnership to its full potential. At the time of drafting this paper, both parties were said to be working on how to take forward the substantive recommendations of the evaluation. Since the time of completing the evaluation, a new government has taken office in South Africa, with significant changes in Cabinet Ministers and the reconfiguration of government departments. On the UN side, a new Resident Coordinator was appointed in the second half of 2009.

Challenges experienced

Given the complexity of the evaluation, it is not surprising that there were a number of challenges at various stages of the evaluation. These are detailed below.

Time constraints

Conducting an evaluation within a limited time frame is a challenge for most evaluations, and more so in the case of this joint evaluation. The reason for this was that the planning and scoping of the evaluation took significantly longer than originally envisaged. This left limited time for the actual conducting of the evaluation and completion of the report. Although initial discussions began as early as June 2006, the evaluation team only commenced work in August

2008 and submitted their final report in March 2009. There was pressure to complete the evaluation before the South African general election which was expected to take place in April 2009. It was imperative to conclude the evaluation in advance of the election as it would have been difficult to secure interviews, and retain the interest of stakeholders, in the run-up to what was to be a watershed election in South Africa's short history of democracy.

As a consequence of the time constraints, the evaluation team had to narrow the scope of the evaluation. Some areas considered important by the government or by the UN simply could not be covered within the time available. Global, continental and regional initiatives could not be explored in sufficient detail, and some initiatives had to be excluded in their entirety.

Methodological challenges

The evaluation faced a number of methodological challenges which had to be addressed effectively to ensure that the evaluation was credible.

It was imperative that the sampling of programmes selected for evaluation was credible to the UN and to the government. The evaluation team spent an inordinate amount of time defining the sampling universe, in order to strengthen the rigor of the sampling. This did not necessarily improve the final selection of the sample and some valuable time was lost in the process.

Although the intention of the evaluation was to provide strategic insights, the unit of analysis was a programme. Moving from the level of programme to a strategic level is often challenging in an evaluation. It proved to be very challenging in this joint evaluation, given the complexities built into the evaluation design. Strategic insights had to be garnered from programmes in four different thematic areas. Not surprisingly, the quality of analysis varied across the four thematic areas as each area had varying qualities of information available on which to base its analysis.

The evaluation was intended to cover the period 1994 to 2008. During the inception phase the evaluation team found that programme documents and reports between 1994 and 2000 were not always available, even where in some cases it was known that the documents existed. There were many instances of officials from government and the UN side no longer being in the department or the country. This meant that the majority of programmes evaluated covered only the last eight out of the fourteen years.

Securing interviews with government officials proved to be a major challenge. It appeared that the existence and purpose of the evaluation had not been communicated adequately to government departments. Even though focal point officials had been identified for government departments targeted for interviews, in a number of instances officials to be interviewed were unaware of the evaluation. There were also events which affected the availability of officials, for example, a major cholera outbreak meant that some senior health officials were not available for interviews. Other important stakeholders such as parliamentarians could not be interviewed as parliament was in recess and parliamentarians had returned to their constituencies.

Misinterpretation of the intentions of the evaluation

A major challenge faced by the evaluation was that some UN officials in South Africa did not fully understand the intentions of the evaluation. They interpreted the evaluation as an assessment of the performance of their individual agencies, so they expressed concern when they felt that the coverage given to their agency in the evaluation was insufficient. Some expected the evaluation team to review all of their major programmes.

Within the South African government also, some officials had the impression that it was the performance of the South African government that was being evaluated. In addition, they did not understand why South Africa's response to UN support was being evaluated.

Creating coherence in the evaluation team

The evaluation team members were all senior professionals who were clearly experts in their respective fields. Given their different disciplines, it was a challenge with the team members to create coherence on the evaluation approach, analysis and reporting. It took considerable effort to develop an evaluation framework that all team members were agreeable to. It also required strong leadership from the team leader and his deputies to ensure coherence in the execution of the evaluation so that it would not be too onerous to integrate the results into one final report. In an effort to facilitate this process, one deputy team leader developed templates for the team.

Lessons learned

Commissioning a joint evaluation as complex as the one discussed here was a high risk for the Government of South Africa and for the UN especially as there was no precedent for this type of evaluation. If not managed effectively, there was a risk to the relationship between the two partners. Any such rift could affect important decisions on future engagement between the UN system and the Government of South Africa. Many lessons were learned in the process and this section outlines the ten we consider most important.

A joint evaluation requires a long lead time

There was a long lead time between the idea of a joint evaluation and the implementation of the evaluation. Given the significance of the evaluation, its potential impact on the relationship between the UN and South Africa, and its potential impact on how the UN conducted its activities in South Africa, there was extensive consultation between the parties. It was also necessary for the parties to have internal consultations with the government departments (South Africa) and other agencies (UN). These consultations usually take a long time and need to be factored into the planning of a joint evaluation. It is important to allow sufficient time for these consultations and discussions as they assist parties in articulating what they want from the evaluation. It is preferable to have major differences in understanding addressed in the early phases of the evaluation so as to avoid disappointment at the end of the evaluation.

Whilst a joint evaluation is likely to have a long lead time, this needs to be balanced in relation to the time available for the implementation of the evaluation. In the case of the South African-UNEG evaluation, the implementation time was very limited relative to the lead time. If too much time is devoted to planning, the scope tends to expand and the available implementation time decreases. This runs the risk of having a rushed evaluation that does not meet the expectations of stakeholders. In a joint evaluation the process is as important as the product.

Keep the governance arrangements clear and simple

Complex evaluations do not necessarily need complex governance arrangements. There is a temptation to make governance structures complicated in a bid to be inclusive. In this joint evaluation the governance structure was simple and so avoided getting bogged

down in unnecessary bureaucratic processes and decision-making. The roles and responsibilities of the JEMG were articulated clearly in the Terms of Reference for the evaluation. There was extensive consultation between the UNEG and the Government of South Africa on the governance structure and management arrangements.

There is often the temptation to want to have the most senior officials in government (Directors-General in the case of South Africa) serve on committees. The membership of the JEMG was based on the potential contribution of the official rather than rank. Officials from the South African side were not the most senior in rank. Because they brought in-depth knowledge of their departments and of their own specific area of work, be it in development assistance, monitoring and evaluation, or international relations, they could make a solid contribution to the JEMG. The UNEG members of the JEMG were evaluation advisers with ample experience in evaluation and clear understanding of the operational activities of the UN.

Contributing funding and other resources reinforce a sense of ownership

The equal sharing of funding of the evaluation reinforced joint decision-making on critical matters, and a sense of ownership on the part of the Government of South Africa and the UN. This ownership of the evaluation increases the probability of the parties taking action on the recommendations made by the evaluation. While such equal sharing of the funding may not be feasible in low income countries, it is feasible for middle income countries to contribute to the funding of the joint evaluation.

Aim to keep the scope of the joint evaluation focused

The Government of South Africa had high expectations of the evaluation that they had initiated. The National Treasury saw the evaluation as an opportunity to redefine the development cooperation relationship between South Africa and the UN, in line with the *Paris Declaration on Aid Effectiveness*. The Department of Foreign Affairs had an interest in the political relationship between South Africa and the UN system. From the UN side the evaluation represented an opportunity for a different approach to evaluating its activities which would inform the ongoing reforms of the UN system. Not surprisingly, the scope of the evaluation was ambitious. There is nothing inherently wrong with an ambitious evaluation scope, provided the scope can be matched with the required resources of time, money and capacity.

In the case of this joint evaluation, compromises were made on the methodology as a result of time and budgetary constraints. For example, in evaluating the UN's role in supporting South Africa's peace and security efforts on the African Continent, time and budgetary constraints did not permit travel to peace-keeping initiatives. The team had to rely on documents and interviews in South Africa.

The scope of the evaluation also made for a large evaluation team. This not only makes management and coordination more challenging, it also makes consistency and coherence harder to achieve. As noted earlier, there was a degree of inconsistency in the level of analysis amongst the four thematic areas, so integrating the findings from these four areas was a challenge.

A joint evaluation requires good in-country capacity

One of the challenges of joint evaluation is to ensure that the "jointness" does not compromise good evaluation standards. As has been shown in other joint evaluations, the partner countries tend to play a marginal role in the evaluation, often due to the lack of good national evaluation capacity. Also, in countries dependent on development aid, the relationship between donor and recipient is not equal. In the case of the South African-UN evaluation, the government representatives on the JEMG were skilled and experienced officials. Collectively, they had extensive experience in engaging with multi-lateral institutions in the political sphere and in the development cooperation sphere. With South Africa playing an influential role in the regional, continental and global arena, these officials had no difficulty in asserting their role in the JEMG. The South African JEMG membership included senior officials from the Presidency, the Public Service Commission and the Department of Public Service who were experienced in managing monitoring and evaluation, and were leading the drive to develop the national monitoring and evaluation system and capacities.

The JEMG was also in a position to draw from a considerable pool of national experts who enjoyed a high degree of professional credibility with the Government of South Africa, and members such as the team leader were well regarded by the UN. Most of the national team members also had experience of conducting assignments in other countries. The experience and expertise of the evaluation team allowed for a highly professional and independent evaluation.

A joint evaluation requires a common understanding of context and evaluation

The South African-UN evaluation was expected to contribute to an important policy dialogue on the future partnership between South Africa and the UN system in South Africa. In order for the evaluation to produce insightful conclusions and recommendations, it was essential that there was a common understanding of the historical, political, economic and cultural specificities of South Africa, as well as an understanding of the context and realities of the UN system as it operated in South Africa and at UN headquarters. This common understanding of context had to be developed within the JEMG, the evaluation team, and between the JEMG and the evaluation team. Without this common understanding between South Africa, the UN and the evaluation team, the joint evaluation ran the risk of producing an evaluation that provided no new insights and that would have produced a litany of recommendations that were of limited value to the policy dialogue. It was also essential that there be a common understanding of evaluation and the evaluation standards of the United Nations so as to ensure an evaluation that complied with UN standards and that was of acceptable quality. The JEMG played a key role in facilitating this common understanding of context, and of evaluation, by bringing together technical expertise on South Africa, the United Nations and on monitoring and evaluation.

A joint evaluation requires mutual trust

Opening the partnership to such comprehensive scrutiny required a high level of trust between the Government of South Africa and the UN. Although the Government of South Africa initiated the evaluation, there was a genuine attempt to implement the evaluation jointly, thus reinforcing the principle of mutual accountability. The JEMG made decisions at critical points of the evaluation on a consensus basis. These critical points included the drafting of the evaluation; identification and selection of the evaluation team members; feedback on the inception report; the organisation of the stakeholder workshop; and, the quality assurance process.[10]

---

10   United Nations and the Government of the Republic of South Africa (June2009). Joint Evaluation of the Role and Contribution of the United Nations System in the Republic of South Africa, p. ii.

A joint evaluation of this kind should be independent and be seen to be independent

In a joint evaluation such as the one undertaken by South Africa and the UN, the independence of the evaluation is paramount. Any perceived lack of independence and objectivity would have damaged the credibility of the evaluation and reduced the prospect for the parties to act on its recommendations. As the majority of team members were South African, it was all the more important for them to be seen to by UN officials to be acting objectively. The evaluation team took concrete actions to ensure the evaluation was independent, and seen to be independent:

- The evaluation team determined the final evaluation design, following comments from the JEMG on the inception report. The decisions on where to limit the scope were not always viewed favourably by the JEMG, but the evaluation team felt strongly that time constraints and the evaluability considerations meant having to reduce the scope of the evaluation.

- The evaluation team organised the stakeholder workshop and made the decision on who should be invited to the stakeholder workshop. As the purpose of the workshop was to test emerging findings and draft recommendations, it was important for the team to identify a very broad range of stakeholders, including those who had not been included in the inception phase or interviewed during the main phase of the evaluation. The team also determined the format and content of the workshop.

- An external panel subjected the draft evaluation report to rigorous peer review. The three member panel were experts in their field and were well-placed to provide a thorough critique of the draft report.

- Each comment received, from the stakeholders on the South African side and on the UN side, was logged and reasons were recorded if a particular comment was not incorporated. This was to demonstrate that the evaluation team gave careful consideration to all the comments received.

By demonstrating their independence, the evaluation team was well positioned to address issues of sensitive to either party. The evaluation team could raise these issues without fear or favour. The team's overt independence engendered a degree of trust that allowed for frank responses on these sensitive issues from the UN officials and the government officials.

A joint evaluation requires leadership from the main players

A joint evaluation of this size and complexity requires leadership from both sides of the partnership, as well as from the evaluation team. This leadership was evident from the South African side with both the Department of Foreign Affairs and the National Treasury playing a key role throughout the process. They played a critical role in shaping the evaluation and bringing their in-depth understanding of the political terrain within which the evaluation was being conducted. The perseverance of the Department of Foreign Affairs, in overcoming internal bureaucratic obstacles, was also important.

On the UNEG side the task force led by UNICEF, FAO and the UNDP provided leadership on technical aspects of the evaluation, and importantly, in championing the evaluation at UN headquarters, thereby facilitating access by the evaluation team to senior officials in UN headquarters. The UN Country Team in South Africa also provided leadership to the process, by convening discussions at appropriate times between the country team and the evaluation team. Changes in leadership in the UN Country Team during the evaluation hampered a more fluid operation.

The quality of the leadership of the evaluation team was also critical for the evaluation process. The leadership of the team was decisive and asserted the independence of the team. The leadership was not vested in one individual. The team leader and deputy team leaders operated as a collective on three different fronts, drawing on their respective strengths:

- The team leader was responsible for leading the engagement with the JEMG and UN Headquarters, and for ensuring that the team delivered on the Terms of Reference.

- The South African deputy team leader provided strategic and critical analysis of the thematic areas, challenging the team experts and facilitating the team's formulation of recommendations.

- The international deputy team leader provided leadership on the technical quality of the evaluation and the report.

Conclusion

The joint evaluation described in this paper was certainly unique and a first of its kind in the UN system. The utilisation of the evaluation has not been fully tested as the new leadership in the South African Government and in the UN Country Office is still reflecting on the recommendations. The experience of the evaluation has been instructive for the UNEG. This experience is influencing the UNEG's thinking towards using joint approaches to evaluation in other middle income countries. The concept of a joint evaluation that covers all UN agencies in a particular country resonates with the "Delivering as One" UN reform initiative.

There are increasing demands from other middle income countries to conduct joint evaluations; how replicable is the South African experience? As demonstrated in this paper, the way a joint evaluation unfolds is very specific to the particular context in which the evaluation is taking place. The existence of good national capacity to lead a joint evaluation, and the existence of independent national evaluation capacity are important determinants of an effective joint evaluation. If there is a genuine desire on the part of donors to encourage joint evaluations, they must support partner countries to build national evaluation capacity. They can do this by beginning with joint evaluations of smaller scope. This will provide opportunity for learning. Donors can also support the development of national capacity by agreeing to allow partner countries to use a percentage of donor funds for joint evaluations.

The South African experience does not provide a blueprint for joint evaluations, but it does provide important principles which can inform other countries wishing to embark on joint evaluations.

The joint South Africa-UN evaluation recommended an independent follow-up evaluation in 2012, to ascertain what progress has been made with the implementation of key recommendations. It is hoped that the lessons learned from the first evaluation will inform the follow-up evaluation in 2012.

References

Department for International Development. (June 2009). *Building Evidence to Reduce Poverty: The UK's policy on evaluation for international development*.

Jensen, R.I. and Eriksson, J. (May 2009). *Evaluation study on experiences with conducting evaluations jointly with partner countries*, DANIDA, Ministry of Foreign Affairs of Denmark.

Khayri Ba Tall, O. (2009). *The role of national, regional and international evaluation organizations in strengthening country-led monitoring and evaluation systems*, in Segone M et al (2009), UNICEF, World Bank, IOCE, IDEAS, MICS and Devinfo, Country-led Monitoring and Evaluation Systems: better evidence, better policies, better evaluation results, pp 119-134.

Segone, M. (2009). *Enhancing evidence-based policy-making through country-led monitoring and evaluation systems,* in Segone M et al (2009), UNICEF, World Bank, IOCE, IDEAS, MICS and Devinfo, Country-led Monitoring and Evaluation Systems: better evidence, better policies, better evaluation results, pp 17-31.

United Nations General Assembly. (2007). *Triennial comprehensive policy review of operational activities of the United Nations development system: conclusions and recommendations*. Report of the Secretary General. A/62/253. August 2007.

United Nations and the Government of the Republic of South Africa. (2009). *Joint Evaluation of the Role and Contribution of the United Nations System in the Republic of South Africa.*

THE ROLE OF STATISTICAL OFFICES IN MONITORING AND EVALUATION SYSTEMS. THE ETHIOPIAN PERSPECTIVE

by Yakob Mudesir,
Deputy Director General,
Central Statistical Agency of Ethiopia

Introduction

Although it is very difficult to trace back exactly when mankind began monitoring and evaluation to improve their daily activities, one can guess that humans have been applying the monitoring and evaluation system since the early Stone Age. This guess is close to reality since development cannot happen without a proper monitoring and evaluation system being in place. In order to understand the close link between the monitoring and evaluation system and development, it is important to understand the basic concepts related to monitoring and evaluation.

Monitoring is the systematic collection and analysis of information as a programme or plan progresses. It is aimed at improving the efficiency and effectiveness of a programme and it helps to keep the work on track. It can let management know when things are going wrong. Monitoring involves establishing indicators of efficiency, effectiveness and impact, and setting up systems to collect information relating to these indicators. It also involves collecting and recording the information, analyzing the information and using the information to inform day-to-day management.

On the other hand, evaluation is the comparison of actual programme impacts against the agreed aims in the strategic plans. It looks at what the aim was, what has been accomplished and how it was accomplished. Evaluation involves looking at what the project or organization intended to achieve, answering the question: "*What difference did it intend to make? What impact did it want to make?*". Moreover, it also focuses on assessing progress towards what was to be achieved, the impact targets and the project strategy and management. It addresses the question "*Was the project effective in following the strategy?*".

Monitoring and evaluation enables one to look at the "bottom line" of development work. Through monitoring and evaluation one can easily review progress, identify problems in planning and/or implementation and so, make adjustments which "make a difference", as planned. Monitoring and evaluation should not be seen as donor requirement; rather it is a management or decision-maker's tool.

Establishing a good monitoring and evaluation system helps to: identify problems and their causes; suggest possible solutions to problems; raise questions about assumptions and strategy; review the direction and approach; provide information and insight; encourage action on the information and insight; and, increase the likelihood that the project will make a positive difference to development.

The importance of statistical data for monitoring and evaluation systems

The development process involves implementing a coherent monitoring and evaluation system that uses statistical data obtained through censuses, surveys, administrative records or vital registrations.

Given the importance of statistical data for monitoring and evaluation systems and hence for a significant role for development, due emphasis should be given to each step in the chain of data acquisition and use: data collection; data processing; analysis; dissemination and archiving. In particular, the *UN Principles for Official Statistics* must be strictly followed during statistical data production.

These are:

1. Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.

2. To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

3. To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.

4. The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.

5. Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.

6. Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

7. The laws, regulations and measures under which the statistical systems operate are to be made public.

8. Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.

9. The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.

10. Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.

In the Ethiopian context, the country's poverty reduction strategy: *Programme for Accelerated and Sustainable Development to End Poverty*, needs a coherent monitoring and evaluation system supported by accurate, timely and easily accessible statistical data.

How the Central Statistical Agency supports the monitoring and evaluation system of the Programme for Accelerated and Sustainable Development to End Poverty

The *Central Statistical Agency of Ethiopia*, as the official national statistical data provider, has been producing, analyzing and disseminating information obtained from surveys, censuses and administrative

records. This activity started in 1960, when the then named Central Statistical Office was established under the Ministry of Commerce, Industry and Tourism. Since then the office has undergone several reorganizations up until 2005, when the office was re-named as the Central Statistical Agency of Ethiopia, and made accountable to the Ministry of Finance and Economic Development.

As with many statistical offices across the world, the Agency is mandated to conduct socio-economic surveys and to compile national statistics from administrative records. The Agency has also been given a mandate to coordinate the national statistical system of the country and build statistical capacity for various ministries, departments and agencies. Since launching the *Integrated Household Survey Programme* in the 1980's, the agency has each year conducted about 11 different surveys. The agency also conducted three consecutive *Population and Housing Censuses* (1984, 1994 and 2007), and the first ever *Agricultural Census* in 2001/02.

The programmes set out in the *Programme for Accelerated and Sustainable Development to End Poverty* require a proper monitoring and evaluation system in order to observe their effects in the different sectors across the country. This monitoring and evaluation system has to be supported by accurate, timely and easily accessible data which helps corrective measures to be taken to ensure successful and timely implementation of the *Programme for Accelerated and Sustainable Development to End Poverty*.

The *Central Statistical Agency of Ethiopia* is expected to enhance its capacity to provide a wide range of socio-economic and demographic data. In particular, at an increased level of disaggregation in a more timely manner and to the necessary quality standard to facilitate the overall monitoring and evaluation process of the *Programme for Accelerated and Sustainable Development to End Poverty*. This will involve reorienting the *Central Statistical Agency* activities to prioritize surveys and other data sources necessary for the *Programme for Accelerated and Sustainable Development to End Poverty*. In addition, it will be important to strengthening the commitment of government and development partners, to improve both the quality and performance of the overall statistical system, through provision of institutional support.

Developing the roadmap document for the statistical work in Ethiopia

The international renowned forum on statistical development: *Partnership in Statistics in the 21st Century (PARIS21)* agreed to ensure that every country develops an integrated plan for the strategic development of their national statistical system, covering all the main data sectors and users. The *National Strategy for Development of Statistics* (NSDS) is being prepared in order to define the best national, long term strategy for developing an efficient and sustainable statistical system within the country. This *National Strategy for Development of Statistics* will have the following main features:

- *promoting a strategic approach* to the development of statistical capacity to help the implementation of nationally prepared and owned strategic statistical development plans;

- *providing an intervention mechanism* to help initiate changes in policy and decision-making through improved availability and use of information and statistical data;

- *improving the use of resources for development*, through better informed decision- making resulting from improved statistical systems and

- *promoting partnerships and improved coordination* among local and international agencies and civil society organizations.

Although they are not directly in accordance with PARIS21, the *Central Statistical Agency* has developed two separate documents namely, the *Medium-term National Statistical Programme* document, for the years 2003/04 – 2007/08, and the *Statistical Strategic Planning and Management* document, which covers the years 2005 – 2009. The first document aims to set up systems and mechanisms to ensure a sustainable flow of statistical data in Ethiopia and thereby, wherever possible, bridge the existing statistical data gap in the medium-term. The second document defines the vision and mission of the Agency and analyzes the strengths and weaknesses of the Agency and the stakeholders. The document provides goals for the next five years, the opportunities and the strategic issues to be addressed, and the financial and manpower requirements for alternative approaches in 2005-2009.

The *Central Statistical Agency*, being the central core governmental institution for the country's statistical work, is responsible for guiding the country's overall statistical activities and should lead

the statistical system of the country. To achieve this, the *Central Statistical Agency* has developed a comprehensive document for the *National Strategy for Development of Statistics* in Ethiopia. The new *National Strategy for Development of Statistics* document examines the overall statistical system for the country in general, and the technical and managerial functions of the *Central Statistical Agency* in particular. In examining the technical and managerial functions, the enterprise-wide vision of the *Central Statistical Agency's* services and management of the enterprise portfolio was considered in the *National Strategy for Development of Statistics*. This document is fully up to standard and comparable with other *National Strategy for Development of Statistics* documents from countries with the greatest experience of statistical work.

Developing this *National Strategy for Development of Statistics* provided a clear roadmap for the statistical work of the country, helped identify a more efficient and directive approach to statistical work in the country, and to improve the statistical information provision of the Agency, all of which is key to measuring delivery of the *Programme for Accelerated and Sustainable Development to End Poverty.*

Improving data quality, timelines, standards and data dissemination system

In order for the Central *Statistical Agency* to answer effectively the various demands for statistical information from the government; donor agencies; researchers; investors; and others, the Agency has to change the way it operates. In particular, it needs to consider data quality, timeliness, data standardization and dissemination system. Improving data quality involves, *inter alia*, revising the survey methodology used by the *Central Statistical Agency*, and developing a system for checking accuracy during data collection and processing. Improving the timeliness of data involves not only producing statistical data in a reasonable period of time after field work, but also producing quarterly, bi- annual and annual data as input for short and medium term economic management and development. Only when the above issues are well addressed and implemented does the investment in collecting and analyzing survey and census data give the best return and respond adequately to the various demands of data users.

Recognizing this fact, these issues were emphasized in the Agency's 2004 *Statistical Strategic Planning and Management* document. In this strategic document an attempt was made to analyze

the current situation with data quality, timeliness, standards and dissemination system. In addition this document set out the *Central Statistical Agency's* future milestones and plans for improvement in these areas. These are explained clearly in the goals stated under: "What the *Central Statistical Agency* envisages achieving during the next five years".

Data storage and dissemination for the *Central Statistical Agency's* purposes has also been enhanced. With the financial support from the UNDP-DAG (Development Assistance Group) and government budget, a number of activities have been initiated with encouraging results. The *Central Statistical Agency* has already started disseminating data through electronic media including the *Welfare Monitoring Survey* (WMS) results. To facilitate information sharing and dissemination, a reliable *Local Area Network* (LAN) using fiber optics technology has already been installed.

In addition, the *Central Statistical Agency* has reached a number of milestones towards establishing the *Central Data Bank*, which is intended to serve as a one–stop data warehouse. It will include census data, household and non-household based survey data, and administrative data with a state of the art data management system. The Agency was able to archive micro and macro level data, in the *Central Data Bank*, together with their related metadata. An *EthioInfo database* was also developed containing indicators from the previous major surveys.

Conducting, processing and analyzing surveys for use by monitoring and evaluation systems

The *Central Statistical Agency* has been involved in providing various socio-economic and demographic data which have now been used for more than four decades as a basis for policy formulation and information-based decision-making. Today, the role of statistical data appears to have reached its climax and every decision made in various sectors must make effective use of these statistical data. Due to this, the demand for statistical data from the Agency is expected to increase dramatically. This information will be needed in the formulation, evaluation and monitoring of the country's overall development programme, and, in particular, for the monitoring of the *Programme for Accelerated and Sustainable Development to End Poverty* and the MDGs.

To satisfy the data needs of the monitoring and evaluation systems, the Central Statistical Agency, as the main data provider of the country, has been and will be conducting a number of surveys in accordance with *Programme for Accelerated and Sustainable Development to End Poverty* process.

Capacity building for data management, dissemination and surveys to monitor and evaluate the Programme for Accelerated and Sustainable Development to End Poverty

Various activities have already been undertaken by the *Central Statistical Agency* to improve data quality, timeliness, standards and dissemination system to achieve the stated objectives. In addition to these, the Central Statistical Agency is now building its capacity for the following:

i. Improving data quality and timeliness by:

 a. establishing a Wide Area Network (WAN) connecting branch offices, which will have a major role;

 b. providing more timely data by decentralizing the data processing activity;

 c. improving the quality of data by making spot correction to the processed data;

 d. making use of Computer Aided Personal Interviews (CAPI), starting with small scaled establishment surveys and price survey.

ii. Developing a statistical standard guideline document containing definitions of concepts, the national coding system, and the universally accepted standards for various classifications and for metadata. This will have a large role to play in harmonizing various statistics produced by different organizations. The linking of administrative records with surveys will also be addressed.

iii. Improving the data archiving and dissemination system. This could involve:

 a. the development of a well secured data warehouse with an archive back-up system (on and off site);

 b. a web-based dissemination system through a dynamic web site;

 c. implementation of a GIS for spatial data analysis and dissemination of various surveys;

d. poverty mapping based on the results of the upcoming census and subsequent surveys.

The capacity building in the dissemination system also needs to deal with improving the Agency's existing printing service.

iv. Strengthening the *EthioInfo* database to provide basic indicators on the *Programme for Accelerated* and *Sustainable Development to End Poverty* and MDGs will help provide time-series data for selected indicators at different levels of aggregation.

In order to provide reliable data for monitoring and evaluation of the *Programme for Accelerated and Sustainable Development to End Poverty*, there will be a need to build the *Central Statistical Agency's* capacity to conducting regular and new surveys through its *National Integrated Household Survey Programme* and to compile reliable Administrative Records.

Provision of reliable data for monitoring and evaluation will also involve improvement of the *Central Statistical Agency's* survey methodology to: provide data with minimal sampling errors; improve field work and data processing activity; ensure quality data by minimizing the non-sampling errors; increasing timeliness of data; and, implement new techniques and technologies for conducting surveys. Timely and reliable data from these surveys and compilation of administrative records will be crucial for the effective implementation of the programmes set out in the *Programme for Accelerated and Sustainable Development to End Poverty*.

It is important to note the need for skilled manpower, as part of the overall capacity building. Fulfilling the logistic and equipment gaps is clearly also important in providing for effective monitoring and evaluation of the *Programme for Accelerated and Sustainable Development to End Poverty*.

The Ethiopian national strategy for the development of statistics

The *National Statistical Development Strategy* for Ethiopia covers the period 2009/10 to 2013/14. It follows on from the *Medium Term Statistical Programme* for the *Central Statistical Agency* which ended in 2008, but it differs in scope. The *National Strategy for Development of Statistics* covers statistical development through the entire statistical system, not just that of the *Central Statistical Agency*.

Preparations for the National *Strategy for Development of Statistics* started early in 2008 with the formation of sector working groups which were charged with identifying the gaps in national statistics. The gaps identified were extensive and not sufficiently prioritised for immediate action. The work was therefore refined by asking key statistical user organisations to identify their three priorities for improvement, and to identify three new, as yet un-met data needs.

To gain a better understanding of the likely problems in implementing a *National Strategy for Development of Statistics*, the outcomes of the *Medium Term Statistical Programme* were examined. The major reasons for non-implementation of the plan objectives were said to be capacity constraints and de-prioritization after the publication of the *Medium Term Statistical Programme*.

In order to improve the *National Statistical System*, six Strategic Themes have now been identified and a number of proposals made under these themes, as set out below.

Strategic Theme 1. Implementation of the Statistics Law

The first theme relates to the coordination of the *National Statistical System,* and requires the establishment of a *National Statistical System* coordination unit in the *Central Statistical Agency.* This unit would manage the development of common standards, classifications and definitions; negotiate legal agreements between each *National Statistical System* partner and the *Central Statistical Agency;* agree any support to be given to partners by the *Central Statistical Agency*; and, manage the data quality assessment process.

Strategic Theme 2. Develop data quality procedures

Quality assessments would be the responsibility of a special *National Statistical System* coordination unit located at the *Central Statistical Agency*. Each statistical collection should be quality assessed periodically, according to a pre-determined timetable. The results should be made available, and an improvement programme agreed. Members of the quality assessment panels may include experts in the subject matter from academia or research establishments as well as suitably qualified civil servants trained in the use of the quality assessment framework. To aid quality improvements and adherence to the statistical standards, ministerial statistical units should be established in each *National Statistical System* partner.

Strategic Theme 3. Enhance advocacy and use of statistics

The *National Strategy for Development of Statistics* preparatory process highlighted problems in stakeholders' knowledge of the availability of statistics. More publicity for statistical products is required, publication calendars should be made available for all *National Statistical System* producers, and all statistical releases should be accompanied by press conferences to raise the users' awareness of statistical products. Access and use of data should be enhanced by a common website for the *National Statistical System* which provides a one-stop-shop to users, and more training given to users. The training would include hands-on support to use survey databases and software.

Strategic Theme 4. Methodological improvements and statistical modernisation

The gaps which were agreed as priorities were for statistics relating to the environment and commercial farming; non-sedentary populations about whom little is known; and for a range of price and economic statistics, particularly for the construction sector and import and export indices. Improvements in the business register, and integrating register data sources, are recommended as a first step, and developing modelling techniques for 'difficult' sectors is recommended. Improving demographic projections were a priority for many *National Statistical System* partners who need accurate denominators in order to express statistics obtained from registers and administrative records as ratios or percentages.

The agricultural and population censuses would continue to be conducted on a 10 yearly cycle. Sample survey programmes would continue on a cycle similar to the *Medium Term Statistical Programme*, but the contents of different surveys would be rationalised and expanded to include more information about the informal sector, cottage industries and small-scale farming of spices and vegetable crops.

Strategic Theme 5: Capacity developments in the National Statistical System

The supply of trained statistical staff will need to be increased, as staff retention is becoming problematic and demand for information and communication technology (ICT) and statistical staff is growing among the international community in Ethiopia. ICT staff is particularly difficult to appoint and retain in the government

service and they are essential to upgrading and improving statistics in ministerial statistical units. Continually upgrading skills is important, and training can act as an incentive to staff to remain within the government service. A full training needs analysis of the *National Statistical System* will be required, and an in-service training programme put in place.

Due to the high turnover of staff, knowledge management is vital. Systems are required to document fully the statistical value chain and to maintain up-to-date metadata. This will ensure continuity even when staff moves on from their current positions.

Technological improvements can vastly improve data quality, timeliness and accessibility. A programme of technological improvement is proposed, including computer assisted data capture and improving networking in the *National Statistical System*. Satellite imagery has proved to be invaluable in undertaking statistical activities, and will be used for estimating environmental and agricultural variables.

Strategic Theme 6: Relationship of the National Strategy for Development of Statistics to the Programme for Accelerated and Sustainable Development to End Poverty monitoring and evaluation system

The process of defining official statistics is an important consideration requiring constant review. The *National Strategy for Development of Statistics* is concerned with regular sustainable statistics and ad hoc statistical activities fall outside the remit of official statistics. While ad hoc surveys are outside the remit of the *National Strategy for Development of Statistics* they may still be published by government bodies and development partners should be alerted about the role of the *Central Statistical Agency* in coordinating statistics and commenting on statistical quality. It is important that all research of a statistical nature is included on the *National Statistical System* website. The policy needs of statistics should be under regular review to ensure that priority needs are met, and changing priorities catered for.

System wide improvements

A number of system wide improvements are included in the strategy. These include coordination of all official statistics, the establishment of ethical and quality standards; and the agreement of common classifications and definitions for the *National Statistical System.* A data quality assessment framework will be developed

and agreed with stakeholders. The framework will be fully tested, staffed and funded before being recommended to the Council for approval as a legally valid instrument.

Improvements are needed in the business register, which will form the basis for better data collection in economic statistics. Modelling for difficult industries will be introduced and training in these techniques will be needed for members of the *National Statistical System.*

Household surveys will be better integrated and more carefully designed to include users' needs in the minimum number of surveys. No more than one major survey will be conducted in any one year. The introduction of new technologies will be tested to improve the quality of fieldwork and to reduce the time taken for data capture and editing. Informal sector and detailed labour force information will be collected more regularly and systematically. A modular survey programme will be considered.

The introduction of technological improvements including a Global Positioning System (GPS) and satellite imagery are planned. All the *Central Statistical Agency* regional offices should be networked before the end of the plan period.

Development partners are asked to support the *National Strategy for Development of Statistics* in a manner which meets the principles of the *Paris Declaration.* To promote donor harmonisation and alignment a donor committee for statistics is recommended, and to ease negotiations and reporting arrangements common systems should be agreed.

Implementation Plan

Business process re-engineering (BPR) is due to start in the *Central Statistical Agency* in the near future and the *National Strategy for Development of Statistics* should provide input into this process. The *National Strategy for Development of Statistics* provides a work programme and performance targets against which progress can be monitored. These targets should be monitored on a six-monthly basis and an annual report prepared on progress, for the Council and for development partners. Performance management will be considered as part of the BPR process.

The capacity building needs are set out in the strategy, as is the budget for the implementation.

Innovation in data dissemination at the Central Statistical Agency of Ethiopia

We are in the era of globalization and, in order to plan properly the progress and to monitor development, every piece of information useful for policy-making and strategic socio-economic decisions should be readily available at the country, regional and international levels. Indicators such as the MDG information on the country's *Sustainable Development and Poverty Reduction Programme*, or data on the country's socio-economic characteristics, should be easily accessible in a timely way.

To be an efficient statistical organization and to accomplish the vision of becoming a recognized statistical office in the region, the capacity around the Agency activities must be strengthened through effective use of the existing ICT for data collection, data processing, analysis and dissemination. This implies an effective utilization of ICT such as Internet, databases, Geographic Information System (GIS), and electronic methods of data archiving and dissemination.

In addition, provision of data involves: harmonizing and integrating statistical data; filling the gap between data production and availability; laying down efficient ICT infrastructure; improving the quality and comparability of data; solving the challenges emerging from data and metadata exchange; and, harmonizing the different standards available in data management system. All these require the strengthening and effective use of ICT at the *Central Statistical Agency*.

Recognizing these facts, the *Central Statistical Agency*'s top priority has been to avidly focus on improving its data collection, management and dissemination system through use of ICT. Accordingly, the Agency has raised the profile of ICT in its organizational structure and set a vision towards the improvement of ICT capacity through a range of action plans. One of these action plans deals with setting up an IT based dissemination system for its micro and macro survey data and related metadata.

Framework to improve dissemination system

In order to facilitate data dissemination and metadata exchange at the global level, the dissemination system has been designed to meet international recommendations and best practices for metadata and archiving.

Figure 1: Data management framework

USERS / PRODUCERS NEEDS			
DATA PRODUCTION Planning, collection, cleaning, processing Quality Control	**DATA DISSEMINATION** Publish Traditional Media, Multimedia, Web	**HARMONIZATION** Standard formats, comparability, multilingual	**COLLABORATION** Disseminate and share knowledge, user-producer dialog, feedback
	DATA ARCHIVING Conversion, Packaging Confidentiality	**ACCESS** Metadata, data & documentation + research tools	**ANALYSIS** Retrieve data for analysis Online Analysis Technical Support

DATA MANAGEMENT FRAMEWORK				
ICT UNIT	**IT CORE ARCHITECTURE**			**DATA TOOLS**
PROJECT MANAGEMENT	**HARDWARE** Server, Workstations, Laptops CD/DVD, Scanners, Printers, Backup, Tools	**TELECOMS** Intranet, Internet, Connectivity, LAN/WAN, Security	**SOFTWARE** OS, DBMS, Development, Web, Multimedia, Office, Statistics, Security	**SPECIALIZED SOFTWARES AND GUIDELINES** Toolkit, DevInfo, WinISIS, DDP, Nesstar, CSPro, XML
TRAINIING				

As can be seen from the diagram above, the IT infrastructure is based on users/producers needs and sets up a data management framework that takes into consideration the establishment of an implementation body, IT core architecture and data archiving and dissemination tools.

This IT based archiving and dissemination system is made possible by establishing a central databank to archive all documentation and microdata obtained from various surveys and censuses, and developing a user friendly dissemination system for easy access by users of the Agency's data. This in turn will help establish a one-stop data shop for our users.

To achieve this, the project calls for the adoption of the standards and specialized data management tools including the *International Household Survey Network Microdata Management Toolkit* and the UNICEF supported *DevInfo* package.

The data management framework that we are putting in place takes into account the global recommendations and best practices in micro and macro level data management. It plans for step by step improvement to reach high quality online accessibility to metadata and availability aggregated data, online analysis and microdata access services.

Utilization of new technologies for data capturing

The scanning technology

One of the time consuming process in census data processing is data capture. For instance, data capture for the first population and housing census of Ethiopia took about two years using a main frame system. The second census took over a year using a standalone PC system involving 180 data entry clerks using 90 PCs.

In order to capture the data quickly and provide the result of the census as quickly as possible, scanning technology was implemented for the 2007 *Population and Housing Census*. The actual data capture for the *Census of Sedentary Areas* (excluding Somale and Afar Regions), was started in mid July, 2007 and the scanning work was completed in November 2007. This shows more than 95 percent of the population data was captured within four months.

Before implementing the scanning technology for the third *Population and Housing Census of Ethiopia*, insights were obtained from the study tours made to Tanzania and Ghana where scanning was used for the population and housing census. This study tour helped the *Central Statistical Agency* to learn from the success and failures in these countries and to conclude that scanning could be very applicable for the Ethiopian census data processing.

The Personal Digital Assistants

The *Central Statistical Agency* has deployed Personal Digital Assistants (PDAs) to facilitate the price data collection and electronic transfer from the field to its head office. The price data management process at *Central Statistical Agency* was reviewed and compared with the PDA system. The system satisfies statistical standards' requirements and also timeliness requirements for price data users.

Achievements in the Central Statistical Agency dissemination system

In order to improve its data archiving and dissemination system described in the IT frame work above, the *Central Statistical Agency* has been engaged in various IT related improvement activities. The following are the main achievements obtained so far related to IT based archiving and dissemination systems since 2004.

Establishment of the Central Databank

A Central Databank has been established for the microdata and today contains over 5000 data and documentation files covering about 75 surveys. Over 48 surveys have been archived using the *World Bank Microdata Management Toolkit*, thereby making the metadata compliant with the *Data Documentation Initiative* specifications, as recommended by the *International Household Survey Network*.

Website development

The *Central Statistical Agency* was able to develop its official website and used this website not only to provide adequate information about the Agency's activities but also to provide various statistical data to the users. Most importantly, the *Central Statistical Agency* is using its website to disseminate national statistics, provide access to all documentation and related metadata for all surveys archived in the central data bank, and to serve as a portal for other facilities available to access the Central *Statistical Agency's* data, such as the *EthioInfo database*

CD_ROM Publishing

CD_ROM products have been prepared and a CD based electronic dissemination system has been put in place. Accordingly the *Central Statistical Agency* is able to disseminate its survey results, documentation and metadata archived using the IHSN toolkit on CD-ROMs

The EthioInfo Database

EthioInfo is a customized adaptation of *DevInfo*, a worldwide user-friendly software that helps to organize and present data in a result based environment with unique features linking it to strategic monitoring and evaluation of policies such as MDG and National Poverty Reduction Strategies. As a common platform for indicators related to human development, *EthioInfo* facilitates data sharing and indicator harmonization at global, regional and country level by making statistics available to a wide audience. It allows presentation of data through tables, graphs and maps.

Figure 2: EthioInfo

Introducing the Geographic Information System

The GIS is a collective term commonly used for computer systems that manipulate geographic data. The *Central Statistical Agency* has been heavily engaged in use of GIS. In the last two years the *Atlas of Ethiopian Rural Economy* was produced in collaboration with the Ethiopian Development Research Institute and the International Food Policy Research Institute.

A digitized Wereda (district) map is being produced which shows each Enumeration Area (EA) within the Wereda and shows Urban/Rural Kebele (dwellers association) boundaries, cultural and natural features, educational and health facilities, as well as other infrastructures. This will be made available to users.

The *Rural Facilities Atlas* currently being developed is again one the great achievements. This atlas uses the data obtained during cartographic work for the 2007 population and housing census. It shows the distribution of basic facilities in the rural areas of the country. These facilities include: schools; health posts; water points; agricultural development offices; roads etc. These atlases are produced for all districts and contain a rich geo-referenced data for development planning.

The atlas aims primarily to report at district level so as to give information as accurately as possible on the spatial pattern of basic facilities and services within the boundaries of each district. The atlas will also be highly important to the academic community. By providing students and various researchers with information on the level of services and facilities in rural parts of Ethiopia, these maps will be of value to a range of social sciences. The maps show the development level of rural parts of the country and help to provide a good understanding of local social conditions and basic rural infrastructure in Ethiopia.

Satellite Imagery

In 2000, the UN Statistics Division published a *Handbook on Geographic Information System and Digital Mapping,* to help countries prepare for census activities and, in particular, to apply GIS technologies to census mapping operations. The traditional role of maps in census and survey operations has been restricted to supporting enumeration and to presentation of results in the form of an atlas. However, the accuracy, content and timeliness of satellite imagery make it useful for pre-census activities such as ensuring consistency and facilitating census operations. During the actual census, these maps support data collection and monitor census activities. At the post-census stage, the maps make it easier to present, analyze and disseminate census results in the form of an atlas showing the complete picture in areas of interest (study areas). By adopting a digital approach through the use of satellite imagery, the UN's belief that it will result in planners, policy makers, public administrators etc., being able to make better-informed decisions, is fully endorsed. With multiple years of imagery, archive data will provide a large proportion of the required coverage from previous years, ensuring that a viable product can be produced within a relatively short timescales.

One of the advantages of satellite imagery is in its political immunity and non-invasive sensing of any geographical area, regardless of the conditions on the ground which might otherwise dictate its accessibility, such as natural hazards, remoteness or hostilities. Whilst satellite imagery cannot measure population levels directly, it can help to identify populated areas for census planners and so reduce the potential distortion of the final figures from pastoralism.

Multiple uses of the satellite imagery

As mentioned earlier, it is anticipated that satellite imagery will be a valuable national asset, able to underpin a range of other national, regional and local governance activities. These applications will include

support to activities such as food security; provision of health services; planning of housing; water supplies; and, transportation infrastructure to support the urbanization process in Ethiopia. Current mapping products are a fundamental part of creating a sustainable economic and social society as well as helping to address civil instability by providing accurate information about the natural and manmade environment.

Conclusion

It is very important to note the strong link between development, the monitoring and evaluation system and statistical data. Therefore, an effort should be exerted to improve the monitoring and evaluation system and the generation of statistical data to sustain the development processes. This needs the provision of a due concern from the political leadership of the country; the development partner; academia; and, the professionals involved in the monitoring and evaluation system and the statistical data production. This again involves continuous capacity building in the field of statistics; proper funding for the statistical data production, utilization of new technologies and above all, mainstreaming the statistics work in the country's development agenda.

References

Central Statistical Agency of Ethiopia. *National Strategy for Development of Statistics.*

Central Statistical Agency of Ethiopia. *Medium Term National Statistical Programme 2003/4 – 2007/8.* http://www.csa.gov.et

Central Statistical Agency of Ethiopia. *Statistical Strategic Planning and Management 2005-2009.* http://www.csa.gov.et

Central Statistical Agency of Ethiopia. *Atlas of Ethiopian Rural Economy*, 2006, Ethiopian Development Research Institute & International Food Policy Research Institute

Ministry of Finance and Economic Development of Ethiopia (2007). *Programme for Accelerated and Sustainable Development to End Poverty.* July, 2007, http://www.mofed.gov.et

World Bank. *International Household Survey Network Microdata Management Toolkit,* , http://www.csa.gov.et

UN Statistics Division. (2008). *Handbook of Geographic Information Systems and Digital Mapping.*

UN Statistics Division. *UN Principles for Official Statistics.* , http://www.unstats.un.org

World Bank. *Guidelines for developing NSDS, Partnership in Statistics in the 21st Century*, http://siteresources.worldbank.org

World Bank. *Marrakech Action Plan for Statistics*, www.worldbank.org

BUILDING CAPACITY IN DEVINFO IN THE CARICOM REGION

by Philomen Harrison,
Project Director, Regional Statistics Sub-Programme,
Caribbean Community (CARICOM) Secretariat

Regional training workshops in DevInfo

DevInfo is a database system that helps to organize and present data and metadata. This data and metadata are being compiled across the world, on social development indicators based on the United Nations Millennium Development Goals (MDGs). The DevInfo system consists of: indicators; time periods; geographic areas; sub-populations; units; sources; and, data values. The latest version of DevInfo (version 6.0), is map-enabled and web-enabled. The map-enabled feature implies that data can be easily presented in the form of thematic maps, providing information at the smallest geographic level available, such as communities, villages etc. The web-enabled feature implies that the data can be readily disseminated on the world-wide web. In addition, DevInfo 6.0 is more user-friendly, and so, less training is required for first-time users of the User and Administrative modules. DevInfo is not limited to the dissemination of social indicators. It can disseminate statistics and indicators in any area such as economic and environmental statistics.

In the Caribbean Community (CARICOM), early training in DevInfo began in August 2004, in Panama, with a regional workshop on version 4.0, for Latin America and the Caribbean. It was funded by the United Nations Children's Fund (UNICEF). It was attended by approximately 22 people from international, regional and national agencies from CARICOM. Of the 22 people attending, four representatives were from national statistical agencies within the Region and two were from the CARICOM Secretariat Regional Statistics Programme. The focus of this initial training was to provide a general introduction to DevInfo.

A follow-up "Training of Trainers" workshop, in December 2004, was attended by five people from the CARICOM Region. There were no representatives from the national statistical agencies of CARICOM at this workshop.

In September 2006, the UNICEF's Office in Barbados was the location for 33 participants from 19 Caribbean and Latin American countries to receive training to becoming DevInfoLAC focal points. The focus of this training was on DevInfo 5.0. The workshop was funded by UNICEF, the United Nations Development Programme (UNDP), and DevInfoLAC, in partnership with the CARICOM Secretariat. It was organized in close collaboration with the United Nations Development Group (UNDG).

One of the major outcomes of the workshop was the identification of three countries (Saint Lucia, Barbados and Guyana) as pilot countries for the implementation of DevInfo. Other workshop outcomes were for CARICOM to champion the implementation of DevInfo within the Member States, and for UNICEF to hire a consultant for CARICOM, for a period of six months, to give technical support to these countries. The consultant's role was, amongst other tasks, to enhance the harmonization of data.

Following this 2006 workshop, Saint Lucia and Guyana began the implementation of DevInfo 5.0. As a direct result of this, Saint Lucia's HelenInfo 1.0 desktop version was launched in April 2007. Substantial assistance was given to Saint Lucia from the European Commission and the UNDP. In Guyana, UNICEF assisted in the roll-out of DevInfo by planning workshops and meetings for the Ministries and various agencies in Guyana. These were supported, in terms of facilitation of the training, by the CARICOM Secretariat.

In May 2008, basic and advanced parallel training workshops were held in Barbados in DevInfo 5.0. From CARICOM, 23 participants attended the basic level training, of which 15 were from the national statistical agencies. Six participants attended the advanced level. Of these six, three were from the national statistical agencies and two were from the CARICOM Secretariat Regional Statistics Unit.

In 2009, the Caribbean Community (CARICOM) Secretariat was allocated funding by the United Nations Children's Fund (UNICEF), for activities related to the enhancement of DevInfo. These activities focused on the countries of the Organisation of Eastern Caribbean States (OECS), and on the OECS Secretariat. The activities included in the DevInfo funding for 2009 were implemented by the Regional Statistics Sub-programme of the CARICOM Secretariat in response to a proposal submitted to UNICEF.

The activities in 2009 included a CARICOM Regional DevInfo 6.0 Training Workshop for the Caribbean, held in Georgetown, Guyana.

This training was attended by 35 participants from 19 Caribbean countries, and the CARICOM Secretariat. It focused on the User and Administration application and was facilitated by the DevInfo Support Group, India. The funding provided by UNICEF also supported participation by countries at a back-to-back workshop on CensusInfo, organized by the United Nations Statistics Division (UNSD) in collaboration with UNICEF, the CARICOM Secretariat and the DevInfo Support Group of India.

CensusInfo is based on the DevInfo technology and was developed by the UNSD, the United Nations Population Fund (UNFPA), and UNICEF, in order to help countries disseminate their Census data. CensusInfo was recommended by the Fifteenth Meeting of the CARICOM Regional Census Coordinating Committee (RCCC), as a platform for the dissemination of data from the 2010 round of population and housing censuses. The RCCC is one of the main monitoring mechanisms for coordination and support of census-taking in CARICOM. It comprises Directors of Statistics/Chief Statisticians who have the role of Census Officers in their respective countries; the CARICOM Secretariat; and, representatives from other regional and international organizations.

Training activities undertaken at the national level – development of national adaptations of DevInfo

Jamaica had, effectively, already developed an adaptation of DevInfo, desktop version (JamStats) before the interventions discussed above. This has since been launched on the Internet. Following the training in Saint Lucia in 2006, Saint Lucia, literally on their own, with minimal support from the CARICOM Secretariat, developed their DevInfo desktop version adaptation, called HelenInfo. This was launched in 2007. Trinidad and Tobago was able to obtain funding for more intensive training in order to develop and launch its own adaptation of DevInfo, CTTInfo (2008). This adaptation is intricately linked to the development thrust of this Member State's Vision 2020.

The CARICOM Secretariat, with funding provided by UNICEF, undertook training sessions in approximately seven countries and for the Secretariat of the OECS. In most cases there were at least two training sessions. Arising from this support, adaptations (mainly desktop versions), were launched in Dominica (NatureIsleInfo – 2007);

Barbados (BimInfo – 2008, developed but not yet launched); Guyana (Guy_DInfo-2009); and, in St Vincent and the Grenadines (VinInfo – 2009). Suriname created an adaptation (ABSInfo) for internal use and is moving towards launching a further adaptation of this. The OECS Secretariat also launched an adaptation (OECSInfo – 2009), and Montserrat is currently working on the development of its own adaptation. Grenada also is working towards the launch of its adaptation, SpiceIsleInfo. The CARICOM Secretariat developed its adaptation, CARICOMInfo to disseminate statistics and indicators for all countries of the Region, in a wide range of areas. CARICOMInfo was officially launched in October 2009 at the events commemorating Caribbean Statistics Day.

Conclusions and some considerations for the future

Evidently much has been achieved over approximately a three-year period of regional training, which included post-training activities of countries and training executed mainly by the CARICOM Secretariat in its role as champion in the development of DevInfo in the Region. In the course of the roll-out of the various adaptations, some central issues have emerged. These include: the role of statistics in the development thrust of the CARICOM countries; enabling greater use of the statistics through the DevInfo adaptations; and, the use of DevInfo as a mechanism to increase the accessibility of data which would otherwise be unavailable within the National Statistical Systems. As a consequence of addressing these issues, greater coordination thas been achieved within the National Statistical Systems.

The main features of the DevInfo system can serve to improve the range and quality of statistics available across the Region which can feed into the development process. The fact that the DevInfo system makes provision for documentation of metadata is critical to: monitoring; improving data quality; and, achieving greater harmonization of statistics across CARICOM. The mapping capabilities of DevInfo are highly useful in targeting economic and social interventions by governments, and other organizations, aimed at improving the quality of life of the people within and across Member States. In addition, DevInfo is "Standard Data and Metadata Exchange" (SDMX), compliant in that provides for easy exchange of data from countries to international and regional organizations. It therefore eliminates duplication of effort in having to prepare varying formats for submitting the same data to these different organizations.

Much work remains to be done to enable the implementation of DevInfo in the remaining countries of CARICOM. One critical issue that requires attention is to achieve greater harmonisation/standardisation of the adaptations so as to enable greater linkages at the regional level for data feeding into CARICOMInfo and into OECSInfo. This would eliminate duplication of effort on repetitive data entry at these levels. There is much to be gained by standardization of: variable names; measurement units: format for the description of metadata; and, other such formats in the various template designs.

The creation of a regional standardized template is one crucial factor for moving forward. In addition, a mechanism must be found to assist those countries who do not qualify for funds from UNICEF. Support to those countries to continue or to begin the development of their national adaptations must continue. Efforts must be supported to improve the basic statistics, indicators and other information which are to be disseminated through the DevInfo system.

Enhancement in statistical coordination, across statistics departments and units in line ministries and other agencies, also requires continued support. Other considerations for future support are continued training and updating of skills, and interaction with the DevInfo Support Group from India, as well as the provision of equipment and staffing support.. There is much to be gained from the continued inclusion of the DevInfo Support Group in our regional training efforts.

Finally, the decision to use CensusInfo as one of the main mechanisms to disseminate the 2010 Round of Population and Housing Censuses data poses great challenges, but also fantastic opportunities for rapid dissemination of the census results and the use of these results by the population: policy makers; researchers; school children; and, other users. The efforts of UNICEF in enabling the diffusion of DevInfo must be commended. Fundamentally, the use of the information in achieving improved development outcomes – in, amongst other areas, education, health, crime, housing and poverty alleviation – must be encouraged.

Annexes

ANNEX I: AUTHORS VITAE

ANUSHYANTHAN, Vaidehi is an Assistant Director Planning of the Department of Foreign Aid and Budget Monitoring of the Ministry of Plan Implementation (MPI). She has over 15 years of Public Sector experience in planning, monitoring and evaluation of development programmes. She is currently the lead M&E officer for the water supply and sanitation sector and had extensive experience in working at the sub national level in the District Planning Secretariat. She actively contributed in the institutionalization of Managing for Development Results (MfDR) and actively contributed to design the National Evaluation Information System of the MPI.

BAMBERGER, Michael has a Ph.D. in Sociology from the London School of Economics. He has worked on the evaluation of development programmes in more than 30 developing countries in Africa, Asia, Latin America and the Middle East. He worked for 13 years with non-governmental organizations throughout Latin America. During his 22 years with the World Bank, he worked as Advisor on Monitoring and Evaluation with the Urban Development Department, as Asia training coordinator for the Economic Development Institute, and as Senior Sociologist in the Gender and Development Department. Since retiring from the World Bank in 2001, he has carried out consulting and teaching assignments for the Asian Development Bank; Swedish International Development Cooperation Agency; UK Department for International Development (DFID); United Nations Development Program (UNDP); U.N. Department of Economic and Social Affairs; UNICEF; UN Secretariat for Asia and the Pacific (ESCAP); U.N. Evaluation Office; U.S. Agency for International Development (USAID); World Bank; World Food Program; and, for several private consulting firms. Bamberger has published widely on development evaluation, including a co-authored 2006 Sage publication on: *Conducting evaluations under real-world constraints*; and several recent World Bank publications on: *Conducting quality impact evaluations under budget, time and data constraints; Influential evaluations; Institutionalizing impact evaluations and reconstructing baseline data*.

BESTER, Angela is a Director of Deloitte & Touche (Southern Africa) and focuses on improving government performance through monitoring and evaluation. Angela has extensive experience in the public sector in South Africa and Australia, with a strong focus on human development and governance. Previously, she served as Director-General of the Office of the Public Service Commission, South Africa and was instrumental in establishing the Commission's monitoring and evaluation practice. Angela also served as Director-General of the National Department of Social Development, South Africa and Deputy Director-General in the Department of Land Affairs where she established monitoring and evaluation units and built monitoring and evaluation capacity. Her work as Governance Adviser for the Department for International Development (DFID) deepened her knowledge of international development cooperation. Bester served as a Board Member of the South African Monitoring and Evaluation Association from 2006 to 2007. She holds a Masters in Business Administration from the University of Technology, Sydney, and MA (Sociology) from the University of New South Wales (Australia).

BOYLE, Shanelle is an organizational behavior psychologist with eight years of applied research and evaluation experience. As a Research Assistant at EVALCORP (www.evalcorp.com), Boyle assists clients in all stages of the evaluation process. She has conducted studies to examine the use of various evaluation capacity building strategies to improve organization members' knowledge, skills, and beliefs about evaluation, and currently works with public and private sector organizations to help build their capacity to conduct high quality evaluations and to use findings to improve their programmes and meet organizational goals. Boyle has presented nationally at psychology and evaluation conferences and is co-author with Hallie Preskill of an American Journal of Evaluation article titled, *A Multidisciplinary Model of Evaluation Capacity Building (2008).* She holds Master and Doctorate degrees from Claremont Graduate University, in Psychology in the area of Organizational Behavior, with a specialization in evaluation capacity building.

CASTRO, Manuel Fernando is Senior Evaluation Officer in the Independent Evaluation Group of the World Bank where he has been since 2008. Before joining the Bank he was the Director of Colombia's National Results Based Management and Evaluation System (SINERGIA), Director of Public Policy Evaluation, Special Investments

Director and Advisor to the General Director of the National Department of Planning. In these positions he held various responsibilities in the formulation of national development plans and social development strategies as well as in the design and evaluation of major development programmes. He has over 14 years of experience as an executive, university professor and international consultant in almost 20 countries, for the Inter-American Development Bank (IDB), the World Bank and the IMF in public sector reform, performance management, and monitoring and evaluation.

DORADO, Diego is currently the director of the Office of Evaluation of Public Policy in Colombia's National Planning Department (DNP). In the execution of his duties as director he works jointly with the Office of the President and the Ministry of Finance in the development and institutionalization of the performance based management tools, in order to strengthen the inputs used by policy-makers. Previously he served as manager of the Unified Public Investment System, position he held from 2007 until August 2009. He is also known for his experience as an international consultant in themes related to public investment and project formulation. He has worked as a consultant of the ECLAC and the IDB. He is an economist specialized in evaluation with a Masters Degree in Cooperation and International Affairs.

FEINSTEIN, Osvaldo is advisor to the Spanish Evaluation Agency, professor for the Masters on Evaluation at the Universidad Complutense de Madrid and at the European Institute for Public Administration, and a senior evaluation consultant with the World Bank and the International Fund for Agricultural Development (IFAD). He was a manager at the World Bank's Evaluation Department and a senior evaluator at IFAD, where he created PREVAL (Latin American programme for the development of evaluation capacities). Feinstein was a member of the Monitoring and Evaluation Panel of the Science Council of CGIAR (the Consultative Group on International Agricultural Research), a professor with the FLACSO Ecuador Masters in Development Studies, a visiting professor at several European and Latin American universities and institutes, and a consultant with the United Nations Evaluation Group (UNEG); Global Environment Fund (GEF); UNDP; CEPAL; ILPES; UNESCO; the UN Fund for Science and Technology; the African Development Bank; the Inter-American Development Bank; and the ILO. He has worked in monitring and evaluation and devel-

opment in almost all Latin American and Caribbean countries, and some Asian and African countries. He has written and edited articles and books on evaluation, development and economics.

FETTERMAN, David is President and CEO of Fetterman & Associates, an international evaluation consulting firm. He is concurrently a Professor of Education at the University of Arkansas at Pine Bluff and the Director of the Arkansas Evaluation Center. He is also a Collaborating Professor, Colegio de Postgraduados and Distinguished Visiting Professor at San Jose State University. He has 25 years experience at Stanford University. He was the Director of Evaluation in the School of Medicine; Director of Evaluation, Career Development, and Alumni Relations; and Director of the Master of Arts Policy Analysis and Evaluation Program in the School of Education. He also served in the senior administration at Stanford University. He was formerly Professor and Research Director at the California Institute of Integral Studies; Principal Research Scientist at the American Institutes for Research; and a Senior Associate and Project Director at RMC Research Corporation. Fetterman is a past-president of the American Evaluation Association (AEA) and the American Anthropological Association's Council on Anthropology and Education. He is the co-chair of AEA's Collaborative, Participatory, and Empowerment Evaluation Division. David received the Outstanding Higher Education Professional Award in 2008. He received both the Paul Lazarsfeld Award for Outstanding Contributions to Evaluation Theory and the Myrdal Award for Cumulative Contributions to Evaluation Practice. He was elected a fellow of the American Anthropological Association and the Society for Applied Anthropology. He also received one of the 1990 Mensa Education and Research Foundation Awards for Excellence. Fetterman has contributed to a variety of encyclopedias and is the author of 10 books, including *Ethnography: Step by Step*, and over 100 articles and reports.

GARCIA, Oscar A. is senior evaluation adviser at the Evaluation Office of the United Nations Development Programme (UNDP). He oversees programmatic evaluations for Africa and Latin America and the Caribbean. He co-chairs the UNEG task force on Country Level Evaluations. Prior to joining UNDP he was founder and managing director of *Proactiva,* a management consulting firm specialized in strategic management and development evaluation. He held positions in public administration in Bolivia, where he was

Director General for Trade Policies and resource mobilization manager for the Social Emergency Fund. He consulted throughout Latin America for IADB, WB, UNDP, and UNFPA. Garcia holds a Masters degree in Organizational Change Management from the New School University, in New York and an MBA from the Catholic University of Bolivia in association with the Harvard Institute for International Development. He is co-author of several publications and evaluation reports including Latin American NGOs and their needs for organizational development, the Role of UNDP in Poverty Reduction Strategies, and the evaluation of the UNDG contribution to the Paris Declaration.

GIOVANNINI, Enrico graduated in Economics at "La Sapienza" University of Rome. He continued his studies at the Institute of Economic Policy in the same Faculty, specialising in econometric analysis. In December 1982 he was employed by the Italian National Institute of Statistics (Istat). In December 1989 he became Research Director at the National Institute for Short-term Economic Analysis, where he took care, particularly, of monetary and financial analyses. In January 1992 he moved to the National Institute of Statistics (Istat). From December 1993 to May 1997 he was head of the National Accounting and Economic Analysis Department. In December 1996 he was appointed Central Director of Statistics on Institutions and Enterprises. Up until 2001 he was Director of Statistics and Chief Statistician of OECD and co-chair of the Board of the Global Project on *"Measuring the progress of societies"*. He is now the President of the Italian Statistical Institute.

HAMILTON, Billy is a consultant in tax, fiscal policy and related issues. In this capacity, he works with a wide range of public and private clients. He also writes a regular column on state tax issues for the national publication *State Tax Notes*. He is an adjunct professor at the LBJ School of Public Affairs of the University of Texas at Austin. Prior to January 2007, Hamilton was Chief Deputy Comptroller of Public Accounts of Texas. He has also served on special assignment as Co-Executive Director of the California Performance Review for Governor Arnold Schwarzenegger, *A Government for the People for a Change*, released in July 2004.

HEIDER, Caroline has over 20 years of experience in evaluation. She is the director of the office of evaluation of the World Food Programme (WFP) and vice chair of the UN Evaluation Group. She provides strategic direction to the office and has been the driving force behind improving evaluation at WFP. During this period, she has been called upon to participate in peer reviews of evaluation functions or of evaluations of sister organizations. Prior to her current position, Caroline was the deputy director or the office of evaluation at the International Fund for Agricultural Development, where she was responsible for guiding professional staff and quality assuring their evaluation reports, revising evaluation guidelines, and managing evaluations. At the Asian Development Bank, Caroline was the team leader for most of her evaluations, including country portfolio evaluations, impact evaluations, thematic evaluations, and programme and project evaluations. While at the United Nations Industrial Development Organization she was a central part of a small core team that conducted corporate evaluations and supported the self-evaluation system. Heider authored *Conceptual Framework for Evaluation Capacity Development* in a forthcoming IDEAS/World Bank publication (2010).

HSU, Douglas is a technical advisor with the DevInfo Support Group, focusing on content development and training on behalf of the DevInfo database system endorsed by the UN to assist countries in monitoring progress towards achieving the Millennium Development Goals. He is a graduate of the Woodrow Wilson School of Public and International Affairs, Princeton University. His experience of nearly 20 years in the public and not-for-profit sectors includes serving as a program evaluator with the U.S. Government Accountability Office and working with various U.S.-based NGOs focused on environmental issues, U.S. inner-city development, and international aid assistance, particularly in India.

KHAYRI BA TALL, Oumoul is currently the president (2008-10) of the International Organization for Cooperation in Evaluation (IOCE), past president (2005-07) and board member of the African Evaluation Association (AfrEA) and a founder of the Association Mauritanienne de Suivi-Evaluation (AMSE). She is currently involved in initiatives to organize a network dedicated to strengthen evaluation in French speaking countries around the world (Réseau Francophone d'Evaluation, RFE). She has written several papers and articles, and has delivered speeches on topics

such as Aid and Development, and Evaluation Capacity. She has 21 years of professional experience in various but related fields from auditing, accounting, evaluation, organisational development, micro-entreprise, microfinance, community-based and development fields, including seven years of evaluation experience and 18 years of auditing. She is the Executive Director of her own audit and management consultancy business in Nouakchott, Mauritania. Oumou has a Masters in Business Administration (1995) and is a member of professional accounting bodies in Mauritania and in Senegal.

KUSEK, Jody Zall has provided leadership in the area of monitoring and evaluation at the World Bank for eight years. She is advisor on results and capacity development at the World Bank Institute. Previously, she led the Bank's Global HIV/AIDS Monitoring and Evaluation Group (GAMET) that aims to strengthen the use of HIV/AIDS data to support national and sub-national policy and program decision-making in over 50 countries. Prior to that, she was the Cluster Leader for Getting Results at the World Bank's Africa Region, and co-authored the Bank's business process to design and use a Results-based country assistance strategy that is now in use, Bank-wide. Earlier, Kusek worked for the Clinton-Gore Administration in the United States, designing and implementing the Government performance and results act. She is the co-author of *Ten steps to results-based monitoring and evaluation*. She is also the author of numerous papers on government management, results based management and poverty monitoring system development.

KUZMIN, Alexey is an international evaluation consultant. He is the president of the Process Consulting Company, an organization development and programme evaluation group based in Moscow, Russia. Kuzmin has been working as a management and organization development consultant since 1987. Since the mid-90s he has specialized in programme and project evaluation, conducting over 100 evaluations of projects and programmes in Russia, Commonwealth of Independent States and Central and Eastern Europe. His evaluation clients include UN agencies, international development agencies, private foundations, NGOs, businesses and government entities. He is a lecturer at the Moscow School for Social and Economic Sciences where he has taught a programme evaluation course since 2004. He also conducts evaluation training for various audiences on a regular basis nationally and internationally. Kuzmin is active in professional associations in the region and internationally. He has been a member

of the International Programme Evaluation Network (IPEN) Board of Trustees since 2000 and a member of the American Evaluation Association (AEA) since 1999. He chaired the AEA's International Committee in 2005-2006 and co-chaired the AEA's International and Cross-Cultural Evaluation Topical Interest Group in 2003-2006. Kuzmin has a number of publications in English and in Russian on evaluation and evaluation related issues. In particular, he was a co-editor and a co-author of the Russian-American *book "Program evaluation: Methodology and practice"* published in Russia.

MENON, Saraswathi is currently the Director of the Evaluation Office in the United Nations Development Programme and the elected Chair of the United Nations Evaluations Group that brings together the heads of evaluation of all UN organizations. She has worked in UNDP for more than 20 years. She was a member of the team of authors who wrote the first six Human Development Reports. Subsequently she worked on UNDP programmes as Deputy Chief of the Regional Programme in the Regional Bureau of Asia and the Pacific. She served as UNDP Deputy Resident Representative in Nepal (1999-2000) and as UN Resident Coordinator and UNDP Resident Representative in Mongolia (2000-2003).Prior to joining UNDP she taught sociology in Madras University in India. She has a Ph.D. in Sociology from Jawaharlal Nehru University in New Delhi. Her Ph.D. dissertation was on caste and land control in Thanjavur district during the nineteenth century, and she continues to be interested in multi-dimensional issues of poverty.

MORRA IMAS, Linda is widely known as the "mother of IPDET", the International Program for Development Evaluation Training, a World Bank and Carleton University programme. She has been Chief Evaluation Officer and Evaluation Capacity Building Adviser for the World Bank Group until her recent retirement. She continues to advise on evaluation capacity development for the World Bank's Independent Evaluation Group and to co-direct IPDET and other training programmes. She was a director for the U.S. General Accounting Office (now Government Accountability Office) in charge of all evaluations for the U.S. Congress of education and employment and occupational health and safety programmes. She is co-author of the book published by the World Bank *The Road to Results: Designing and Conducting Effective Evaluations*. She is an Adjunct Professor with the Faculty of Public Affairs at Carleton University and Secretary of the International Development Evaluation Association (IDEAS),

for which she is currently leading an effort to develop competencies for those engaged in international development evaluation.

MUDESIR SEID, Yakob is Deputy Director General of the Central Statistical Agency (CSA) of Ethiopia. He gained his B.Sc Degree in Mathematics from the Addis Ababa University and an M.Sc degree in the field of Mathematical Statistics with distinction from the University of the Orange Free State, Bloemfontein, South Africa. Since his graduation, he has been working at the CSA of Ethiopia in various posts. He worked as a programmer being responsible for developing computer programmes for processing survey data; as a team leader for computer operations and ICT Department Head in the Agency. He was appointed as Deputy Director General in 2006 being responsible for the Operations, Statistical Methodology, Data Processing and ICT development of the Agency. As Deputy Director General, he led, among others, the operations of the 2007 Population and Housing Census of Ethiopia; the development of the Ethiopian National Strategy for the Development of Statistics; the development of the Ethiopian National Data Archive; and the introduction of new techniques and technologies for survey methodology, data capture and dissemination systems.

NAIDOO, Indran is the Deputy-Director General responsible for monitoring and evaluation (M&E) at the Public Service Commission (PSC) of South Africa. Naidoo has provided leadership for M&E at various levels for the past 15 years, having started the country's first M&E unit in 1995 at the Department of Land Affairs, before moving to the PSC in 2000, where he helped develop M&E systems and engaged with stakeholders on evaluation results. He trained as a teacher, and holds degrees in English, Education and Geography (Masters: West Virginia University, USA). He has actively supported the development of M&E in Africa, and was Conference Co-Chair for the 2004 AFREA event, which the PSC co-hosted, as well as the 2007 and 2009 SAMEA conferences. He served as the Deputy-Chairperson of SAMEA and represented AFREA on the network of Networks on Impact Evaluation (NONIE). He has presented papers at several international conferences, the most recent being the Latin American and Caribbean Evaluation Network (Bogota, 2009), the IDEAS Development Evaluation Seminar (Johannesburg, 2009), and the European Evaluation Association Conference (Lisbon, 2008). He has also served on the United Nations Joint Country Evaluation advisory panel. He has co-

taught a course at the International Programme for Development Evaluation Training (IPDET) held in Canada in 2008 and 2009.

NDAMOBISSI, Robert has spent a total of 13 years serving with UNICEF as Chief of Planning, Monitoring and Evaluation in his Country, the Central African Republic (7 years), Niger (3 years) and Mali (3 years). Before joining UNICEF, he led from 1990 to 1997, the Directorate of Demographic and Social Statistics, and the Bureau of Census at the Ministry of Planning in the Central African Republic. He coordinated and completed many national census and surveys (Demographic and Health Surveys, 4 Multiple Indicator Cluster Surveys, Nutrition Survey, etc) and he has managed many Country Programme evaluations and situation analyses of children and women. He has published several articles with *Karthala Edition* (in French), external Universities and UNICEF. He was co-chair of the fourth AfrEA conference organized in Niger in January 2007 involving 600 participants and was the Coordinator of Niger's National Network on Monitoring and Evaluation from 2005 to 2007. He is now strengthening the national capacity for country-led M&E in Mali and promoting the newly established network for evaluation in Mali (APEM). Since 2008, he has been the leader of the inter-agencies monitoring and evaluation technical group for UNDAF in Mali.

NUWAMANYA, Kanengere Joe holds an M.Sc. in Agriculture Economics from Makerere University, Kampala, Uganda, and an M.Sc. in Soil Science from University of Nairobi, Kenya. He has been working in the Uganda Public Service (Office of the Prime Minister) since 1987, as a Commissioner (1987-2009) and as a Director (2009-to-date) for policy coordination and monitoring. He has been instrumental in the policy coordination and monitoring and evaluation capacity development in Government. He was part of the team who designed the National institutional coordination framework; the Poverty Eradication Action Plan (PEAP) results framework; the monitoring and evaluation strategy for the 5-Year National Development Plan (2010/2011- 2014/2015); and, the current sector-based monitoring and reporting framework. He is the chairperson of the National Monitoring and Evaluation Technical Working Group which brings together all the people in the public sector, academia, NGOs and donors, to discuss issues concerning monitoring and evaluation capacity development in the country. He is also the National Coordinator of the Joint Assessment Framework (JAF) which is used to measure performance of Government under the Joint Budget Support Operations by donors.

O'BRIEN, Finbar is Director of Evaluation at UNICEF. He has worked in international development for 25 years, fifteen of which were spent in Africa. He was formerly the Head of Evaluation and Audit with the Department of Foreign Affairs in Ireland and also served as Chair of the DAC Evaluation Network. O'Brien's major interests in recent years have been institutional arrangements for evaluation and the promotion of joint and country-led evaluations.

O'SULLIVAN, Rita is an Associate Professor at the University of North Carolina, where she also is Director of Evaluation, Assessment, and Policy Connection, a unit in the School of Education. O'Sullivan specializes in collaborative evaluation techniques which enhance evaluation capacity and utilization among educators and public service providers. She has successfully used collaborative evaluation approaches with programmes in agriculture, arts, science, education, health, community development, and family support with government, museums, and non-profit organizations throughout the United States, as well as overseas. In 2004, she completed *Practicing Evaluation: A Collaborative Approach* (Sage), and is senior author of *Programs for At-Risk Student: A Guide to Evaluation* (Corwin Press, 1993). Since 1997, O'Sullivan has travelled to Puerto Rico, Kazakhstan, Southern Sudan, Russia, Brazil, Guinea, Barbados, and the U.S. Virgin Islands to conduct evaluations, make presentations, and provide evaluation training. In addition to her scholarly contributions to the field of evaluation via articles and presentations, O'Sullivan served as Secretary/Treasurer of the American Evaluation Association from 1992-1997. In 2001 she was presented the first lifetime Distinguished Service Award from the North Carolina Association for Research in Education, and in 2002 received the Ingle Service award from the American Evaluation Association.

OSWALT, Kris is an international expert in the design and implementation of information systems. He has over 30 years of experience in software application development for database management systems, geographic information systems and knowledge management systems. Oswalt is the President of Community Systems Foundation, a not-for-profit organization founded in the USA in 1963, and the Executive Director of the *DevInfo* Support Group where he has been instrumental in the design of *DevInfo* database technology. Oswalt has provided technical assistance in more than 80 countries to a broad range of international organizations, including: UNICEF,

UNFPA, UNDP, WFP, UN-Habitat, UNESCO, WHO, DFID, USAID, World Bank, UN Statistics Division, OECD, John Snow Inc., International Science and Technology Institute, Management Sciences for Health and the U.S. Library of Congress.

PRESKILL, Hallie is the Executive Director of FSG's Strategic Learning and Evaluation (www.fsg-impact. org). FSG is an international strategy consulting, research, and evaluation firm committed to accelerating the pace of social progress. Prior to joining FSG, Preskill spent more than 20 years in the academic world, teaching courses in programme evaluation, organizational learning, appreciative inquiry and training and organization development. Her books include *Reframing Evaluation through Appreciative Inquiry (2006); Building Evaluation Capacity: 72 Activities for Teaching and Training (2005); Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance & Change (2001, 2009); Evaluative Inquiry for Learning in Organizations (1999);* and *Evaluation Strategies for Communication and Reporting (2005).* Preskill was the 2007 President of the American Evaluation Association. She received the American Evaluation Association's Alva and Gunnar Myrdal Award for Outstanding Professional Practice in 2002, and the University of Illinois Distinguished Alumni Award in 2004. Over the years, she has provided consulting services and workshops, and has conducted evaluations in schools, healthcare, foundations, government, and corporate organizations.

PRON, Nicolas Charles has been working for the United Nations for 16 years, out of which 12 years were spent in the field in Africa and Asia, where he implemented UNICEF Country Programmes. Pron is currently posted in New York where he manages the DevInfo flagship project, a high profile UN inter-agency initiative to monitor progress towards achieving the Millennium Development Goals. He is a national of France. He holds a Masters degree in International law from the Sorbonne University and a Masters degree in Development law from the Rene Descartes University in Paris.

QUESNEL, Jean Serge is Senior Facilitator at the United Nations System Staff College, Adjunct Professor at Carleton University and Professeur associé at the École nationale d'administration publique of Quebec. He was Director of Evaluation at the United Nations Children Fund (UNICEF), the Inter-American Development Bank

(IADB) and the Canadian International Development Agency (CIDA) where he also was Director of Policy coordination, of Management Improvement and of international development programmes in Asia, Africa, Latin America and the Caribbean. He is a senior advisor to the Board of the International Organisation for Cooperation in Evaluation (IOCE) and the first official member of the International Development Evaluation Association (IDEAS).

QUINN PATTON, Michael is an independent evaluation consultant. He is former President of the American Evaluation Association and author of six evaluation books including *Utilization-Focused Evaluation*, 4th ed.; *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*; and *Qualitative Research and Evaluation Methods*, 3rd ed. He co-authored *Getting to Maybe: How the World Is Changed*, a book that applies complexity theory to social innovation. He is an instructor in The World Bank's International Program in Development Evaluation Training (IPDET) in Ottawa each summer; the American Evaluation Association's professional development courses; and, The Evaluators' Institute.

RAMIREZ, Ana Carolina is part of the team of the Office of Evaluation of Public Policy in Colombia's National Planning Department (DNP), were she has worked since 2007. In the execution of her duties she is in charge of leveraging the main responsibilities of the director's office and the management of the international cooperation agenda for the strengthening of SINERGIA. Previously she worked at the IDB Headquarters in the Office of Evaluation and Oversight, were she made part of the Evaluability Project as well as Chile's housing programs' evaluations.

RIDER SMITH, David has been an adviser to the Office of the Prime Minister and Uganda Bureau of Statistics in Uganda since 2007. His work is focused on strengthening the use of statistics and evidence in decision making across Government through supporting policy and systemic reform. He works extensively on building institutional capacity to establish systems for producing timely and credible reviews and evaluations, and in harmonizing budget reforms and budget support of donors, with wider performance measurement efforts of Government. Since 1995, he has worked extensively in sub-Saharan Africa on performance issues, and in social research, within the United Nations system, for bilateral agencies and

academia. He is the author of books on institutionalizing impact orientation in public sector institutions, and on the dimensions of participation in agricultural research.

RIST, Ray has had a distinguished career which includes a range of high profile government and academic appointments. He has been a visiting professor at several prestigious universities, and has been a consultant to many national and international organisations, including the World Bank, OECD, DFID, IADB, and a range of corporations, House and Senate committees in the United States. The focus of much of this consulting has been on public sector performance, especially that of results-based management and measurement. He has been an advisor to senior government officials in more than 50 countries. Rist is currently an advisor to the World Bank, co-director of the International program for Development Evaluation Training (IPDET) and President of IDEAS. He has authored or edited 26 books and has authored more than 140 articles and monographs

RUSSON, Craig is currently a Senior Evaluation Officer with the International Labour Organization in Geneva, Switzerland. He is currently a Co-Chair of the United Nations Evaluation Group Training Taskforce. He previously worked as an Evaluation Manager with the W. K. Kellogg Foundation and as a Senior Researcher at The Evaluation Center on the campus of Western Michigan University. Craig has served on the Boards of Directors of the American Evaluation Association, the Michigan Association for Evaluation, and the Evaluation Capacity Development Group. Together with many others, Craig was involved in the creation of the International Organisation for Cooperation in Evaluation (IOCE) for which he received the Robert Ingle Award for service from the American Evaluation Association. Craig was the founder and is Co-moderator with Jim Rugh of XCeval. Craig holds a Masters degree in educational psychology, a Masters in Business Administration, and a Ph.D. in agricultural education.

SANZ LUQUE, Belen is Head of Evaluation at the United Nations Development Fund for Women (UNIFEM) where she has been since 2007. She is the co-chair of the Taskforce on Gender Equality and Human Rights evaluation of the United Nations Evaluation Group. Her academic background is in Social Anthropology, with postgraduate degrees in Anthropology from the University of London and in Evaluation of Public Policies and Programmes from the

Complutense University of Madrid. During the last 10 years she has specialized in evaluation of development programmes and policies, with a focus on gender equality policies and strategies. She has focused on the development of systems, methods and tools for organizations to use evaluation as a function for learning and improvement, and on the integration of gender equality into evaluation. Previous to the United Nations, Sanz worked as Head of Evaluation in the Spanish Development Cooperation. She acted in 2006 as vice-Chair of the DAC/OECD Evaluation Network, and was a member of the DAC/OECD Gendernet Bureau between 2003 and 2004.

SEGONE, Marco is responsible for the decentralized evaluation system portfolio at the UNICEF Evaluation Office and serves as Senior Advisor to the IOCE Board. Previously, he was Regional Chief, Monitoring and Evaluation in the UNICEF Regional Office for Central and Eastern Europe and the Commonwealth of Independent States (CEE/CIS), during which time he represented UNICEF on the Board of Trustees of the International Programme Evaluation Network (IPEN). During his 17 years in international development, he worked in Bangladesh, Pakistan, Thailand, Uganda and Albania in integrated development projects. In 1996 he joined UNICEF to work for the UNICEF Regional Office for Latin America and the Caribbean; UNICEF Niger, where he founded and for two years coordinated the Niger M&E Network (ReNSE); and UNICEF Brazil, where he was one of the founders and coordinator of the Brazilian Evaluation Network. In 2003 he was elected Vice-President of IOCE and was one of the founders of the Latin America and the Caribbean Network for Monitoring, Evaluation and Systematization (RELAC). Segone authored and/or edited about 30 books and articles, including *Country-led M&E systems*; *Bridging the gap. The role of M&E in evidence-based policy making*; *New trends in development evaluation*; *Creating and developing evaluation organizations: Lesson learned from Africa, Americas, Australasia and Europe*; and *Democratic evaluation.*

SIVAGNANASOTHY, Velayuthan is the Director General, Department of Foreign Aid and Budget Monitoring of the Ministry of Plan Implementation for the Government of Sri Lanka. He has more than 25 years of senior-level public sector experience in development planning, management, and monitoring and evaluation. He served as a Co-Chair of the International Reference and Management Groups on the Evaluation of the Implementation of the Paris

Declaration on Aid Effectiveness, and is member of the Community of Expert Evaluators in the South Asia Region. He represented Asia in the OECD negotiating team that formulated the Accra Action Agenda for the Paris Declaration on Aid Effectiveness. He also serves as a Steering Committee member of the Asia Pacific Community of Practice on Managing for Development Results (MfDR) and Aid Effectiveness Community of Practice (AE-CoP). He has lectured at various universities and development agencies and is an author of *The M&E Concept and Principles in Sri Lanka* and developed a MfDR training module for the Post Graduate Institute of Management.

ANNEX II: ABBREVIATIONS

AAPSCOMS	Association of African Public Service Commissions
ADB	Asian Development Bank
ADEA	Association for the Development of Education in Africa
AEA	American Evaluation Association
AfDB	African Development Bank
AFrEA	African Evaluation Association
ANC	African National Congress
Aim	Africa Impact Evaluation Initiative
AP	President's Advisor, Columbia
AU	African Union
BMAU	Budget Monitoring and Accountability Unit, Uganda
BPR	Business Process Re-engineering
CARICOM	The Caribbean Community
CCAP	Common Country Action Plan
CEDAW	UN Convention on the Elimination of All Forms of Discrimination against Women
CES	Canadian Evaluation Society
CGIAR	The Consultative Group for International Agricultural Research
CONPES	Consejo Nacional de Política Económica y Social. National Council for Political Economy and Society, Columbia
CPA	The Certified Public Accountant
CVP	Constitutional Values and Principles, South Africa
DANIDA	Danish International Development Agency
DBSA	Development Bank of Southern Africa
DEPP	Office of Evaluation of Public Policy, Columbia
DeGEval	German Evaluation Society

DDPD	Departmental Direction of Prospective Development, Benin
DFID	Department for International Development, UK
DIFP	Office of Public Investment and Finances, Columbia
DLA	Department of Land Affairs, South Africa
DNP	National Planning department, Columbia
DTn	Departmental Technical Office, Columbia
ECB	Evaluation Capacity Building
EDPRS	Economic Development and Poverty Reduction Strategy, Rwanda
EES	European Evaluation Society
EFA	Education for All
ePMS	Online Project Monitoring System, of Sri Lanka
EPRC	Economic Policy Research Centre
EVALNET	Evaluation Network, South Africa
FAO	Food and Agriculture Organization
GAD	Gender and Development
GDP	Gross Domestic Product
GIS	Geographical Information System
GOSL	Government of Sri Lanka
GRPP	Global and Regional Partnership Programmes
HIPC	Highly Indebted Poor Countries
HRBA	Human Rights Based Approach
IAS	Information and Analysis Section
IBSTPI	The International Board of Standards for Training, Performance, and Instruction
ICT	Information and Communication Technology
IDEAS	International Development Evaluation Association
IEG	Independent Evaluation Group, of the World Bank

IHSN	International Household Survey Network, Ethiopia
3IE	International Initiative for Impact Evaluation
ILO	International Labour Organization
IPDET	International Program for Development Evaluation Training
IT	Information Technology
IMIS	Integrated Management Information System
INSAE	National Institute of Statistics and Economic Analysis; Benin
IOCE	International Organization for Cooperation in Evaluation
JAF	Joint Assessment Framework, Uganda
JEMG	Joint Evaluation Management Group, South Africa
KPIs	Key Performance Indicators
LAPC	Land and Agrarian Policy Centre
LGs	Local Governments
M&E	Monitoring and Evaluation
MDAs	Ministries, Departments and Agencies
MDG	Millennium Development Goals
MfDR	Managing for Development Results
MFPED	Ministry of Finance, Planning and Economic Development, Uganda
MOF	Ministry of Finance, Columbia
NBO	National Budget Office, Columbia
NDS	National Development Strategy
NGO	Non-Governmental Organization
NIMES	National Integrated Monitoring and Evaluation System of Uganda
NISR	The National Institute of Statistics, of Rwanda
NMES	National Monitoring and Evaluation System, Benin

NONIE	Network of Networks for Impact Evaluation
NSDS	The National Strategy for Development of Statistics
ODA	Official Development Assistance
OECD	Organization for Economic Cooperation and Development
OECD/DAC	Organization for Economic Cooperation and Development/ Development Assistance Committee
OECS	Organisation of Eastern Caribbean States
PARIS21	Partnership in Statistics in the 21st Century
PDA	Personal Digital Assistant, Ethiopia
PEAP	Poverty Eradication Action Plan
PFA	Beijing Platform for Action
PMAU	Poverty Monitoring and Analysis Unit, Uganda
PRSC	Poverty Reduction Support Credit
PRSP	Poverty Reduction Strategy Papers
PSC	Public Service Commission of South Africa
PSRs	Poverty Status Reports
RCCC	Regional Census Coordinating Committee, of CARICOM
SAMEA	South African Monitoring and Evaluation Association
SDMX	Standard Data and Metadata Exchange
SINERGIA	National Monitoring and Evaluation System of Columbia
SEDESOL	National Monitoring and Evaluation System of Mexico
SLEVA	The Sri Lanka Evaluation Association
SOPS	State of the Public Service Reports, South Africa
RSS	Really Simple Syndication web feed technology
ToT	Training of Trainers

TRIAMS	Tsunami Recovery Impact Assessment and Monitoring System
UBOS	Uganda Bureau of Statistics
UN	United Nations
UNCT	United Nations Country Teams
UNDAF	United Nations Development Assistance Framework
UN-DESA	Department of Economic and Social Affairs
UNDG	United Nations Development Group
UNDP	United Nations Development Programme
UNEG	United Nations Evaluation Group
UNESCO	United Nations Education, Science and Culture Organization
UNFPA	United Nations Population Fund
UNHS	Uganda National Household Survey
UNICEF	United Nations Children's Fund
UNORC	United Nations Office of the Resident Coordinator
UNSD	United Nations Statistics Division
UPPAP	Uganda Participatory Poverty Assessment Process

Notes

Notes

Notes

Notes

Notes

The previous publications in this series, *Country-led monitoring and evaluation systems* published in 2009; *Bridging the Gap: the role of monitoring and evaluation in evidence-based policy making* published in 2008; and *New trends in development evaluation* published in 2007, have already made important contributions to the debate on how monitoring and evaluation can enhance evidence-based policy making. This publication furthers the debate by providing new analysis on strengthening country monitoring and evaluation capacities, and in moving from policies to results through country-led monitoring and evaluation systems.

Within the United Nations system, there is a growing recognition that national capacity development for monitoring and evaluation systems is an essential part of the broader support to policy reform and to promoting national ownership of evidence-based policy making.

UNICEF continues to be committed to supporting national monitoring and evaluation capacities development and is looking forward to continuing collaborations and work with major partners at global, regional and country levels to help achieve the goal of developing strong national systems that will enhance sustainable development results.