



Country-led monitoring and evaluation systems

Better evidence, better policies,
better development results

It contains
DevInfo
databases

In partnership with:





Country-led monitoring and evaluation systems

Better evidence, better policies,
better development results

In partnership with:



It contains
DevInfo
databases

The Evaluation Working Papers (EWP) are documents that present strategic evaluation findings, lessons learned and innovative approaches and methodologies. We would like to encourage proposals for relevant papers which could be published in the next EWP issues. Papers can be prepared by UN staff, consultants and partners.

For additional information and details please contact Marco Segone, Senior regional advisor, monitoring and evaluation, msegone@unicef.org

- ISSUE #1: Regional strategy to strengthen the monitoring and evaluation function in CEE/CIS, 2005
- ISSUE #2: Comparative analysis of major trends in evaluations and studies in CEE/CIS region, 2005
- ISSUE #3: Quality matters. Implementing the Regional Evaluation Strategy, 2006
- ISSUE #4: Accessing qualified consultants: The Regional Evaluation Roster, 2006
- ISSUE #5: New trends in development evaluation (joint paper with IPEN, with prefaces by Presidents of IDEAS and IOCE), 2006
- ISSUE #6: Developing UNICEF and partners monitoring and evaluation capacity, 2006
- ISSUE #7: Evaluation of the Family support and foster care project and Prevention of infant abandonment and de-institutionalization project in Georgia. In: Child Protection series, 2006
- ISSUE #8: Evaluation of Global education project in Central Asian Republics, 2006
- ISSUE #9: Knowledge leadership for children. Evaluations, studies and surveys supported by UNICEF CEE/CIS in 2004-2007, 2007
- ISSUE #10: A formative evaluation of Parenting programmes. In: Resources for early childhood, 2007
- ISSUE #11: Evaluation of the Family education project in Uzbekistan. In: Resources for early childhood, 2007
- ISSUE #12: Bridging the gap: The role of monitoring and evaluation in evidence-based policy making, 2008.
- ISSUE #13: What a UN evaluator needs to know. Introductory course on what is evaluation and how it is designed and managed, 2008
- ISSUE #14: Joint Country-led evaluation of Child-focused policies within the Social Protection Sector in Bosnia and Herzegovina. Published by the Directorate of Economic Planning, Government of Bosnia and Herzegovina, and UNICEF Bosnia and Herzegovina, 2008
- ISSUE #15: The Regional monitoring and evaluation facility. An innovative client-oriented technical assistance system, 2008
- ISSUE #16: Regional thematic evaluation of UNICEF's contribution to Juvenile Justice System reform in Montenegro, Romania, Serbia and Tajikistan. In: Child Protection series, 2008
- ISSUE #17: Regional thematic evaluation of UNICEF's contribution to Child Care System reform in Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan and Uzbekistan. In: Child Protection series, 2008
- ISSUE #18: Emerging challenges for children in Eastern Europe and Central Asia. Focus on disparities, 2008

Photo Credits: UNICEF/MOL/00741/Pirozzi

Disclaimer:

The opinions expressed are those of the contributors and do not necessarily reflect the policies or views of UNICEF. The text has not been edited to official publication standards and UNICEF accepts no responsibility for errors. The designations in this publication do not imply an opinion on legal status of any country or territory, or of its authorities, of the delimitations of frontiers.

Extracts from this publication may be freely reproduced with due acknowledgement.

Country-led monitoring and evaluation systems

Better evidence, better policies,
better development results

Editor

Marco Segone

Authors

Marie-Hélène Adrien

Finbar O'Brien

Petteri Baer

Kris Oswald

Michael Bamberger

Robert Picciotto

Oswaldo Feinstein

Nicolas Charles Pron

Enrico Giovannini

Jean Serge Quesnel

Denis Jobin

Ray Rist

Megan Grace Kennedy

Jim Rugh

Oumoul Khayri Ba Tall

George Sakvarelidze

Jody Zall Kusek

Marco Segone

Hans Lundgren

Daniel Vadnais

Keith Mackay

Contents

Prefaces:

<i>Finbar O'Brien, Director, Evaluation Office, UNICEF Headquarters</i>	2
<i>Ray Rist, President, International Development Evaluation Association (IDEAS)</i>	4
<i>Oumoul Khayri Ba Tall, President, International Organization for the Cooperation in Evaluation (IOCE)</i>	6

Editorial:

<i>Marco Segone, Senior Regional Advisor, Monitoring and Evaluation, UNICEF Regional Office for CEE/CIS, and former Vice President, International Organization for the Cooperation in Evaluation (IOCE)</i>	8
---	---

Part 1

Why country-led monitoring and evaluation systems?

Enhancing evidence-based policy making through country-led monitoring and evaluation systems.

<i>Marco Segone, Senior Regional Advisor, Monitoring and Evaluation, UNICEF Regional Office for CEE/CIS, and former IOCE Vice President</i>	17
--	----

Evaluating development. Is the country the right unit of account?

<i>Robert Picciotto, Visiting Professor, King's College, London and former Director General, Evaluation, the World Bank</i>	32
---	----

The strategic intent. Understanding strategic intent is the key to successful country-led monitoring and evaluation systems.

<i>Jean Serge Quesnel, Professor at the United Nations System Staff College, Adjunct Professor at Carleton University and Professeur Associé at the École Nationale d'Administration Publique of Quebec</i>	56
---	----

Supporting partner country ownership and capacity in development evaluation. The OECD DAC evaluation network.

<i>Hans Lundgren, Head of Evaluation Section, Development Co-operation Directorate, OECD</i>	77
<i>Megan Kennedy, Consultant, OECD</i>	

Country-led evaluations. Learning from experience.

<i>Osvaldo Feinstein, Professor at the Master in Evaluation, Complutense University, Madrid, and former Manager, Operations Evaluation Department, the World Bank</i>	96
---	----

Country-led impact evaluation. A survey of development practitioners.

<i>Marie-Hélène Adrien, President, Universalis, and former President, IDEAS</i> <i>Denis Jobin, Vice President, IDEAS, and Manager, Evaluation Unit, National Crime Prevention Center, Public Safety, Canada</i>	102
---	-----

The role of national, regional and international evaluation organizations in strengthening country-led monitoring and evaluation systems. <i>Oumoul Khayri Ba Tall, President, International Organization for Cooperation in Evaluation (IOCE)</i>	119
Bringing statistics to citizens: a “must” to build democracy in the XXI century <i>Enrico Giovannini, Chief Statistician, OECD</i>	135
Proactive is the magic word. <i>Petteri Baer, Regional Advisor, Statistical Division, UN Economic Commission for Europe</i>	158

Part 2

Good practices in country-led monitoring and evaluation systems

Building monitoring and evaluation systems to improve government performance. <i>Keith Mackay, Evaluation Capacity Development Coordinator, Independent Evaluation Group, the World Bank</i>	169
Getting the logic right. How a strong theory of change supports programmes which work! <i>Jody Zall Kusek, Lead Coordinator of Global HIV/AIDS Monitoring and Evaluation Group, the World Bank</i> <i>Ray C. Rist, Advisor, the World Bank, and President, International Development Evaluation Association (IDEAS)</i>	188
RealWorld Evaluation: conducting evaluations under budget, time, data and political constraints <i>Michael Bamberger, Independent consultant</i> <i>Jim Rugh, Independent international program evaluator</i>	200
Strengthening country data collection systems. The role of the Multiple Indicator Cluster Surveys <i>Marco Segone, Senior Regional Advisor, Monitoring and Evaluation UNICEF CEE/CIS</i> <i>George Sakvarelidze, Monitoring and Evaluation Specialist UNICEF CEE/CIS</i> <i>Daniel Vadnais, Data Dissemination Specialist UNICEF Headquarters</i>	238
Strengthening country data dissemination systems. Good practices in using DevInfo <i>Nicolas Pron, DevInfo Global Administrator, UNICEF Headquarters</i> <i>Kris Oswalt, Executive Director, DevInfo Support Group</i> <i>Marco Segone, Senior Regional Advisor, Monitoring and Evaluation, UNICEF CEE/CIS</i> <i>George Sakvarelidze, Monitoring and Evaluation Specialist, UNICEF CEE/CIS</i>	252

Making data meaningful. Writing stories about numbers*.
UNECE, Statistical Dissemination and Communication,
Conference of European Statisticians.....268

Annexes

Authors vitæ294
Abbreviations303
What is DevInfo? 311

*This article was originally published by UNECE. Reprinted with the permission of UNECE.



Prefaces and Editorial

Prefaces:

<i>Finbar O'Brien, Director, Evaluation Office, UNICEF Headquarters</i>	2
<i>Ray Rist, President, International Development Evaluation Association (IDEAS)</i>	4
<i>Oumoul Khayri Ba Tall, President, International Organization for the Cooperation in Evaluation (IOCE)</i>	6

Editorial:

<i>Marco Segone, Senior Regional Advisor, Monitoring and Evaluation, UNICEF Regional Office for CEE/CIS, and former Vice President, International Organization for the Cooperation in Evaluation (IOCE)</i>	8
---	---

PREFACE BY DIRECTOR OF EVALUATION, UNICEF

It is a great pleasure, as Director of Evaluation at UNICEF, to write a preface for this timely publication. The issue of country-led monitoring and evaluation systems has been increasingly recognized as central to the promotion of development effectiveness. The Paris Declaration and the recent follow up in the Accra Agenda for Action, stress the importance of developing and working through country systems, and explicitly refer to national monitoring systems and country-led evaluations.

Within UNICEF, there has long been a recognition that our approaches to monitoring and evaluation have to reflect the nature of our involvement in the development process. The Country Programmes supported by UNICEF are country-led and nationally executed and therefore there will be an increasing emphasis on country-led evaluations and the strengthening of national monitoring and evaluation systems. In supporting countries to uphold and protect the rights of children and women and to achieve the Millennium Development Goals, we recognize the importance of using evidence to shape policy and practice, both internationally and in specific country contexts.

Unfortunately, we have to acknowledge that the reality is often far removed from the lofty ideals of international agreements. So much evaluation work, especially in developing countries, is still donor-driven and designed to meet the needs of outside agencies. The change that is needed is a paradigmatic one if monitoring and evaluation are truly to inform national policy making processes. It will require a change of attitude and behaviour as well as the building of capacity at many levels.

This publication fully recognizes the extent of the challenges ahead. The editor is to be congratulated on bringing together a diversity of perspectives and making an important contribution to the debate on country-led monitoring and evaluation systems and their ability to enhance evidence-based policy making.

Finbar O'Brien, Director

*Evaluation Office
UNICEF Headquarters*

PREFACE BY IDEAS PRESIDENT

It is a pleasure, as President of IDEAS, to write the preface for this book on strategies and approaches for enhancing evidence-based policy making through country-led monitoring and evaluation systems. At least one quarter of the papers presented here have been written by IDEAS members. This fact, yet again, is evidence of the intellectual vitality and focus of IDEAS members on the issues facing all of us working in development evaluation.

Enhancing evidence-based policy making, including through country-led monitoring and evaluation systems has, for some time, been a concern of development evaluators, donors, and government officials. It is good that this book takes us forward, in our thinking and understanding, on how to improve decision making through use of monitoring and evaluation systems, especially in developing countries. We now know much on how it should be done (and sometimes is done) in developed countries. But building the knowledge base on how it should be done in developing countries is still an area with significant gaps in understanding. I commend the editor for taking this inquiry forward.

Country-led monitoring and evaluation systems is an emergent topic with a knowledge base which is slowly growing. Developing country-led monitoring and evaluation systems takes time – just as it has in developed countries. There are, however, additional constraints on building such systems in developing countries. Learning how to cope with these constraints; how to create viable data in countries and locations where it previously did not exist; and, how to get relevant information to relevant decision makers in a relevant time frame, are all challenges that are only slowly being addressed. There are relevant case studies of developing countries where monitoring and evaluation systems are operational, providing good information to decision makers in real time. IDEAS has held several conferences on this topic and the paper here by two IDEAS colleagues, Adrien and Jobin, summarizes much of this work.

Again, the editor is to be congratulated on pulling this group of papers together. They are timely, topical, and to the point. This book also takes us further forward as it starts to forge the link between our learning about evidence-based policy making and the contributions that country-led monitoring and evaluation systems can play in supporting good decision making.

Ray C. Rist, President

*International Development
Evaluation Association*

PREFACE BY IOCE PRESIDENT

As a global evaluation organization, IOCE seeks to promote evaluation as an effective decision making tool that works in different contexts and cultures. IOCE is very much attached to the principles of cultural diversity, inclusiveness and cross fertilization of different evaluation traditions in ways that respect this diversity. It is therefore a great pleasure to welcome the book on “Country-led monitoring and evaluation systems. Better evidence, better policy, better development results”, as we share the same principle of ownership that lies under the concept of Country-led evaluations (CLE).

Whilst the evaluation community agrees on the inherent value and attractiveness of CLE, important challenges arise when it comes to the question of how to do CLE. CLE conveys principles in line with new development theory paradigms which value a bottom-up approach. It puts developing countries in the driver’s seat, and is therefore attractive. Along with capacity and institutional weaknesses, major constraints are the lack of a genuine evaluation demand, and a weak evaluation culture. When we analyze the trends in evaluation worldwide, it is no surprise to see that the traditional and current evaluation practices in the developing world are mainly top-down methodologies, introduced through models with different aid modalities. They are therefore designed and conducted to respond primarily to aid effectiveness. It is also no surprise to observe that evaluation thinking is evolving at a moment when development paradigms are changing priorities and introduce the principles of ownership and mutual accountability.

The CLE concept carries the hope that evaluation systems will be nationally owned. It builds on the Paris Declaration principles and clearly states the rules of the game. It pictures a reversal of the current status which is simply upside down, but there is still a long way to go to make it work effectively. An official in a developing country government commented recently that “ownership of development aid is necessary for the capacity building of the country”, whereas, in many agreements, capacity building is set to come first, usually as

conditionality or pre-requisite before the country's system can be used.

Evaluation networks play an important role in bringing together evaluation stakeholders, not only practitioners, but also commissioners and users, from the north and the south. They meet in networks to share, create and disseminate knowledge around key issues on development results. In this way they raise awareness and interest in the multiple uses of evaluation in development which are the first steps to build capacity.

I invite all networks to use the reflections contained in this book for that purpose, and to continue to enrich research and to advocate for more evaluations that respect the CLE principles.

Oumoul Khayri Ba Tall, President

*International Organization
for Cooperation in Evaluation*

EDITORIAL

This publication offers a number of strong contributions from senior officers in institutions dealing with national monitoring and evaluation systems, such as UNICEF, the World Bank, the UN Economic Commission for Europe, the Organisation for Economic Cooperation and Development (OECD), the International Development Evaluation Association (IDEAS) and the International Organisation for Cooperation in Evaluation (IOCE). It tries to bring together the vision, lessons learned and good practices from different stakeholders on how country-led monitoring and evaluation systems (CLES) can enhance evidence-based policy making.

Why Country-led monitoring and evaluation systems?

The international community agrees that monitoring and evaluation has a strategic role to play in informing policy making processes. The aim is to improve relevance, efficiency and effectiveness of policy reforms. Given this international community aim, why then is monitoring and evaluation not playing its role to its full potential? What are the factors, in addition to the evidence, influencing the policy making process and outcome? How can the uptake of evidence in policy making be increased?

This publication suggests that country-led monitoring and evaluation systems may enhance evidence-based policy making by ensuring national monitoring and evaluation systems are owned and led by the concerned countries. This would facilitate the availability of evidence relevant to country-specific data needs to monitor policy reforms and national development goals, whilst at the same time, ensuring technical rigour through monitoring and evaluation capacity development. However, effective country-led monitoring and evaluation systems will also have to address a second challenge: to bridge the gap between policy-makers (the users of evidence) and statisticians, evaluators and researchers (the providers of evidence).

Segone introduces the concept and dynamics of evidence-based policy making, underling that the main challenge is matching technical rigour with policy relevance. For policy-makers, good evidence has to be technically sound – that is, good quality and trustworthy evidence - as well as policy relevant – that is, addressing their policy questions. This is why country-led monitoring and evaluation sys-

tems may be the right strategy for national development decision making processes. Country-led evaluations (CLE) are evaluations in which the country which is directly concerned leads and owns the evaluation process by determining: what policy or programme will be evaluated; what evaluation questions will be asked; what methods will be used; what analytical approach will be undertaken; and, how the findings will be communicated and ultimately used. CLE serves the information needs of the country and, therefore, CLE is an agent of change and instrumental in supporting national development results. Finally, Segone assesses the challenges which remain in implementing country-led monitoring and evaluation systems despite the Paris Declaration principles of national ownership and leadership, and proposes a way forward.

Picciotto, acknowledging the increasing amount of evaluation of development activities at country level, explains why the shift in the unit of account, from individual operations to the higher plane of country assistance strategies, took place. In addition, he analyses what the new orientation implies for aid management and what challenges it creates for evaluation methods and practices. Finally, Picciotto assesses whether a country-based approach to development evaluation will remain relevant, given the spread of multi-country collaborative development programmes.

Quesnel explains how an understanding of the strategic intent is an essential prerequisite for any relevant and efficient country-led monitoring and evaluation system. The strategic intent makes explicit the aim of the developmental intervention being pursued and provides coherence to country efforts and external support. It fosters greater effectiveness of the scenario being implemented and facilitates the measurement of achievements. Academic literature tends to present the strategic intent using a monolithic view. Quesnel presents a generic definition and illustrate various applications of the strategic intent at different levels of management, using different results-based paradigms. He then concludes that country-based monitoring and evaluation systems need to start with an explicit enunciation of the strategic intent.

Lundgren and Kennedy describe some of the opportunities and challenges in promoting partner country leadership in development evaluation. In the context of the aid effectiveness agenda, the authors provide an overview of donor efforts to promote joint and partner-led evaluations; support evaluation capacity development; disseminate evaluation standards and resources; and, to better

align and harmonise aid evaluation. The article shares some lessons on the role of donors in supporting partner ownership of evaluation drawn from the experience of the DAC Evaluation Network members. Finally, several outstanding issues are raised, including: the challenge of balancing the evaluation needs of the donor, partner and beneficiary; the need to integrate aid evaluation into partner governance and management systems; and, the limitations posed by the lack of an enabling environment for evaluation in many contexts.

Feinstein analyses a country-led evaluation experience, presents a rationale and vision for country-led evaluations, and assesses opportunities, achievements and lessons learned. He explains why the experience so far with CLE has been mixed if not disappointing. Finally, he concludes by proposing a wider approach which shifts the focus from a specific type of evaluation to country-led evaluation systems which generate country-led evaluations as products.

Adrien and Jobin explore the relationship between Country-led evaluations and good governance, suggesting CLE directly impacts three component of good governance: voice, accountability, and the control of corruption. The authors analyze a specific type of CLE: country-led impact evaluations (CLIE), introducing a discussion on impact evaluation, and presenting the results of a survey on impact evaluation. Finally, they present the challenges ahead, based on the debate generated at the recent conference on “Evaluation under a managing-for-development results environment” organized by IDEAS and the Malaysian Evaluation Society.

Khayri Ba Tall analyses the role of national, regional and global evaluation organisations in strengthening country-led monitoring and evaluation systems. She gives an overview of the evaluation networks world-wide, and elaborates on the different functions of evaluation. Finally, Khayri Ba Tall proposes some strategies to strengthen country-led monitoring and evaluation systems, such as creating a domestic demand for evaluation; extending the evaluation object and scope beyond aid; and, improving the supply side through evaluation capacity development.

Giovannini identifies some key challenges for official statistics in terms of relevance, legitimacy and, therefore, their role in modern societies. He investigates how citizens see and evaluate official statistics and the role played by the media in this respect, using empirical evidence concerning several OECD countries. Giovannini argues that the value added of official statistics depends on

its capacity for creating knowledge in the whole society, not only among policy-makers. The development of a culture of “evidence-based decision-making”, together with the transfer of some decisions from the State to individuals and the growing opportunities created by globalisation, has stimulated an unprecedented increase in the demand, by individuals, for statistics. Some conclusions are drawn about the need to transform statistical offices from “information providers” to “knowledge builders” for the sake of democracy and good policy.

Baer argues how development of services, marketing and dissemination of statistical information are issues of strategic importance for any statistical institution. Understanding customers, marketing and building relationships are not just side functions or minor activities, they are closely linked with the reputation, future role and viability of statistical agencies. To develop better interaction with existing and new users it is vital to be proactive. Agencies must define potential user groups and describe their likely needs. The relative importance of each potential user group must be decided before developing a dissemination strategy. There is limited time and resources to provide services to all user groups and so prioritization will be necessary.

Good practices in Country-led monitoring and evaluation systems

Mackay examines the various ways in which monitoring and evaluation systems can, and are, used to improve government performance. He reviews key trends which are influencing developing countries in building or strengthening existing monitoring and evaluation systems. He also discusses the numerous lessons from international experience in building monitoring and evaluation systems, including the important role of incentives to conduct, and especially to make use of, monitoring and evaluation information. Mackay also presents ways to raise awareness of the usefulness of monitoring and evaluation creating incentives for its utilization and how such incentives can help to create demand for monitoring and evaluation. Finally, he examines the importance of conducting a country diagnosis, to provide a shared understanding of the strengths and weaknesses of existing monitoring and evaluation systems, and to foster a consensus around an action plan for its further strengthening.

Kusek and Rist present the importance of a strong theory of change. They explain how to successfully build a strong evaluation

culture in developing countries and the need for an emphasis on how evaluation can help deliver information and analysis that strengthen programme delivery. In short, how evaluation can provide coherent and useful theories of change which countries can deploy as they seek to address the problems they have. Kusek and Rist finally present the COREL approach, that is, five questions which need to be answered when thinking through the logic of a programme, or its theory of change.

Bamberger and Rugh explain how the RealWorld Evaluation (RWE) approach may assist the many evaluators, in developing, transition and developed countries, who must conduct evaluations within budget, time, data and political constraints. Determining the most appropriate evaluation design under these kinds of circumstances can be a complicated juggling act involving a trade-off between available resources and acceptable standards of evaluation practice. Often the client's concerns are more about budgets and deadlines, and basic principles of evaluation may receive a lower priority. Failure to reach satisfactory resolution of these trade-offs may also contribute to a much lamented problem: low use of evaluation results. RWE is a response to the all-too-real difficulties in the practical world of evaluation.

Segone, Sakvarelidze and Vadnais present the contribution of household surveys in general, and the Multiple Indicators Cluster Survey (MICS) in particular, in strengthening country-led monitoring and evaluation systems. The authors explain how MICS3 was instrumental in enhancing national statistical capacity and quality assurance systems, through national ownership and a technical assistance system. They also present good practices in data dissemination, as well as some examples of how MICS3 data have been used at national, regional and global level to inform evidence-based policy advocacy and to stimulate further analysis on specific topics, such as child poverty analysis.

Pron, Oswalt, Segone and Sakvarelidze argue that to achieve sustainable development outcomes, country-led development strategies must be backed by adequate financing within the global partnership for development. However, this is only possible if timely evidence is available from policy-relevant and technically-reliable country-led monitoring and evaluation systems. The evidence provided by such systems, owned by developing and transition countries, should inform necessary policies and strategies to ensure progress. The authors present how DevInfo – a user-friendly data

dissemination system which the UN offers to countries – was designed to facilitate ownership by national authorities and is being used by hundreds of countries world-wide – including more than half the countries in Eastern Europe and Central Asia - within national and decentralized monitoring and evaluation systems. Selected good practices from Belarus, Moldova, Kyrgyzstan, Serbia and Tajikistan – among others – are presented.

Last but not least, the **UNECE** article is a practical tool to help managers, statisticians and media-relation officers to use text, tables, graphics and other information to bring statistics to life using effective writing techniques.

I wish you an interesting and inspiring reading.

Marco Segone, Editor



Part 1

Why country-led monitoring and evaluation systems?

Enhancing evidence-based policy making through country-led monitoring and evaluation systems. <i>Marco Segone, Senior Regional Advisor, Monitoring and Evaluation, UNICEF Regional Office for CEE/CIS, and former IOCE Vice President</i>	17
Evaluating development. Is the country the right unit of account? <i>Robert Picciotto, Visiting Professor, King's College, London and former Director General, Evaluation, the World Bank</i>	32
The strategic intent. Understanding strategic intent is the key to successful country-led monitoring and evaluation systems. <i>Jean Serge Quesnel, Professor at the United Nations System Staff College, Adjunct Professor at Carleton University and Professeur Associé at the École Nationale d'Administration Publique of Quebec</i>	56
Supporting partner country ownership and capacity in development evaluation. The OECD DAC evaluation network. <i>Hans Lundgren, Head of Evaluation Section, Development Co-operation Directorate, OECD</i> <i>Megan Kennedy, Consultant, OECD</i>	77
Country-led evaluations. Learning from experience. <i>Oswaldo Feinstein, Professor at the Master in Evaluation, Complutense University, Madrid, and former Manager, Operations Evaluation Department, the World Bank</i>	96

Country-led impact evaluation. A survey of development practitioners
Marie-Hélène Adrien, President, Universalis and former President, IDEAS,
Denis Jobin, Vice President, IDEAS and Manager, Evaluation Unit, National Crime Prevention Center, Public Safety, Canada 102

The role of national, regional and international evaluation organizations in strengthening country-led monitoring and evaluation systems.
Oumoul Khayri Ba Tall, President, International Organization for Cooperation in Evaluation (IOCE) 119

Bringing statistics to citizens: a “must” to build democracy in the XXI century
Enrico Giovannini, Chief Statistician, OECD 135

Proactive is the magic word.
Petteri Baer, Regional Advisor, Statistical Division, UN Economic Commission for Europe 158

ENHANCING EVIDENCE-BASED POLICY-MAKING THROUGH COUNTRY-LED MONITORING AND EVALUATION SYSTEMS

Marco Segone,

*Senior Regional Advisor, Monitoring and Evaluation,
UNICEF Regional Office for CEE/CIS,
and former IOCE Vice President*

Introduction

The international community agrees that evidence is, and should be, instrumental in informing policy-making processes. The aim is to improve relevance, efficiency and effectiveness of policy reforms. Given this international community aim, why then is evidence not playing its role to its full potential? What are the factors, in addition to the evidence, influencing the policy-making process and outcome? How can the uptake of evidence in policy-making be increased? This paper is a preliminary attempt to give some answers to the above questions.

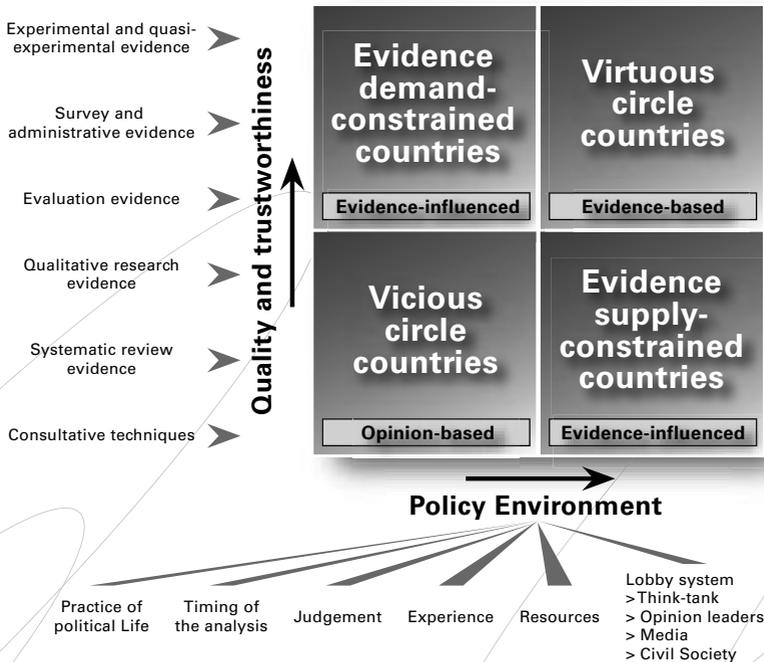
The dynamic of evidence-based policy-making

Evidence-based policy has been defined as an approach which “helps people make well informed decisions about policies, programmes and projects by putting the best available evidence at the heart of policy development and implementation” (Davies, 1999a). This definition matches that of the UN in the Millennium Development Goals (MDG) guide. Here it is stated that “Evidence-based policy-making refers to a policy process that helps planners make better-informed decisions by putting the best available evidence at the centre of the policy process”.

This approach stands in contrast to opinion-based policy, which relies heavily on either the selective use of evidence (e.g. on single survey irrespective of quality) or on the untested views of individuals or groups, often inspired by ideological standpoints, prejudices, or speculative conjecture.

Many governments and organizations are moving from “opinion-based policy” towards “evidence-based policy”, and are in the stage of “evidence-influenced policy”. This is mainly due to the nature of the policy environment as well as national technical capacity to provide good quality and trustworthy evidence. The policy environment may vary from a closed and corrupted society to an open, accountable and transparent one. Political and social systems influence use of evidence. Issues such as the timing of evidence and availability of resources; values, beliefs and ideology affect its use. Personal experience and expertise also influence the judgment of policy-makers. In addition, the lobby system existing in the country, including think-tanks, opinion leaders, non-governmental organizations and mass media have an impact.

Figure 1: Dynamic of policy-making



Public policies are developed and delivered through the use of power. In many countries, this power is ultimately the coercive power of the state in the hands of democratically accountable politicians. For politicians, with their advisers and their agents, securing and retaining power is a necessary condition for the achievement of their policy objectives. There sometimes seems to be a tension between power and knowledge in the shaping of policy. A similar tension appears to exist between authority and expertise in the world of practice. Emphasizing the role of power and authority at the expense of knowledge and expertise in public affairs seems cynical; emphasizing the latter at the expense of the former seems naïve.

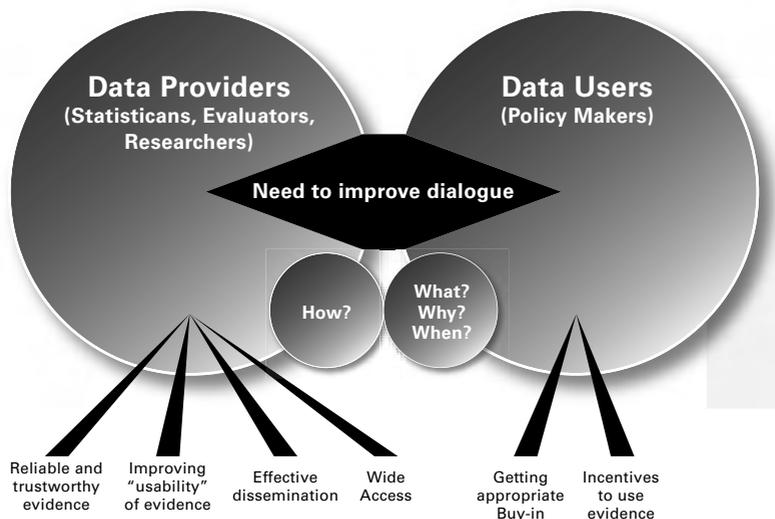
Power and authority versus knowledge and evidence, maybe more complementary than conflicting. This interdependence of power and knowledge is perhaps more apparent if public policy and practice is conceived as a continuous discourse. As politicians know too well, but social scientists too often forget, public policy is made of language. Whether in written or oral form, argumentation is central in all stages of the policy process. In this context, evidence is an important tool for those engaged in the discourse, and must be both broad enough to develop a wide range of policy options, and detailed enough for those options to stand up to intense scrutiny.

Matching technical rigor to policy relevance

For policy-makers, good evidence is technically sound. That is, good quality and trustworthy evidence which is policy relevant and addresses their policy questions. If evidence that is technically sound is not policy relevant, then it will not be used by policy-makers. The opposite also applies, that is, policy-makers may be forced to use poor quality evidence, if this is the only evidence available that address their policy questions.

A stronger commitment to make evidence not just useful but useable, and increasing the uptake of evidence in both policy and practice, has become a preoccupation for both policy people and service delivery organizations.

Figure 2: Good quality evidence. Matching technical rigour to policy relevance



The need to improve the dialogue between policy-makers and evidence providers

Getting policy-makers and practitioners to own the evidence needed for effective support and implementation of policy is an important strategy. This is in contrast to the position where evidence is solely the property and domain of evaluators, statisticians and researchers, or, perhaps even worse, managers and bureaucrats who try to impose less than transparent evidence upon practitioners and front line staff. Ownership of the best available evidence can enhance its use to make well informed and substantiated decisions.

To improve ownership and uptake of evidence, in both policy and practice, developing better ongoing interaction between evidence providers and evidence users is the way forward. Much of the more recent thinking in this area now emphasizes the need for dialogue if common ground is to be found. This is strategic because, at the end of the day, policy-makers know what evidence they need, why they need it, and when they need it. Statisticians, evaluators and researchers know how to provide that evidence.

The advantages for an enhance dialogue are clear. However, the professional autonomy of statisticians, evaluators and researchers needs to be maintained to ensure the trustworthiness of evidence produced, and therefore its use by policy-makers as well as the

public. Therefore, getting the right balance between both the principles of professional autonomy and accountability, and the relevance of evidence produced, is paramount.

Matching demand with supply of appropriate evidence

A distinction can be made between people who are users of evidence and those who are providers of evidence. Whilst it may be unrealistic for professional decision-makers and practitioners to be competent doers of statistics and evaluations, it is both reasonable and necessary for such people to be able to understand and use statistics and evaluations in their professional practice. Integrating evidence into practice is a central feature of professions. An increasingly necessary skill for professional policy-makers and practitioners 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. Joint training and professional development opportunities for policy-makers and analysts may be one way of taking this forward and for matching strong demand with a good supply of appropriate evidence.

Making evidence “usable” for the policy-making community

A further challenge for statisticians and evaluators is making data and information “usable” for the policy-making community. Statisticians often need to ‘translate’ statistics into a language that is useful to the users of evidence, without distorting or mis-representing data.

Effective dissemination and wide access

A key issue is how to communicate findings to those who need to know. The strategies used to get evidence to their point of use involve both dissemination (pushing information from the centre outwards), and provision of access (web based and other repositories of information which data users can tap into). DevInfo, the UN common platform to monitor MDGs, has proven to be an effective tool in this regard.

Incentives to use evidence

Policy-makers may need incentives to use evidence and to do what has been shown to be effective. 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 use of evidence, such as integrating analytical staff at all stages of the policy development process.

Civil society organizations may also advocate the use of evidence in policy-making. Think-tanks, with the support of mass media, may also make evidence available to citizens, and citizens may demand that policy-makers use it.

Evidence-based policy-making in different country settings

Developing and transition countries vary greatly in the quantity and quality of information available to policy-makers, and in the extent to which this information is used. Paris 21, a partnership for strengthening statistics led, by the Organization for Economic Cooperation and Development (OECD), distinguishes four types of country (as in figure 1). These are:

- **Vicious circle countries.** Evidence is weak and policy-makers make little use of it. Evidence-based policy-making is not practiced, which results in poor policy decisions and poor development outcomes. 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 weak, it is increasingly used by policy-makers. However, evidence deficiency reduces the quality of decision-making which results in poor development outcomes. Policy-makers are likely to resent being held to account on the basis of inadequate evidence. The priority is to adopt measures to increase the quantity and quality of evidence, which will require additional technical assistance for capacity development, as well as to improve the dialogue between producers and users of data. The challenge is to strike a balance between generating improvements to evidence quickly, while laying the foundations for better performance of the national monitoring and evaluation system in the long-run. What should be avoided are actions which offer short-run benefits, but generate long-run costs.
- **Evidence demand-constrained countries.** The quantity and quality of evidence is improving, but it is not used for decision-making because policy-makers lack the incentives and/or the

capacity to utilize it. This results in poor policy design and poor development outcomes. Policy-makers are likely to be at the very least wary of (or may even actively dislike) having more and better figures pushed at them when these data may not support decisions they have 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 improving and is being increasingly used 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 mutually reinforce each other, resulting in better policy design and better development outcomes.

This situation of virtuous circle countries serves more as a goal to be achieved, even in some developed nations, than as a description of events in a particular group of countries. Nevertheless, 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. Strengthening the democratic process by requiring transparency and accountability in public sector decision-making, together with the establishment of clear accounting standards and an effective regulatory framework for the private sector, are essential elements for sustaining a virtuous circle linking statisticians, evaluators and researchers to policy-makers.

Country-led monitoring and evaluation systems. Better evidence, better policies, better development results.

As acknowledged by the 37th Development Assistance Committee (DAC) working group on aid evaluation, the fact that most evaluations of development aid have been led by donors and were done to satisfy donors' requirements had at least two significant consequences. These are lack of country ownership of these evaluations and, therefore, under utilization of evaluation findings and recommendations and, a proliferation of donor evaluations leading to high transaction costs for the countries. In addition, the primary purpose

of donor-led evaluations is to ensure donor accountability and learning, and not to address the information needs of national and local decision makers and governance systems.

To address the above situation, a number of joint-evaluations by donor and partner countries have been carried out since early the 1990s. However, many of them were led by donors, and the role of partner countries tended to be confined to supporting data collection and commenting to evaluation findings drafted by donors.

It is therefore clear that simply tweaking the existing donor-led monitoring and evaluation systems is not enough. A new approach to country-led monitoring and evaluation systems is needed. The shift called for is not only a technical one, but a socio-organizational one.

Country-led monitoring and evaluation systems

At the 2008 virtual international workshop held by IDEAS on country-led evaluation (CLE), which I had the honor to facilitate, CLE was defined as evaluation which the partner country (and not the donors) leads and owns by determining:

- what 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,
- how the findings will ultimately be used.

CLE serves the information needs of the country and, therefore, CLE is an agent of change and is instrumental in supporting national development results. This is possible because it builds on the culture and values of the country. If values and beliefs of one exogenous society are imposed on another through evaluation, we have a situation that is likely to lead to error, resentment and misunderstanding.

It should be noted that, while governments have a key role to play in CLE, civil society could be actively involved by evaluating the performance of public services – and thus allowing them to articulate their voice. In this context, professional evaluation organizations have a potentially significant role to play. This is especially so given the dramatic increase in the number of national and regional professional evaluation organizations. In the last 10 years, the number

grew from half a dozen in 1997 to more than 70 in 2008, with most of the new organizations located outside Western Europe and North America¹. Moreover, two global organizations have been created. These are the International Organization for Cooperation in Evaluation (IOCE), the world federation of regional and national evaluation organizations, and the International Development Evaluation Association (IDEAS), a world association of individual evaluators.

The Joint Country-led evaluation in Bosnia and Herzegovina

Within the cooperation with UNICEF, the Directorate for Economic Planning (DEP) of the Council of Ministers of Bosnia and Herzegovina (BiH) attended the IDEAS's regional workshop on Country-led evaluation held in Prague. As outcome, it was decided to carry out a joint country-led evaluation (CLE) of the child-focused policies within the social protection sector.

The scope of the joint CLE was multi-faceted. Rather than evaluating the effectiveness, relevance, efficiency, sustainability and impact of one specific policy area, the decision was made to combine an assessment of child and family-focused policies as defined in the Mid Term Development Strategy (MTDS), with an evaluation of the effectiveness of the UNICEF contribution to child-focused policies. This dual approach allowed for an evaluation of governmental and UNICEF interventions both individually and, more importantly, the interaction between them. Further objectives related to the implementation of Paris Declaration targets by national stakeholders and donors, as well as documenting the methodology used in the joint CLE for its further application in BiH.

The joint CLE provided a strategic opportunity for DEP to demonstrate increased leadership in the field of monitoring and evaluation of national development strategies. The DEP's leadership in the CLE was strategic as that same year, 2007, they began the process of preparing a new MTDS, the Social Inclusion Strategy and the National Development Plan. DEP ability to apply the lessons learned in the joint CLE process proved to be particularly valuable.

In addition, the joint CLE further strengthened the existing partnership between UNICEF and DEP in the area of strengthening national monitoring and evaluation capacities.

Source: 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: Segone, M, Bridging the gap. The role of monitoring and evaluation in evidence-based policy making. UNICEF

1 See Segone, M. and Ocampo, A. (2006), IOCE (International Organization for Cooperation in Evaluation). *Creating and Developing Evaluation Organizations. Lessons learned from Africa, Americas, Asia, Australasia and Europe, Peru.*

National ownership and capacity development: the key ingredients of country-led monitoring and evaluation systems

As mentioned above, national ownership is the best strategy to ensure policy relevance, and therefore use of evidence, while national capacity development is needed to enhance the technical rigour of evidence.

The Paris Declaration on aid effectiveness was endorsed in 2005 by more than one hundred ministers, heads of agencies, and other senior officials from a wide range of countries and international organizations. It lays out five principles to improve the quality of aid and its impact on development: ownership; alignment; harmonization; managing for results; and, mutual accountability. The explicit commitment to ownership was an addition in Paris to the previous aid effectiveness agenda, and it was intentionally placed first on the list. The prominence of ownership reflects the understanding that national ownership and leadership is the most important overarching factor for ensuring good development outcomes.

The ownership principle in the Paris Declaration states that partner (developing and transition) countries will exercise effective leadership over their development policies and strategies and co-ordinate development efforts themselves. Donors are responsible for supporting and enabling partner countries' ownership by respecting their policies and helping strengthen their capacity to implement them.

The implication for the monitoring and evaluation function is fundamental. The principle of ownership means that partner countries should own and lead their own country-led national monitoring and evaluation systems, while donors and international organizations should support sustainable national monitoring and evaluation capacity development. Donors and international organizations should also take into consideration the value of diversity in evaluation approaches and help to ensure the information and data produced are in compliance with monitoring and evaluation standards.

Challenges facing country-led monitoring and evaluation systems

The Central and Eastern Europe regional workshop on CLE held in Prague² acknowledged that experience so far has been mixed,

2 The workshop was organized by IDEAS in cooperation with Development Worldwide, Institute of international relationship and UNICEF.

due to a range of issues (IDEAS, 2006). "One element is that the drive towards ownership is partly supply-driven. A second element may be the perceived risk, on the side of partner countries, that independent evaluations of donor support may have political and financial consequences. A heavy aid dependency could translate into a reluctance to evaluate the role of donors independently. A third element may be the time frame. Starting up a process towards a country-led evaluation may require much more time than expected because of the necessary internal negotiations among different stakeholders, such as different ministries, civil society and evaluators. Last but not least, a fourth element is the perceived risk by donors of weak national capacities and, in some cases, of weak independence of national monitoring and evaluation systems".

This perceived risk is confirmed by the 2008 Evaluation of the implementation of the Paris declaration, which found that strengthening capacity and trust in country systems is a major issues. The evaluation revealed that the real and perceived risks and relative weakness of country systems are serious obstacles to progress on alignment. Efforts by most countries to strengthen national systems are not yet sufficient and not enough donors are ready to help strengthen these systems by actually using them. This limits the capacities of partner countries to exercise leadership.

The 2008 UNDG evaluation of the implementation of the Paris declaration also found that donors continue to rely on their own monitoring and evaluation systems due to weak and fragmented country systems, despite commitments to support countries in strengthening their systems. Helping build national statistical capacities is seen as a key requirement. Almost all donors seem to be engaged in some sort of capacity development assistance that should strengthen managing for results. This assistance can be support to development of statistics, help in developing results frameworks, or the introduction of a "results culture". However, these efforts appear piecemeal and are often tied to the specific needs or areas of intervention of donors.

This situation was confirmed by the joint country-led evaluation carried out by the Government of Bosnia and Herzegovina and UNICEF in 2007. The evaluation found that donors often have difficulties in addressing weak capacities and governance issues within their partnership approaches and they tend to take an over dominant role. As a result, national stakeholders have only a limited sense of ownership of donor-funded programmes and the resulting policy

changes. In turn, donors face difficulties in implementing partnership approaches with multiple levels of government.

The way forward

Despite the challenges above, important efforts have been made, and lessons learned during the first generation of country-led monitoring and evaluation systems:

- **Cooperation among middle income, transition and developing countries**

Middle Income Countries are successfully implementing national monitoring and evaluation systems. The ECOSOC Development Cooperation Forum recommended, in 2008, that south-south cooperation should be strengthened to enhance national capacities, as many emerging eastern and southern countries have a great deal of experience that can be better utilized.

- **National evaluation organizations fostering endogenous demand and supply for monitoring and evaluation**

National evaluation organizations are potentially important players in creating and strengthening national demand for monitoring and evaluation by, for example, setting culturally-sensitive evaluation standards³, enhancing quality implementation, and providing a national forum for greater dialogue on evaluation among civil society, academia, governments and donors. A clear example is the Niger Monitoring and Evaluation Network (ReNSE), which led to the organization of the 2008 African Evaluation Association in Niamey and contributed to the creation of the Government's Monitoring and evaluation department.

IOCE, IDEAS, the Regional evaluation associations in Africa, the Commonwealth of Independent States and Latin America, as well as international development organizations such as the UN, have an important role to play in supporting national evaluation organizations, as described in the book "Creating and developing evaluation organizations".⁴

3 For example, a presenter from China at the 2006 European Evaluation Society stated that his country is exploring the possibility to include two "national" evaluation standards, to measure the extent to which the policy/programme evaluated a) fostered Equity among stakeholders and b) enhanced Innovation.

4 See Segone, M. and Ocampo, A. (2006), IOCE (International Organization for Cooperation in Evaluation), *Creating and Developing Evaluation Organizations. Lessons learned from Africa, Americas, Asia, Australasia and Europe, Peru.*

- **International organizations strengthening national capacities to design and implement national monitoring and evaluation systems**

The Paris Declaration's principles of managing for results, mutual accountability, alignment and ownership are developing an enabling environment. Partner countries and International organizations should therefore take advantage of this historical momentum.

While Partner countries should drive and own the process, international organizations should support them by developing national capacities and facilitate the sharing of international good practices. This book, as well as the previous one on the role of monitoring and evaluation in evidence-based policy-making published in 2008⁵, is an initial small step in this direction, presenting methodologies and good practices on how to strengthen national monitoring and evaluation systems.

References

- Campbell, S., Bettina, S., Coates, E., Davis, P. and Penn, G. (2007), *Analysis for Policy: evidence-based policy in practice*. GSR (Government Social Research Unit) UK.
- Davis, H. T. O., Nutley, S. M. and Smith, P. C. (Eds.), (2000), *What works? Evidence-based Policy and Practice in Public Services*, UK.
- Davies, P. T. (2003), *Systematic Reviews: How Are They Different From What We Already Do?*
- Defra (Department for Environment, Food and Rural Affairs) UK. *Science. Evidence Based Policy-making*. Available at: < <http://www.defra.gov.uk/science/how/evidence.htm> >
- ECOSOC Development Cooperation Forum. (1 July 2008), *Note for the Record. Round Table Discussion on National Capacities to receive Aid*.
- ECOSOC Development Cooperation Forum. (30 June 2008), *Note for the Record. Round-table Discussion on South-South and Triangular Development*.
- Feinstein, Osvaldo. (2006), *Country-led Evaluation: Learning from Experience. Presented at the IDEAS Workshop on Country-Led Evaluation in Prague, June 2006*.
- 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*.
- IDEAS. (2006), *Country-led Evaluations and Systems: Practical experiences of the Central and Eastern European Region. Regional Workshop, June 19-20, 2006. Prague*.

5 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.

- Majone, G. (1989), *Evidence, argument and persuasion in the policy process*, USA.
- NESF (National Economic and Social Forum) Ir-eland., *Evidence-based Policy-making: Getting the Evidence, Using the Evidence and Evaluating the Outcomes*.
- 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*.
- Nutley, S. M., Davies, H. T. O. and Walter, I. (2003), *From Knowing to Doing: A Framework for Understanding the Evidence-into-Practice Agenda*. In: Evaluation 2003, (9, 2, pp.125-148). Available at: < <http://www.stand.ac.uk/~cppm.htm> >
- ODI (Overseas Development Institute) Civil Society Partnerships Programme., *Evidence-based Policymaking: Lessons from the UK for Developing countries*.
- OECD PARIS 21 Secretariat, (2007). *Advocacy for Statistical Capacity Building and Evidence-based Policy-making*.
- Paris Declaration Secretariat, (2008). *Evaluation of the implementation of the Paris Declaration*. France.
- Pawson, R. (2001), *Evidence based policy: I. In: search of a method*, UK, pp.22. Available at: < <http://www.evidencenetwork.org/Documents/wp3.pdf> >
- Pawson, R. (2001), *Evidence based policy: II. The promise of 'realist synthesis'*, UK, pp.20. Available at: < <http://www.evidencenetwork.org/Documents/wp4.pdf> >
- Perri. (2002), *Can policy-making be evidence-based?* In: MCC: Building Knowledge for Integrated Care Feb 10(1), UK, pp.3-8. Available at: < http://www.elsc.org.uk/bases_floor/managecc/feb2002/feb2002.htm >
- 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., 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 (International Organization for Cooperation in Evaluation), *Creating and Developing Evaluation Organizations. Lessons learned from Africa, Americas, Asia, Australasia and Europe, Peru*.
- Segone, M. (2006), UNICEF Regional Office for CEE/CIS and IPEN. *New Trends in Development Evaluation*, Switzerland.

Solesbury, W. (2001), *Evidence Based Policy: Whence it Came and Where it's Going* (Pre-publication version) Submitted to Planning Policy and Practice, ESRC Centre for Evidence Based Policy and Practice, UK.

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

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. Available at: < <http://evi.sagepub.com/> >

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*.

EVALUATING DEVELOPMENT. IS THE COUNTRY THE RIGHT UNIT OF ACCOUNT?

*Robert Picciotto, Visiting Professor, King's College, London
and former Director General, Evaluation, the World Bank*

Increasingly, evaluation of development activities is taking place at the country level. What explains the shift in the unit of account from individual operations to the higher plane of country assistance strategies? What does the new orientation imply for aid management? What challenges does it create for evaluation methods and practices? Will a country based approach to development evaluation remain relevant given the spread of multi-country collaborative development programs?

The origins

Arguably, economic aid has always been country focused. The 'development' idea that grew out of the ashes of World War II was deliberately targeted towards national goals when the victorious allies turned swords into ploughshares. Thus, the Marshall Plan aimed at restoring European countries shattered by conflict. Thereafter aid was explicitly aimed at nation building in the zones of turmoil created by the breakup of European colonies.

In particular, the historic contest between the western countries and the Soviet Union helped to generate resources for aid programs designed to influence the development trajectories of individual developing countries. Competing ideologies were tacitly embedded in aid operations that sought to demonstrate to the leaders of the newly independent countries that progress and modernization would best be achieved through adoption of donor countries' economic and social doctrines.

To be sure, altruism also played a role in development assistance and the discipline of evaluation that came into being at about the same time helped to moderate the ideological excesses of the cold war. This is because, in development as in other public policy areas, the evaluation pioneers intended their nascent craft to act as a transmission belt from the social sciences to public affairs. Indeed, the new evaluation profession was conceived as a source of contin-

gent, fallible and corrigible knowledge that would help bridge the gap between theory and practice.

In particular, Donald T. Campbell's conception of the 'experimenting society' raised expectations about the utility of evaluation for sound policy making. This was a time of heady optimism about the capacity of the social sciences to provide relevant knowledge for the conduct of public policy – whether directed to the reconstruction of war devastated nations, the promotion of prosperity in poverty-stricken regions or the creation of a peaceful global order through international collaboration. Towards these ends, the new development assistance business was conceived a multidisciplinary venture and evaluation acted as a connecting thread among the disciplines.

At country level, planners and economists constructed models designed to guide public investment decisions. On the ground, public administration specialists busied themselves with nation building tasks and financial analysts, economists, engineers, agronomists and other professions worked together to design projects for external financing. The project cycle explicitly included evaluation as part of a learning cycle and in 1970 Robert S. McNamara set the stage for the advent of the development evaluation profession when he instructed the 'whiz kids' of the World Bank's Programming and Budgeting Department to evaluate the *'contribution of the Bank's operations to the development of member countries'*.

A period of intensive experimentation began that drew on the lessons drawn by an evaluation system that gradually matured to address the multiple challenges associated with the development assistance profession (Willoughby, 2003). By then, the intellectual innocence of the pioneering years had dissipated and the public demanded accountability for the performance of aid projects. Accordingly, the evaluation function was entrusted with two distinct mandates – performance auditing and organizational learning.

The same mandate holds today but by now as the rest of this chapter will show development evaluation has expanded its range and its scope beyond its initial focus on discrete investment projects. It now addresses policies and institutions at the national level – and beyond. A vast literature dedicated to the effectiveness of aid has emerged and development evaluation has reached out to the other public policy disciplines.

The rise and decline of cost-benefit analysis

As the policy environment changes so do evaluation concepts and methods. The advent of the project as the main unit of account for development assistance and its subsequent demise parallel the rise and fall of the production function as the preferred metaphor of economic policy makers. In the pioneering years of the development business, input-output tables drove resource allocation decisions. Projects, privileged particles of development, were conceived as convenient vehicles for donor engagement with poor countries as well as building blocks for the design of five year plans by aid recipients.

In both of these contexts, cost-benefit analysis emerged as an indispensable tool of investment programming and project screening. The methodology was endorsed by academia since it was grounded in public finance theory and the 'new welfare' economics. The use of discounted cash flow techniques was novel, seductive and well adapted to the mindsets of planners and aid managers. Numerous operational instructions and training manuals were issued by international organizations, aid agencies and planning ministries to help planners and aid givers in the allocation of scarce national resources.

The new approach to investment planning and project evaluation rested on three pillars: (i) cash flow comparisons of costs and benefits attributable to the project in comparison to the counterfactual (the differentials between the 'with and without investment' scenarios); (ii) opportunity costs for production factors (product prices, labor, foreign exchange, capital, etc.) estimated with reference to national parameter and international markets; and (iii) variable weights applicable to project costs and benefits to take account of social welfare considerations and income distribution impacts.

Remarkably, the economic evaluation techniques used at project level were congruent with those used to estimate gross national products at the macroeconomic level: they were designed to measure the net returns that project investments yielded for the national economy. At the macro level, capital output ratios were plugged into dynamic input-output models to ascertain the effectiveness of public investment programs. Heroic efforts were made to disseminate the technique, train staff, generate data and estimate national parameters and shadow prices.

Needless to say, there was controversy about the practical value of these new fangled techniques. Esoteric methodologies were

proposed (but rarely adopted in practice) to take account of social vulnerability considerations and probe intergenerational effects. Fulsome debates took place about the reliability of the approach, its burdensome information and analytical requirements, the misleading precision of point estimates and the risks associated with the centralized decision making protocols that the method implicitly favored (given the need for consistency in methods, estimates of reference prices and quality assurance).

Nevertheless, an irresistible intellectual momentum swept all objections aside. At the country level, cost benefit analysis provided a logical and convenient intellectual construct that provided technocrats with a ready made management tool for public expenditures and aid programs. At the project level, the very same technique was used for identification, preparation, appraisals and *ex post* evaluations. At both levels the goal was to enhance the impact of public investment on economic growth and social cost benefit analysis provided a consistent analytical scheme that brought together all the relevant disciplines.

Thus, technical specialists provided the input-output coefficients needed to operate the models, financial analysts ensured that risk sharing was appropriate to the resources and responsibilities of participants, macro-economists estimated the shadow prices used to value factors of production and project outputs, while sociologists were consulted in ascribing different weights to project benefits flowing to the rich and the poor and trade offs between income growth and distribution were quantified to facilitate political decision making.

For more than two decades, the technique served as an emblem of rationality and professionalism even though it failed to capture the immense complexity of economic progress and social change. Given the intellectual credentials of the approach and the relative ease of its introduction within the bureaucracy, its decline cannot be explained simply by technical limitations. The objections raised with respect to the impact of uncertainty on the reliability of estimates, the prohibitive costs associated with their systematic use, the poor quality of the underlying data, the lack of comparability of estimates across sectors and the inherent difficulties of quantifying the counterfactuals ('without project' scenarios) provided ample fodder for academic speculations and methodological refinements.

Thus, while major drawbacks were acknowledged and much effort went into mitigating them, the staying power of the approach ultimately rested on the iconic status of the cost-benefit doctrine. Its

symbolic function helped to sustain its popularity even as development practice evolved and the number of operations that could meaningfully be justified through discounted cash flow estimates gradually shrank to a third of those financed by aid. It took a revolution in development thinking to shatter the exalted status of cost benefit analysis in the methodological pantheon of policy makers.

Following the debt crisis, a market fundamentalist wave engulfed the development industry and a gigantic macroeconomic experiment was launched to connect all developing countries to the mighty engine of the global economy. The shift in the unit of account from the project level to the country level occurred in the early 1980's when development policy doctrines evolved from the micro-economics of project appraisal to the macro-economics of the Washington consensus. From that time onwards, some cost benefit calculations would be carried out at the project level but they no longer had much influence in decision making.

The sudden decline of cost benefit analysis was connected to the disillusionment with state-led approaches to development and the shift in policy research priorities towards macro-policy reform. Once the neo-liberal economists captured the commanding heights of development assistance, the basic analytical instrument that the discipline of economics had provided to development evaluation became obsolete.

The aid enterprise having shifted its focus from the plan to the market, the limitations of economic modeling and the technical drawbacks of cost benefit methods (e.g. with respect to projects that dealt with policy reform or institutional development) were suddenly highlighted as fatal flaws. Policy blueprints reflecting the tenets of the Washington consensus replaced project cash flows. New aid vehicles were introduced and conditionality became focused on aligning prices to the market (thus making shadow prices redundant).

Paradoxically, much progress had been made by then in refining the technique and improving access to data. But nothing could stop the juggernaut of the policy adjustment craze that provided aid donors with leverage over major economic management decisions at country level. With the triumph of neo-liberalism, the project instrument that had been ideally suited to multi-disciplinary work fell into disfavor, macro policy conditionality came to the fore and the role of public investment in development was downgraded. Macro economics displaced micro-economics and country level results became the preferred tests of development performance.

Paradoxically, evaluation at project level contributed to the change in paradigm by highlighting the failure of a significant share of investment operations to meet their relevant objectives efficiently and by stressing the critical role of a good policy environment for effective development performance. Conversely, the paradigm shift exerted a powerful impact on evaluation methods. Cost benefit analysis lost its intellectual allure and innovations in evaluation moved to the higher plane of country level and sector wide policy assessments.

The overhaul of the development assistance tool kit, the emphasis on quick disbursing, policy based loans and grants (conditional on changes in policy and the reconsideration of evaluation methods) reflected the lessons of experience as well as the findings of public choice theories that highlighted the failures of government and elevated the prestige of market based solutions. Accordingly, the development evaluation profession began to retool itself to provide objective retrospective assessments of adjustment loans and country assistance strategies.

The combination of financial resources, advisory services and partnership arrangements that made up country assistance strategies became the main focus of development evaluation. In parallel, the discourse of development economics shifted from a predilection with planning to a preoccupation with economic policy and from an assessment of centrally planned public investments to a decentralized approach to economic management emphasizing market friendly policy frameworks and private sector led development.

The retreat of market fundamentalism

By the early 1990's the hubris associated with policy adjustment generated a backlash. The civil society put the spotlight on the intrusive, misguided and counterproductive conditions that had been imposed on some poor countries. This helped to reorient the development agenda: the market based approaches of the prior era were not altogether abandoned but they were made part and parcel of a comprehensive approach that gave equal weight to environmental and social development concerns.

Once again, the ground for the new policy shift was prepared by evaluation studies that exposed the excesses of coercive interference in economic management by aid donors, the social costs of adjustment and the limits of policy change without prior institutional reform. Faced by disappointing development trends, market funda-

mentalism retreated and lessons of experience were used to craft operational principles better adapted to the complex and multi-faceted challenges of the development enterprise.

Specifically, externally imposed conditions over reluctant governments were moderated and development assistance conditionality became less burdensome. Ex-ante policy sticks were replaced by ex-post policy carrots. At the same time development assistance vehicles were reshaped to spawn innovation and greater adaptability to volatile and risky operating conditions.

Eventually, poverty reduction became the overarching goal of development aid. By the mid-nineties, the stage was set for the translation of a new set of principles for effective aid into operational practices¹. A comprehensive development paradigm² took hold. It combined results orientation, domestic ownership of improved policies, partnerships between governments, the private sector and the civil society and a long term holistic approach that recognized explicitly the interaction between development sectors and themes.

The advent of this new consensus was formally consecrated by the endorsement of Millennium Development Goals by developed and developing countries' governments at the turn of the century. Specifically, a universal compact was forged at the United Nations Conference on Financing for Development held in Monterrey (Mexico) in March 2002. It was agreed that poor countries would take primary responsibility for governance reforms and poverty reduction programs and rich countries would provide them with more and better aid, more generous debt reduction and improved access to global markets.

Once again, evaluation had contributed to the re-orientation in thinking that had laid the groundwork for the policy transformation (Nagy, 1999). It did so by providing new evidence for policy making and crafting development effectiveness concepts that facilitated the shift to a new development consensus. Conversely, once the shift occurred, evaluation had to adapt its methods and practices to a more demanding set of requirements and the new consensus raised the importance of country program evaluations geared to the achievement of global development objectives.

- 1 By the late nineties, the new principles had been mainstreamed into general practice through the preparation of *Poverty Reduction Strategy Papers* by low income country governments as a standard requirement of aid and debt reduction programs.
- 2 A paradigm arises when a professional community adopts new beliefs about reality and subscribes to common symbolic generalizations about its expert discipline.

First, the traditional 'results chain' (linking inputs, outputs, outcomes, and impacts) had to be re-shaped to capture program results so that they conform more closely to the indicators associated with the Millennium Development Goals. Second, country program evaluations had to be connected to the objectives and modalities of Poverty Reduction Strategy Papers. Third, development outcomes were attributed to the joint contributions of governments, the civil society, the private sector and external development agencies, i.e. to partnerships geared to the achievement of shared objectives taking account of the distinctive accountabilities and reciprocal obligations of partners in performance assessments.

Shifting involvement of evaluation disciplines and methods

Engineering was dominant during the reconstruction phase of the 1950's, project finance came into its own during the pioneering days of the sixties; micro economics and sector expertise dominated the heyday of development during the seventies when planners and project economists held sway. The baton passed to macro economists in the eighties and to operational 'integrators' in the nineties. In the first decade of the new millennium no single discipline seems to be in charge since only a holistic approach can tackle the global issues that have risen to the top of the development agenda. Thus, from decade to decade, changes in development paradigm induced shifts in the pecking order of the social science disciplines used by development evaluation (Box 1).

Box 1: The impact of the development agenda on evaluation and the disciplines

Decade	Main objective	Main instrument	Main discipline
1950's	Reconstruction	Technical assistance	Engineering
1960's	Growth	Projects	Finance
1970's	Basic needs	Sector investment	Micro-economics
1980's	Adjustment	Policy based loans	Macro-economics
1990's	Capacity building	Country assistance strategies	'Operational integrators'
2000's	Human security	Global policy coherence	Multi-disciplinary

In turn, changing shifts in the development agenda and the discipline mix had a deep impact on evaluation methods and processes. For example, once the unit of account shifted from the project to country programs and policies, development evaluators had to invent new techniques and broaden their concentration on individual projects to the higher plane of policies and institutions (Box 2).

Box 2: New disciplines in evaluation respond to a changed policy context

	Before	After
Context dependent concepts	<ul style="list-style-type: none"> • Project focus • Investment driven growth • Import substitution • Central planning 	<ul style="list-style-type: none"> • Country focus • Policy reform • Outward oriented policies • Decentralized decision making
Evaluation disciplines	<ul style="list-style-type: none"> • Project evaluation • Cost-benefit analysis • Shadow pricing • Self-evaluation 	<ul style="list-style-type: none"> • Portfolio evaluation • Policy evaluation • Risk analysis • Participatory evaluation

To be sure, project evaluations were not abandoned and the micro-economic disciplines used to assess projects as free standing investments were not jettisoned. They were simply reoriented to address sector policy issues. In parallel, project evaluation procedures were reshaped to generate ‘building blocks’ for the evaluation of sector based and country based programs and policies. Aid operations became vehicles for policy reform and instruments of capacity building.

Thus, changing evaluation purposes and new policy agendas dictated the choice of disciplines and the selection of evaluation methods – not the other way round. This was in line with the pragmatic principles that have governed evaluation management since the pioneering days (Chelimsky and Shadish, 1997). Whereas prior evaluation capacity building efforts focussed on the organisational incentives needed for effective monitoring and evaluation at project level, the emphasis was now directed towards public expenditure evaluations using logical frameworks, tracking surveys, and participatory methods.

Equally, the results chain logic that used to link project inputs to project outcomes and project impacts became directed towards the complex connections that relate budget support operations to the socio-economic outcomes envisaged by Poverty Reduction Strategy Papers. Conversely, just as data constraints inhibited cost benefit analysis, poverty reduction strategists were handicapped by yawning gaps in national data gathering and interpretation.

Assessing development effectiveness: from projects to country programs

Until macroeconomists captured the commanding heights of the development profession, projects were “where the action was”. For Albert Hirschman, projects had “*much in common with the highest quests undertaken by human kind*”. They were “*units or aggregate of public investment that, however small, still evoke direct involvement by high, usually the highest, political authorities*”. They produced visible results that taxpayers in rich and poor countries alike can understand and appreciate.

Unsurprisingly, projects have continued to be essential vehicles of development assistance. The positivist assumptions that underlie projects are that (i) national leaders can be influenced through the visible impact of specific investments; (ii) societies can learn from experience and (iii) development interventions can overcome the legacy of conditions over which decision makers have little or no control (e.g. geographical handicaps, lack of skills or limited natural resource endowments).

But projects are not implemented in a vacuum. Just as they impact on the institutional environment, their beneficial impact varies according to the country context. Conversely, projects are not ends in themselves. They are levers of country development, symbols of international cooperation, metaphors for modern management, platforms for social learning and incubators of national leadership. To be sure, development effectiveness is easier to evaluate at the project level since projects have clear objectives, well defined features and a systematic approach to getting things done. They specify the shared goals, distinct accountabilities and reciprocal obligations of the partners.

As the role of good policy came to light, the project instrument was reshaped to promote explicit reforms and fashioned to generate development knowledge. Later, as governance emerged as a critical determinant of country performance, the institutional devel-

opment impact of projects emerged as a notable criterion of aid effectiveness. In short, projects have always been used as policy tools and their designs have gradually adapted to changing conceptions of development. But they involve substantial transaction costs and have no comparative advantage in countries that have acquired the institutional strength to manage effectively large scale poverty reduction programs. In such countries, budget support makes sense. Instrument selectivity is critical to aid effectiveness.

While shunned by macroeconomists who look at aid as a resource transfer, projects remain popular with politicians keen to fly the national flag of donors. They also appeal to a group of social scientists who conceive of development as microeconomic in nature and embedded in society. For them, the transformation processes associated with development are local phenomena that take place at the community level where social relationships are forged³.

By now, it has become an article of faith within the aid establishment that the success of development operations (project aid as well as program aid) should be measured in terms of their cumulative effects at the country level. Up-scaling of operational results has become a major preoccupation of aid managers. For the development community today, it is the direct and indirect impact of the portfolio of externally funded operations (along with the other services funded by the aid) rather than the aggregation of benefits from individual operations measured case by case that matters: the country has become the privileged 'unit of account'⁴.

The realization that development requires a sound policy framework and sound institutions rather than simply more and better public investment funded by aid has had a major impact on the aid industry. All aid agencies now shape their operations and sequence their interventions to achieve strategic results at the country level. Thus, the design and implementation of country assistance strategies has come to the centre stage in aid management. Typically, the design of a country assistance strategy involves the judicious structuring of

-
- 3 This perspective underlies the participatory development doctrine, the fruit of disappointment with centralized, top-down initiatives and highlights the information advantages of local actors. However, these may be offset by the risks of elite capture and misappropriation of funds in weak states (Roland-Holst and Tarp, 2002).
 - 4 While serving at the World Bank in the nineteen fifties, Paul N. Rosenstein-Rodan advocated a broadening of the project approach to encompass the entire economy – through investment in country development programs. Only when macroeconomic policy conditionality took centre stage did his vision prevail. By then, however, the 'big push' public investment driven growth theory that he had consistently promoted was discredited.

operational portfolios combined with technical cooperation and an explicit dialogue with country authorities about the policy objectives of donor involvement.

In this context, it is no longer sufficient to measure development effectiveness project by project or even program by program. Individual operations must now be conceived as building blocks of the country assistance strategy. They are expected to fit within a coherent design: the country program edifice is expected to rest on sound institutional foundations; to be buttressed by the beams and pillars of good policies and to be held together by the cement of partnership. Only then do aid projects and programs contribute to large-scale social transformation and sustainable development.

Explaining the micro-macro paradox

Once the focus moved towards country assistance strategies the goal posts of the aid enterprise were shifted to a higher plane. But since projects have remained a major vehicle for aid delivery, the *micro-macro paradox* (which holds that project results and country results diverge) has proved exceptionally damaging to the aid industry. It first came into view when the debt crisis of the early 1980's unfolded and development economics gave way to the neo-classical resurgence. Suddenly, basic questions about the premises on which aid had been provided emerged.

A cottage industry of cross-country studies came into existence. It failed to establish meaningful correlations between aid volumes and growth at country level. Three overarching conclusions emerged: (i) aid has a small impact on savings and investment behaviour; (ii) aid and growth are positively correlated in the aggregate but the effect is modest, volatile and of dubious statistical validity; and (iii) the hypothesis that good policy generates good aid outcomes has not been proven: multiple regressions and attempts to replicate the positive results with new data have failed to achieve statistical significance.

Several explanations have been offered. Each contains a grain of truth. First, it has been asserted that aid funds are fungible and therefore that donors are not financing the activities they intend to finance: at the margin, the domestic resources liberated through aid are applied to other purposes (e.g. prestige projects or military expenditures) by recipient governments. The counterargument is that projects are not neutral channels of funds. They invariably embody 'trait making' characteristics, e.g. capacity building features,

technology transfers or improved management methods. These aid effects are not fungible. Furthermore, diversion of domestic funds to low priority uses can be restrained by sound aid management that ensures that funds are used for the purposes intended and that public expenditure programs are adequately managed.

The second explanation of the micro-macro disconnect concentrates on the aggregate macroeconomic consequences of aid and suggests that, in highly aid dependent countries, aid harms the economy by creating volatility in public revenues, contributing to inflation and raising the real exchange rate so that export competitiveness suffers⁵. Thus, research by the International Monetary Fund finds that the impact of aid on growth reaches diminishing returns when the intensity of aid becomes excessive. But there is no mystery about how to control this phenomenon through competent monetary and fiscal policies and judicious economic management advice can be provided along with the aid.

The third and closely related explanation deals with the political economy dimension. Allegedly, aid in large amounts creates a 'resource curse'. Competition for control of rents aggravates social tensions. Aid becomes addictive and reduces the incentives to reform. It undermines the social contract between public authorities and citizens, hinders budget discipline and substitutes donor preferences for country priorities. Some studies even purport to show that excessive aid weakens economic and political institutions. But it stands to reason that in most cases the volumes of aid are too small to have such a pervasive and insidious effect.

The fourth explanation of the micro-macro paradox has to do with the fact that many aid agencies and nongovernmental organizations do not have credible aid evaluation systems so that the paradox may be illusory. This highlights the need for independent, high quality and rigorous aid evaluation systems.

The fifth and especially powerful explanation of the micro-macro paradox has to do with quality of aid on the supply side. Transaction costs are high: administrative costs absorb 6-7 percent of aid flows. Tying of aid generates needless mark-ups for goods and services

5 This phenomenon has been labeled the *Dutch disease*: it refers to the negative economic impact that rapid exploitation of a natural resource may have on the rest of the economy by triggering an abrupt rise in the value of the currency that makes other export products uncompetitive. The phenomenon was first observed in the Netherlands in 1634-37 when over-reliance on tulip exports diverted resources away from other productive pursuits. The discovery of large natural gas reserves in the North Sea in the 1960's evinced a similar phenomenon.

that reduce the aggregate value of the aid⁶. The quality of technical assistance funded by aid and the high cost of resident expatriates imposed by donors is another source of frustration among aid recipients. To be sure, the economic returns on well targeted and well managed technical cooperation can be astronomical since knowledge transfers can have multiplier effects and contribute to greater effectiveness of the overall financial assistance package. On the other hand, much of the technical assistance funded by aid has been provided as a quid pro quo for the assistance and it has not always been effectively used⁷.

In some countries, excessive aid flows can overwhelm the domestic administration⁸. This is made worse by aid fragmentation through numerous channels and multiple projects that siphon skills away from core government functions through the use of salary supplements, vehicles and other perks. Poor aid coordination further contributes to the inefficiency of aid delivery⁹. Here again, aid policy reform and prudent aid management could limit the damage¹⁰.

Finally, very detrimental to aid effectiveness are the distortions associated with geopolitical considerations, e.g. the global war on terror. These political imperatives help explain why the poorest countries get less than 30 percent of the aid and also why the share of aid allocated to basic social services is about half of that recommended by the United Nations (20/20 principle).

6 According to Oxfam (http://www.oxfam.org.uk/what_we_do/issues/debt_aid/mdgs_price.htm), "too often domestic interests take precedence: almost 30 percent of G7 aid money is tied to an obligation to buy goods and services from the donor country. The practice is not only self-serving, but highly inefficient; yet it is employed widely by Italy and the USA. Despite donors' agreements to untie aid to the poorest countries, only six of the 22 major donor countries have almost or completely done so".

7 According to a recent review carried out by the Independent Evaluation Group, the internal watchdog department of the World Bank, the organization "does not apply the same rigorous business practices to its capacity building work that it applies in other areas. Its tools – notably technical assistance and training – are not effectively used, and its range of instruments – notably programmatic support, Economic and Sector Work, and activities of the World Bank Institute – are not fully utilized. Moreover, most activities lack standard quality assurance processes at the design stage, and they are not routinely tracked, monitored, and evaluated".

8 Tanzania alone receives funding from 80 donors for 7,000 projects.

9 The Development Gateway, an independent foundation sponsored by the World Bank, provides internet services and information to development practitioners. It includes information on 340,000 projects.

10 Ninety one countries, twenty six donor organisations and partner countries, representatives of civil society organisations, and the private sector met in Paris on February 28-March 2, 2005 and committed their institutions and countries to harmonisation, alignment, and managing for results.

To summarize, while the micro-macro paradox has been used to discredit aid, a sober assessment of research results suggests that well managed aid does work albeit with diminishing returns as absorptive capacity constraints are reached. Thus, sound aid administration and effective aid delivery could overcome most of the obstacles that stand in the way of bridging micro and macro results.

The greatest value of the micro-macro paradox theme is that it has helped to focus on the need to reform the aid industry. The task is multifaceted: (i) to reduce the fragmentation of aid' (ii) to rely on domestic processes of aid coordination centred on poverty reduction strategy papers; (iii) to favour pooling of aid for sector wide program and budget support where country performance warrants it; (iv) to avoid political interference in aid management.

The other useful contribution of the aid effectiveness debate triggered by the micro-macro paradox has been the rediscovery of some important truths about the reality of aid. First, it is less about money than about ideas and institutions. Second, it requires sound aid policies and efficient administration. Third, it calls for effective coordination. Fourth, it needs proper alignment with country needs and priorities.

How can country assistance strategies be evaluated?

It is by now clear how shifts in development doctrines have characterized the history of aid and impacted on development evaluation. The numerous swings in the authorizing environment of aid and the evolving conceptions of development that they have generated have had a major impact on development programs. Is it possible, in this charged context, to assess objectively the development impact of country programs funded by aid?

On the one hand, workmanlike evaluation instruments have been designed and they have been tested with credible results for individual country assistance programs. On the other hand, independent and professional evaluation is still the exception rather than the rule within the aid system. Ironically, evaluation arrangements are weakest in the nongovernmental organizations (NGOs) that have been most critical of the international financial institutions. Yet the share of aid flowing through them is substantial and the proliferation of voluntary agencies has contributed to inefficiency in aid delivery.

Aid fragmentation means that the sum of individual country assistance programs by diverse donors is less than the sum of its part.

This highlights the need to carry out fully integrated evaluations of all official development assistance at the country level. This kind of evaluation has yet to be tested. But there is every reason to believe that it is feasible and that the time is ripe for carrying out such evaluations of the total impact of aid on individual countries.

Thus, in his 2003 Development Cooperation Report, the Chairman of the Development Assistance Committee of the OECD outlined a fourfold hierarchy of evaluations of aid effectiveness (impact of all aid on one country; effectiveness of the development cooperation system; evaluation of an individual donor contribution to the total system; and development effectiveness of an individual donor agency). Initial proposals for piloting evaluations focusing on the uppermost levels of this hierarchy are being reviewed by the DAC Network on Development Evaluation¹¹.

Finally, there is growing consensus within the profession regarding the basic approach to country assistance evaluations. First, the quality of country assistance strategies should not be judged merely through aggregation of project results, important though these are. High quality country programs are more than a collection of disparate projects and the interaction of projects and other aid instruments must be taken into account. It is the impact of the full package of projects and services that needs to be identified, i.e. the difference between actual outcomes and the outcomes that would have materialized without donor intervention.

In principle, this requires the estimation of counterfactuals, but the methodology of scenario building is not mature¹² and the generation of meaningful counterfactuals is still in its infancy. Therefore, the best that can be done within the budget constraints faced by evaluators is to use a mix of program evaluation methods including those that have long been in use in the assessment of social programs in industrial countries. This means in the first instance judging country assistance strategies against common criteria.

11 The World Bank joined forces with the European Bank for Reconstruction and Development (Kazakhstan); the African Development Bank (Lesotho); the Inter-American Development Bank (Peru and Rwanda) and the Islamic Development Bank (Jordan and Tunisia) while Norway and Sweden and Australia and New Zealand teamed up for reviews of their Malawi and Papua and New Guinea programs respectively.

12 Long term growth models (let alone large-scale econometric models) are expensive to construct and they are not very reliable. Country comparisons can provide useful pointers but the performance of one country cannot be used as a reliable benchmark for another since no two countries are alike in their factor endowments and their institutional frameworks.

First, high quality country assistance strategies should be selective. Their priority areas should be selected with care so that projects and other development services included in country programs form a synergistic whole both relative to one another and to the interventions of other donors. The right instruments should be selected. The design of operations should be grounded in a constructive dialogue with country authorities and should take account of the interests and capabilities of other partners. Projects and other services should be competently managed in line with the operational policies of the donor and backed by professional analyses of development potentials, policy constraints and capacity building needs.

Second, verifying compliance of country strategies with the development doctrines currently in vogue is not a useful test: each developing country is unique and the track record of grand development theories has proven to be mediocre. The pertinence of country assistance goals must be judged case by case taking account of country potentials and needs, implementation capacities and the determination of country authorities to address policy obstacles.

Third, development results do not always equate with aid performance not only because aid accounts for a small part of the government's budget in most instances¹³ but also because country level outcomes are ultimately shaped by the host of historical, geographical, political and policy factors.

In the absence of resilient hypotheses about the linkages between policy inputs and development performance, country assistance strategies cannot be evaluated by simple linear methods that examine the extent to which operations are geared to pre-ordained policy tenets. More reliable is triangulation of evaluation methods focused on three major dimensions¹⁴:

- the quality of individual operations, country dialogues, coordination with partners and analytical/advisory services;
- a development impact assessment, involving a "top-down" analysis of the principal program objectives and their achievements

13 Aid accounts for less than 10 percent of public expenditures in over 70 percent of recipient countries.

14 Whereas this approach reflects international financial institution experience, other development agencies use somewhat different approaches. For example, the European Union considers the impact of aid and non aid policy vectors in assessing the relevance, quality and size of its country program and the resulting influence on the recipient country and its partners. The Swiss Development Corporation emphasizes participatory techniques and country involvement in the evaluation process.

in terms of their relevance, efficacy, efficiency, resilience to risk and institutional impact; and;

- an analysis of attribution (or contribution) in which the evaluator assigns responsibility for program outcomes to the various actors according to their distinctive accountabilities and reciprocal obligations.

In evaluating the expected development impact of an assistance program, the evaluator should gauge the extent to which major strategic objectives are relevant and are likely to be achieved without material shortcomings. Programs typically express their goals in terms of higher-order objectives, such as poverty reduction or attainment of the millennium development goals. The country assistance strategy may also establish intermediate goals, such as improved targeting of social services or promotion of integrated rural development, and specify how they are expected to contribute toward achieving the higher-order objective.

The evaluator's task is then to validate whether the intermediate objectives have produced (or are expected to produce) satisfactory net benefits, and whether the results chain specified in the country assistance strategy was valid. Where causal linkages are not adequately specified upfront, it is the evaluator's task to reconstruct the causal chain from the available evidence, and assess relevance, efficacy, and outcome with reference to the intermediate and higher-order objectives.

Evaluators should also assess the degree of client ownership of international development priorities, such as the Millennium Development Goals, at national and, as appropriate, sub-national levels. They examine compliance with donor policies, such as social, environmental and fiduciary safeguards. Ideally, conflicting priorities are identified in the strategy document thus enabling the evaluator to focus on whether the trade-offs adopted were appropriate. However, the strategy may have glossed over difficulties or avoided addressing key development priorities or policy constraints. This inevitably affects the evaluator's judgment of program relevance.

The efficacy of program implementation should be judged by the extent to which program objectives are expected to be met in ways that are consistent with corporate policies. Efficiency ratings concern the transaction costs incurred by the donors and the country in connection with the implementation of the country assistance program. Finally, sustainability has to do with the resilience of country

assistance achievements over time and institutional development impact refers to the capacity building benefits of the country assistance strategy.

Global changes will affect the future of development evaluation

The shift in development paradigm is not over: we have not yet reached the end of development history! The evaluation profession is in transition as it seeks to fulfil the demanding circumstances of an increasingly interconnected global order: the ascent of development evaluation to a higher plane continues. Having moved from the project level to the country level, it is now poised to move to the global level. The interconnectedness of markets, nations and non-state actors is gradually changing the focus of development cooperation as vertical aid programs geared to the resolution of the diverse “problems without passport” that hinder development across country boundaries multiply.

The planet is getting smaller and now more than ever the diverse peoples of the world are living a single history. OECD countries rely on developing countries for a third of their export sales and one half of their oil consumption and developing countries depend on OECD countries for over 60% of their trade and about half of their commodity imports. Large mismatches between economic and political organization have emerged at community, national and transnational levels.

Rich countries exercise control over the institutions that oversee the global economy. It is their rules and their standards that regulate the flows of capital, people and ideas. It is their production and consumption patterns that pose the greatest threat to the global environment. Only new rules of the game can create a level playing field between rich and poor countries in the global market place. During the eighties and nineties the development evaluation community concluded that national policies in poor countries exert a crucial impact on aid outcomes.

Accordingly, aid managers acted on this finding by promoting national policy reform. In the new millennium, the same logic will have to be applied at the higher plane of global policy. The policies of rich countries matter quite as much for global poverty reduction as the policies of poor countries. Civil society activists and policy researchers have long highlighted the need to make globalization work for the benefit of all. They have finally succeeded in inducing

policy makers in OECD countries to conceive of development cooperation as a 'whole of government' endeavor.

The critical role that rich countries policies play in development means that the social sciences and development evaluation will have to address policy coherence for development far more than in the past. Richard Manning, then DAC Chairman, addressing OECD aid ministers put it this way: "*Coherent policies for development ... cannot be mandated by the development community. But we have both a need and a responsibility to ensure that the development dimension is indeed fully understood and taken into account, since if it is not, much of our spending will be merely offsetting the costs imposed on our partners by other policies of our own governments.*"

Thus, development cooperation is being redefined to extend beyond aid and policy coherence for development has become the new leitmotiv of the development enterprise.

Accordingly, the time has come for evaluators to devote more resources to the higher plane of global policy. Just as project level results cannot be explained without reference to the quality of country policies, country level evaluations are incomplete without reference to the international enabling environment.

This is because new mechanisms of resource transfer are dwarfing the 'money' impact of aid and creating brand new connections between rich and poor countries (as well as among poor countries). The private sector is already vastly outpacing the public sector both as a source and as a recipient of loans and grants. Worker remittances are growing rapidly and were expected to exceed \$230 billion in 2005. Another \$260 billion worth of foreign direct investment, equity flows and commercial loans is directed at poor countries. Thus, total private flows are at least four times as high as aid flows. The net welfare benefits that could flow from trade liberalization also represent a multiple of aid flows especially if punishing tariffs against labour intensive products are reduced, workers of poor countries are allowed temporary access to rich countries and food importing countries are induced to generate a successful agricultural supply response through 'aid for trade' schemes'.

Knowledge flows need liberalization too. The intellectual property rules imposed during the Uruguay round involve a reverse flow of the same order of magnitude as current aid flows. While some relaxation of the TRIPS agreement was introduced under the Doha round for life saving drugs and technological development does require

patent protection, special provisions for encouraging research relevant to poor countries, for bridging the digital divide and for filling the science and technology gaps of the poorest countries are warranted to level the playing field of the global knowledge economy. Finally, the environmental practices of rich countries and the growing appetite for energy of the Asia giants may induce global warming costs for developing countries likely to exceed the value (4-22 percent vs. 7 percent of national incomes) through losses in agricultural productivity.

In combination, all of these trends mean that (except for the smallest, poorest and most aid dependent countries where coordination will continue to pose major challenges) the relative importance of aid flows compared to other policy instruments (trade, migration, foreign direct investment, etc.) has been reduced as a direct result of globalisation. But aid will remain critical to attend to emergency situations and post conflict reconstruction, as a midwife for policy reform, as a vehicle for knowledge, technology and management practices, as an instrument of capacity building (especially for security sector reform) and as a catalyst for conflict prevention.

Programmatic aid and budget support are useful aid vehicles in well managed countries. But wielded with skill and professionalism, the project instrument is regaining some of the allure it lost when the neo classical resurgence required a massive diversion of aid flows towards policy based quick disbursing loans and budget support operations. Already infrastructure development and natural resource extraction projects equipped with social and environmental safeguards are making a comeback, mostly through support to private enterprises and voluntary agencies, especially in weak states. Aid for community based social protection schemes is also rising given continuing public support for the notion that development is a bottom up, micro-process.

In brief, through the revival of investment lending geared to the creation of institutions, the promotion of private investment and the mobilization of communities and voluntary organizations, the micro-macro paradox could be exorcised since it only haunts the money dimension of aid. Not that policy based lending will disappear altogether. Many poor countries still need to improve their macroeconomic and their structural policies, especially those related to trade facilitation and the enabling environment for private enterprise. But they may elect to do so through free standing advice and capacity building assistance rather than repeated and addictive dollops of quick disbursing funds.

What is to be done?

First and foremost, aid should no longer be viewed as the only tool in the development cooperation kit. Coherence among conflicting aims remains a major challenge for development cooperation¹⁵. A whole of government approach is needed to ensure that policy coherence for development becomes the driving force of donor countries' relations with poor countries. This means that trade, migration, foreign direct investment, intellectual property and environmental policies should all be shaped to benefit poor countries or at least to avoid doing them harm. From this perspective, aid should be viewed as the connecting thread between all policies that connect the donor country with each developing country. This implies different kinds of country assistance strategies. To help support the reorientation, multilateral agencies should use their analytical skills to evaluate and monitor the quality of rich countries' policies towards poor countries.

Second, the downside risks of current development patterns should be acknowledged and conflict prevention, conflict management, post conflict reconstruction and security sector reform should move to centre stage in country assistance strategies. In parallel, multilateral agencies and regional organizations should use their convening power and their management skills to organize mission oriented networks involving governments, the private sector and the civil society to design and implement collaborative programs. They would aim at global or regional threats to peace and prosperity and they would be implemented at global, national and sub-national levels.

Already, major coalitions of donors are seeking to address such development challenges as HIV/AIDS that do not respect national borders. Increasingly, they will be mobilized to tackle the myriad illegal activities that constitute the dark side of globalization (e.g. the booming trafficking of drugs, arms and people) by combining law enforcement with development alternatives. In a nutshell, dealing with the downside risks of globalisation will require adopting a human security model of development that continues to favour growth but with greater priority to economic equity, social inclusion and environmental sustainability.

15 In the United States and among some of its allies the war on terror has replaced the anti-communist crusade as a geopolitical rationale for development assistance and this constitutes a major threat to development effectiveness as well as a potentially destabilizing approach to international relations.

Third, aid should no longer be conceived and evaluated as a resource transfer mechanism. Instead, it should be conceived as a transmission belt for ideas, a device to train development leaders, an instrument to build state capacity and a platform for policy experimentation and dissemination based on good analytical work and sensitive advisory service. In the poorest, aid-dependent countries, the convening power of multilateral institutions should be used to help overcome the growing fragmentation of aid. Towards this end, the commitments made by donors to improve aid quality, eliminate tied aid, reduce transaction costs, harmonize policies across donor agencies and align aid objectives with country felt needs and public expenditures processes should be met. But this does not mean that the project vehicle should be jettisoned. Well designed and professionally implemented through donor coalitions it can yield considerable benefits. Instrument selectivity is central to aid effectiveness.

Fourth, country assistance programs should be tailored to the political economy. Human security considerations should be prominent in strategy design. Governance should be professionally assessed and conflict analysis should ensure that aid does no harm and that horizontal inequalities are taken into account in project designs. Standard, blueprint models reflecting doctrinal positions (e.g. with respect to privatization) should be jettisoned and transfer of good practice properly adapted to the country context should be emphasized. Where government authorities are not committed to development, non aid instruments should be used and aid should emphasize infrastructure, the private sector and civil society channels as well as local government and community level organizations where good leadership can be identified and future leaders trained. Budget support has its place but not always and not everywhere.

Fifth, given limited resources, selectivity is essential but the current aid allocation system short-changes fragile states. Policy research has established that they are currently receiving 40% less than they should even if policy performance considerations are taken into account. Combining the potential conflict prevention benefits to the satisfactory outcomes at project level confirmed by independent evaluations of almost sixty percent (60% of projects approved by the World Bank in fragile states during 1998-2002¹⁶) would suggest that high risks can lead to high rewards. It is also notable that the performance of private sector projects funded by the International

16 Furthermore, current aid allocation rules do not take account of the benefits of preventing conflict. Research by Paul Collier suggests that, on the average, preventing a single war would save USD 64 billion a year.

Finance Corporation has been as good in fragile states as elsewhere¹⁷.

Sixth, development education should have high priority. The public in the industrial democracies should be exposed to the reality of aid, its inevitable challenges and its exciting opportunities. Currently voters vastly overestimate the share of government budgets allocated to aid¹⁸. Most are unaware that total aid flows declined from about 0.65 percent of the national incomes of OECD countries in 1967 to 0.25 percent today¹⁹ or that aid absorbs only a twentieth of the resources absorbed by the military. The self interest rationale of development cooperation in the era of globalization should be clearly articulated. In an interconnected world the problems of others have become our own. There is no prosperity without peace and there is no peace without justice.

Finally, it was right and appropriate for the unit of account for development evaluation to move from the project to the country. But the time will soon come when it will have to move again to a still higher plane – the regional and global level of the development cooperation enterprise.

References

- Chelimsky Eleanor, and Shadish William R., (1997) *Evaluation for the 21st Century. A handbook*. Sage Publications: Thousand Oaks, London, UK and New Delhi, India.
- Development Assistance Committee, (2002) *Principles of Effective Aid*. OECD, Paris, France.
- Kuhn, Thomas S., (1996) *The Structure of Scientific Revolutions*. University of Chicago Press. Chicago and London.
- Nagy Hanna, et. al., (1999) *Annual Review of Development Effectiveness*. Operations Evaluation Department. The World Bank. Washington D.C.
- Willoughby Christopher, (2003) *First Experiments in Operations Evaluation: Roots, Hopes and Gaps in World Bank Operations Evaluation Department, The First 30 Years*. Washington D.C., USA.

-
- 17 This conclusion is based on the degree of loss reserves, historic write-offs, default rates, equity investment measures, and independent ratings of development outcomes, normalised for the class of investment.
- 18 Americans think that the US spends 24 percent of the federal budget on aid. They believe that 10 percent should be spent in this way whereas, in fact, the US dedicates less than 1 percent of the federal budget to aid.
- 19 The United States that allocated 2 percent of its national income to the Marshall Plan now contributes less than 0.2% of its national income for aid to poor countries.

THE STRATEGIC INTENT. UNDERSTANDING STRATEGIC INTENT IS THE KEY TO SUCCESSFUL COUNTRY-LED MONITORING AND EVALUATION SYSTEMS

Jean Serge Quesnel, Professor at the United Nations System Staff College, Adjunct Professor at Carleton University and Professeur Associé at the École Nationale d'Administration Publique of Quebec

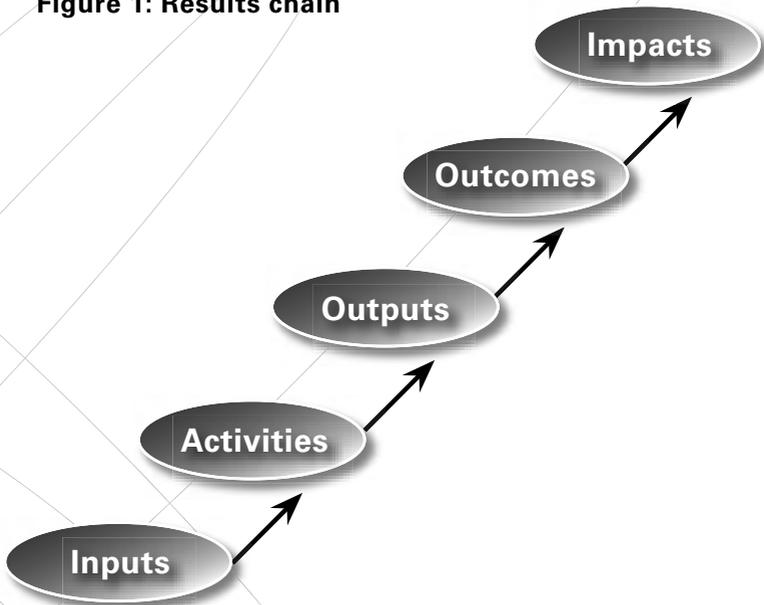
Understanding the strategic intent is an essential requisite for any relevant and efficient country-led monitoring and evaluation (M&E). The strategic intent makes explicit the aim of the developmental intervention being pursued, and provides coherence to country efforts and external support. It fosters greater effectiveness of the scenario being implemented and facilitates the measurement of achievements. Academic literature tends to present the strategic intent using a monolithic view. This article will present a generic definition and illustrate various applications of the strategic intent at different levels of management, using different results-based paradigms. This article will then conclude that country-based M&E systems need to start with an explicit enunciation of the strategic intent.

What strategic intent is not

In 2000, when I had just joined UNICEF, I made a presentation on the vision that I had for use of evaluation in UNICEF and the United Nations. During the presentation, I kept referring to RBM (results-based management). After the presentation a few colleagues told me that they did not understand why I kept referring to roll-back malaria. The same anecdotal situation repeated itself when, in 2003, I asked for greater clarity of the *strategic intent* of UNICEF interventions. I was told to use the currently popular generic term *result*. To no avail, I explained that one uses different terms to express different concepts. Let us review the term *result* and others like outputs; outcomes; impacts; goals; objectives; mission; vision; and, determine that they all fall short of being the expression of a strategic intent.

The Working Party on Aid Evaluation of the Development Assistance Committee of the Organisation for Economic Cooperation and Development (OECD-DAC), defines *Results* as being the “output, outcome or impact (intended or unintended, positive and/or negative) of a development intervention”. This is a widely referred to definition which is based on the logical framework analysis approach (LFA), developed by Practical Concepts Inc in 1971. The LFA is a corner-stone tool used to define project expectations. Its modern version has led to the results chain, now being used globally and illustrated as follows:

Figure 1: Results chain



In this conceptual model, inputs of resources and efforts yield targeted outputs which are results normally under the full control of the manager of the intervention. The outputs in their turn generate intermediate results (outcomes), some under direct control and others under indirect influence. These outcomes, conditional to critical assumptions, are expected to provide the final outcomes of the intervention which are the desired impacts.

The main criticism of the results chain conceptual model is that it is simplistic, linear and does not reflect the reality of multivariate factors at play. It responds to a supply-driven propulsion, with the

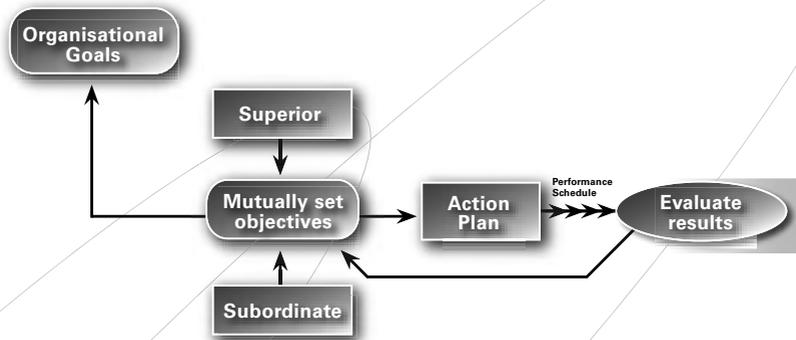
assumption that access to resources (inputs) will suffice to provoke a causal chain of events. The model does not reflect the complexities of factors at play, or the involvement of many actors who often have different motives. The emphasis is put on the outputs assuming that having achieved the outcomes, the impacts logically will materialise. However useful and relevant impacts may not be sufficient to fulfill the nevralgic¹ attraction of a strategic intent.

In the LFA, outcomes may be the *goals* of the intervention. The goals state what is to be achieved and when. They are the immediate results expected once the intervention has been implemented. They give a description of the expected situation upon completion of the implementation of the intervention. They also provide evidence to fund-providers that value-for-money is gained in the short run. Strategic intent is at a different level from goals; it is *super ordinate* to them.

The LFA is an institutionalised expression of the popular management approach called *MBO -Management by Objectives*. Paul Mali describes MBO as a strategy whose basic idea is the setting and pursuit of attainable objectives. MBO is a practical way to facilitate a cascading down, of planning for results, by management. It enables organisational alignment and discipline around strategic goals and it fosters bottom-up initiatives. When all levels of management participate in a strategy, a system emerges in which key individuals are coordinated to move in a given direction. Objectives tend to be the improvements a manager wishes to initiate in his/her area of responsibilities. Once missions and goals are established by an organisation, a superior and a subordinate at the beginning of a time period, participate mutually in setting and agreeing on performance objectives to be completed during the period, as shown by the figure below. The mutual setting of objectives starts at the top of the organisation and continues down to the lower levels of management. Each objective is supported with an action plan and implementation schedule. At the end of the time period, superior and subordinate mutually evaluate actual results and proceed to set objectives for the next cycle. Application of MBO is almost universal in the organisation since all tasks, activities, projects and programmes, from the simplest to the more complex, must have an objective.

1 The term nevralgic is borrowed from the French term *névralgique*. It is used in this article analogically to its military use, meaning the focused purpose of a decisive strategic intervention aiming at having the highest intended effect. The medical etymology point to the Greek words neuron (nerve) and algos (pain).

Figure 2: Management by objectives



Some of the disadvantages of MBO are:

- MBO requires many planning cycles, often with several iterations between levels of management;
- pursuit of objectives is often disrupted by unforeseen circumstances;
- MBO tends to be closed to new opportunities within the implementation period because changes may overshadow stated and agreed objectives;
- it also tends to focus on quantifiable objectives, ignoring important non-quantifiables;
- after a while, MBO tends to be applied mechanically and objectives loosen, because all concerned do not share the same sense of drive and commitment as expected when pursuing a strategic intent.

Long term objectives are frequently known as the *mission statement* which defines the purpose of the organisation. George Odiorne (1981) says the mission describes the condition that will exist if one succeeds. It answers the question what are we in business for. The mission may define the client, the product/ service and the expected quality. It defines these as *indicators* by which decisions may be taken and resource allocation chosen. These indicators are criteria around which all subsequent actions of the organisation and its managers may be judged to have succeeded or failed. In defining the mission, one identifies at the same moment any *gaps* that might exist between the mission and the actual conditions both

internal and external to the organisation. Optimisation of the mission is sought with the use of management techniques such as cost-effectiveness studies, profit planning for the private sector and zero-based budgeting for the non-profit sectors.

Goals, as subordinates of the mission, are the basis for keeping the organisation on course; identifying strengths and weaknesses; allocating resources most effectively; isolating alternative courses of action; providing decision rules for operations; appraising new business proposals; identifying and minimizing the impact of external factors in the environment that could affect the mission; developing plans for bad times; and, maintaining flexibility in operations without losing sight of the main purpose of the organisation. The mission is different from the strategic intent because its strategic drivers remain within the organisation, whereas, the latter aims at making a difference in a reality external to the organisation.

A *vision*, on the other hand, is defined by Gardner & Avolio (1998) as a set of desired goals and activities. It has connotations of encouraging strong organisational values in the strategy process. Therefore it is similar to strategic intent in its emotional effects. The vision goes beyond mere planning and strategy by challenging organisational members to go beyond the status quo. It offers long term direction. Mantere & Sillince (2006/7) wrote that the difference between visions and strategic intents is the degree of collectivism, as many ascribe a strategic intent as a phenomenon diffused at multiple organisational levels while a vision is more clearly a top management leadership tool, often accredited to a single visionary leader.

Acceptance of a future vision, entailing a new set of beliefs about the identity and capability of the organisation, unleashes the creative thinking necessary to invent ways of achieving the strategic intent. Peter Senge (1990) wrote that there are only two possible ways for creative tension to resolve itself: pull current reality toward the vision or pull the vision toward reality. Which of these occurs will depend on whether one holds steady to the vision.

What strategic intent is

In 1989, Gary Hamel and C.K. Prahalad made known the expression *strategic intent* when they published an article of the same name in the Harvard Business Review. They argued that in order to achieve success, a company must reconcile its end to its means through strategic intent. In their book "Competing for the future" they define strategic intent " as an ambitious and compelling...dream

that energizes...that provides the emotional and intellectual energy for the journey... to the future.” Hamel and Prahalad (1989) provide three attributes for the strategic intent:

- Sense of direction. Strategic intent implies a particular point of view about the long term market or competitive position that a firm hopes to build over the coming decade or so. It is a view of the future – conveying a unifying and personalized sense of direction.
- Sense of discovery. Strategic intent is differentiated. It implies a competitively unique point of view about the future. It holds out to employees the promise of exploring new competitive territory.
- Sense of destiny. Strategic intent has an emotional dimension. It is a goal that employees perceive as inherently worthwhile.

Figure 3: Basic steps to identify, enunciate and implement a strategic intent



Adapted from Arthur Schneiderman, *Total Quality Management*, 1990.

A typical strategic intent process starts with the three attributes. The leader sets challenges and communicates them to the entire workforce. The challenges are a means to get into the strategic intent. A key dimension is the realisation that the strategic intent involves everyone. In order to set the right challenges that will yield the strategic intent, it is important to have an insightful and incisive perception of the problem to be addressed and its root causes. One has to be able to identify the key factors that will have a nevralgic effect. The following chart illustrates graphically the steps required to identify, set, implement and assess the achievement of a strategic intent.

When Charles Smith teased out the essence of the strategic intent, he referred to the *Merlin Factor*. Merlin the Magician was King Arthur's mentor (White, 1958). He had the ability to know the future because "he was born on the other end of time and had to live backward from in front, while surrounded by people living forward from behind..." The Merlin Factor is the ability to see the potential of the present from the point of view of the future. It enables a "future-first" perspective adopted by leadership that successfully instills strategic intent in their organisation. Charles Smith explains that the characteristics of the Merlin Factor expressed in leadership are what make the difference in organisational change. The process is one in which leadership teams transform themselves and the culture of their organisation through creative commitment to a radically different future.

Leading from the premise of a strategic intent requires one to think and plan backwards from that envisioned future in order to take effective action in the present. Leaders who employ the Merlin Factor are engaged in a continual process of revealing the desired future in the competitive opportunities of the present. In this sense a leader works rather like the sculptor who, when asked to explain how he had turned a featureless block of marble into a wildlife tableau, replied: "I just chipped off all the parts that didn't look like an elephant."

Merlin leadership starts with personal vision of the organization's future which confronts the shared reality of its existing culture. As other members of the organization make their own commitments to this vision it becomes a strategic intent. In many cases, commitment to the strategic intent preceded the development of the requisite methods for accomplishing it. Managerial 'Merlins' play a critical role in this process by consistently representing the stra-

tegic intent in an ongoing dialogue with the existing organizational culture. The leader is an 'attractor' in the field of creative tension between the entrenched culture and the new strategic vision.

Strategic intent obviously implies intentionality. John Searle (1994) says that "intentionality is that property of many mental states and events by which they are directed at or about, or of object or states of affairs in the world." Intent is a psychological concept which is possessed by a conscious actor. Mantere & Sillince (2006/7) say that organizations are not conscious and cannot possess intent in a strict sense, i.e., organizational intent needs to be possessed by some or all of its members. Organizations are often pluralistic and fragmented, which underlines the necessity to be explicit regarding subjectivity when addressing mental phenomena on the organizational level of analysis. Key to making sense of collective intentionality is the question of what is meant by the pronoun 'we'. Authors on strategic intent seem to be in disagreement over whether the "we" of the strategic intent is the top management team or, whether the "we" is more plural and diffused. The literature appears to miss an important issue: the possibility that the same intent(s) may exist in different variations within the organisation. The literature also misses a potentially important role for organisational strategic intent: the building of coherence between multiple intents. Everett Rogers (2003) wrote: "Strategic intent, when communicated to an organisation, is reinvented as multiple intents as it is diffused among lower level managers and operative employees."

The Merlin factor enables a clear strategic intent. One starts by looking at the endgame –where one wants to go. This is not just talking about SMART² goals. It's about what kind of legacy the organisation wants to build in its community and its professional environment. By starting at the end, one can crystallize organisational and personal dreams and together identify strategic thrusts, long term milestones and actionable steps to reach them. One has to step back at critical junctures to make certain that present endeavours are aligned with the long-term objectives. The plan of the strategic intent becomes the guideline for how efforts get aligned, results assessed and value generated in a synergetic fashion.

Vadim Kotelnikov puts it this way. "The strategic focus is the starting point for developing a statement of strategic intent. A statement of strategy must become then a statement of design through

2 SMART means Specific and Simple, Measurable, Achievable and Attributable, Relevant and Realistic, Time-bound, Timely, Trackable and Targeted.

which the principles, processes and practices of an organisation are developed. These statements must represent the whole as seen from any location in the organisation.” Strategic intent is a high-level statement of the means by which the organization will achieve its vision. It is a core component of a dynamic strategy. Hamel and Prahalad (1989) say that the strategic intent cannot all be planned in advance. It must evolve on the basis of experience during its implementation. As Melissa Kelly-McCabe (2007) writes: “Imagine the power of people working together toward a common aim, uncovering possibilities and leveraging strengths.”

In his article *What Really Matters*, Andrew Spany (2003) provides eight principles that enlighten the business process of the strategic intent. They are:

- The first essential principle is to look at the organisation from the outside-in, from the customer’s perspective, as well as the inside-out’.
- The second principle states that the organisation’s strategy needs to be tightly integrated with business process management.
- The third principle is to articulate the organisation’s strategy in such a way that it inspires, from the boardroom to the lunchroom, and remains front and center throughout the year.
- The fourth principle is the launching pad for organizational alignment. It states that action needs to be taken to assure that the organization’s core business processes are designed to deliver on its strategic goals.
- The fifth principle says that the organization design must enable business process execution. In this context, organization design is defined as the composite of structure, measures and rewards.
- The sixth key principle addresses the need to assess and deploy enabling technology based on the value added through enhanced business process performance.
- The seventh principle states that it is essential to hard-wire the enterprise-wide performance measurement system to budgets and operating reviews.
- The final principle is sustained focus and alignment.

According to Andrew Spany (2003), old solutions don’t work anymore. The time for functional thinking, with all of its attendant weaknesses, is past. The organizational capability approach offers

a contemporary, engaging, and action-oriented approach. Achieving superior, sustainable performance isn't easy at the best of times, and the current business environment makes it that much more difficult. Strategic focus, organizational alignment, and operating discipline will appeal to those leaders who are passionate about winning, challenging them to think systematically as well as systemically. Spany also quotes Miyamoto Musashi, a Sumurai warrior as having said: "In Strategy, it is important to see distant things as if they are close and to take a distanced view of close things."

Frank Greif believes "...that organisations are more successful when they take the time to create a clear sense of purpose. The strategic intent is defined as a compelling statement about what you are doing and where you are going. It's really more than a statement: it becomes a core element in the motivational DNA of the organisation. Yet strategic intent is not enough by itself. To succeed in today's rapidly changing and multidimensional reality each of us must learn to communicate in ways that are deliberate, challenging and inclusive. We have to talk to each other and listen to each other with clarity, honesty and integrity. For leaders, there are no more important skills than developing and communicating purpose, passion and commitment." Pamela Lovell and Julie Kelly wrote: "Intentional leadership aims to address the fragmentation that many people experience and move toward wholeness so that you can give your best to each interaction."

Robert Barthelemy said: "When I think of transformation of airplane to spaceplane, to me that's kind of like the Holy Grail, in the technology world. I think that conjures up images of alchemy, or magic. If you look at when magic occurs in the mythologies, it's always because there's a quest in progress that forces magic to occur. No quest, no magic."

Charles Smith (1994) clarifies Barthelemy's statement writing:

"In the quest to achieve your organization's strategic intent, the destination is fixed but the path is opportunistic. Unpredictable things happen on quests. Helpers, hindrances and tests of resolve appear unexpectedly, as if by magic. To lead through the Merlin Factor one must be a master of change, sensitive to the interaction of long range strategy and emergent circumstance. You will want to be armed with the normal range of business disciplines as you pursue your strategic quest, but remain alert for irregularities, exceptions and other interruptions in your plans. They may conceal the one

thing you never realized you would need in order to achieve your goal. That's where the magic of strategic intent lurks: in the possibilities you couldn't have foreseen when you made your initial commitment. Merlin-like leaders cultivate a mental state of search rather than certainty. If you refuse to be seduced by the understandable desire to feel in control at all times, serendipity will often assist you on your way. But you have to be looking for the magic of unanticipated opportunity before you can recognize it."

Saku Mantere and John Sillince (2007) summarised well the definition of strategic intent. They say that "Strategic Intent is a set of social constructions, governing future-oriented behaviour, which is (1) super ordinate to a goal; (2) long term or very long term; (3) uncertain in its achievability; (4) linked to core competences; (5) of high significance; (6) prospective; (7) inspirational; (8) directional; (9) integrative; and (10) a process."

Strategic intent pursued at different levels of management

In management literature the propensity is to view the *strategic intent* as a beacon that is set and comes from the senior management. The strategic governance of the organisation focuses on a clear enunciation of the key strategic deliverables. It is much a supply-driven endeavour with controls resting on the side of those who propose initiatives. The hierarchy under the strategic intent trickles down to the working levels.

For illustrative purpose, the hierarchy of the strategic intent at Cobleskill (State University of New York³ (SUNY)), is projected as follows on its website:

At the top of the hierarchy is the organization's Vision and Mission, both of which are long-lasting and motivating. At the bottom of the hierarchy are the projects and short-term tactics that faculty and staff members use to achieve the Mission.

3 Source : <http://cobyweb.cobleskill.edu/StrategicPlan/03.html>

Anyone inquiring as to why a SUNY Cobleskill representative is acting in some way should be able to look up the hierarchy to find the reasoning. If seeking to determine how SUNY Cobleskill will accomplish something, one should be able to look down the hierarchy.

Another exemplary enunciation of this approach may be found in the document called *Strategic Intent* published by the Central Intelligence Agency of the

Government of the United States of America⁴. The ALNAP Strategy 2008-2013 is also a good reference.⁵

In most complex organisations that have a decentralised governance system, one does not find the same monolithic management approach, as described above. In global multi-dimensional international organisations such as UNICEF, there are distinct levels of management. These levels respond to different levels of risk appetite and forms of participatory management. From risk adverse management frameworks to bold approaches to experimentation, one may identify five levels of management. At each level, versions of the strategic intent approach are implemented with different execution paradigm. Members of a multinational organisation are not likely to formulate a very deep understanding on the whole organisation role set. Indeed, Jarzabkowski (2005) explains how studies of larger and more pluralistic organisational context portray strategic intent as a distributed, fragmented and contested concept.

Figure 4: Hierarchy of the strategic intent



4 A good practice of such an enunciation may be found at the following website: <https://www.cia.gov/about-cia/strategic-intent-2007-2011.html>

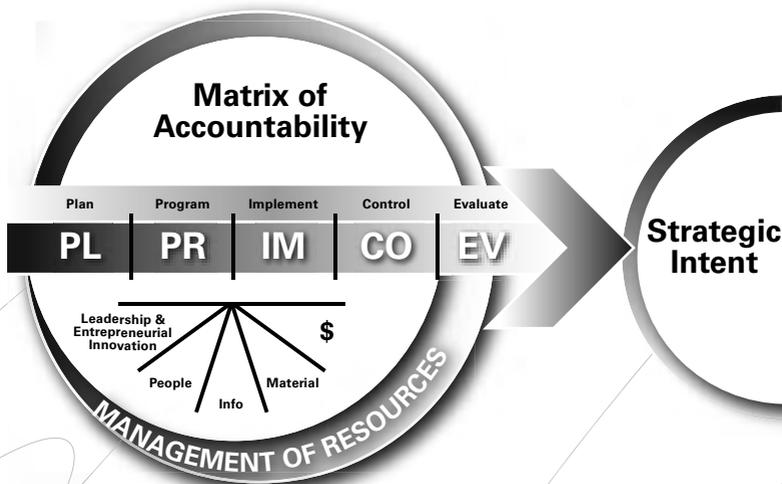
5 The Strategy of the Active Learning Network for Accountability and Performance in Humanitarian Action may be found at the website http://www.alnap.org/pdfs/alnap_strategy_2008-2013.pdf

In order to find a common denominator that will enable alignment and connectedness of the various levels of management the following management framework is useful. Here are its components:

- It starts with the enunciation of a SMART strategic intent.
- Then it draws the management process supporting the achievement of the strategic intent. To do so, it uses the classic management process of planning, programming, implementing, controlling and evaluating.
- After it takes into consideration the actors involved, their roles and their accountabilities.
- Finally it considers the use and optimisation of resources necessary to achieve the strategic intent.

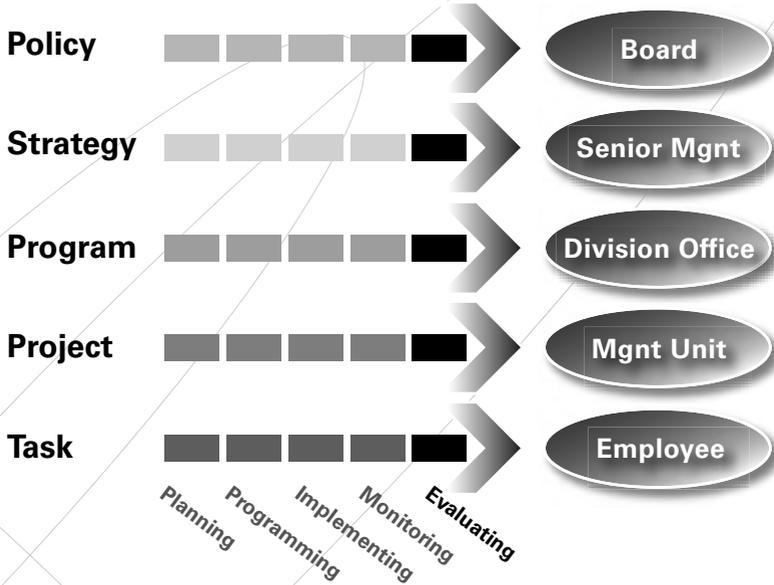
Graphically it can be summarised as follows.

Figure 5: Management framework



Using this management framework as the common denominator, one may view its application at five levels of management within multilateral complex organisation, as illustrated below.

Figure 6: Management levels



This management framework illustrates that the governing body of an organisation sets the overall strategic intent which is implemented by means of various strategies defined by senior management. The strategic intents of the organisation's strategies are implemented by programmes. The strategic intents of the programmes are in turn implemented by projects and activities. The project objectives ("strategic intents") are achieved by the execution of orchestrated tasks.

If we look from end to end of the hierarchy, at the bottom we see the tasks level. There, the purpose ("strategic intent") is well defined and the procedure aims at optimising the efficiency of the delivery of that intent. At the top of the hierarchy we see the policies level. There, the challenge is to define the strategic intents in a SMART fashion, enabling concerned actors to implement scenarios with flexibility adapting them in the light of opportunities and hindrances. The tasks and projects levels usually adopt a closed system approach. The policies and strategies levels require an open system approach because too many factors escape the immediate control of stakeholders. Usually, at the programmes level, a semi-structured approach is followed, defining basic parameters yet enabling different implementation scripts depending on internal and external factors at play.

Management paradigms at each level are different. At the tasks level, the procedure dictates the way the "strategic intent" is reached. The highly structured approach is heavily anchored in ways and motions and systematic processing, leaving little space for adjusting the scope of the intent. At the project level, task sequencing is plotted for the optimal use of resources aiming at delivery within the shortest time period and at least cost. The "critical path" serves a roadmap for the implementation of the optimum scenario maximising value-for-money and risk minimisation. The management emphasis at these two lower levels is on the delivery process. Because planning is done within a closed system approach, one assumes that the "strategic intents" will be achieved if the implementation processes are correctly executed.

At the programmes level the "strategic intent" aims at creating an intended change from a situation "1" to a situation "n". The achievement of the strategic intent implies a collaborative understanding among stakeholders. The underpinnings of the strategic intent structurally rest on a logic model, explicit or implicit, that involves factors causally affecting each other. The optimum programme design entails the identification of the key factors that

have synergetic influence on the systemic configuration of the logic model, and address the root causes of the problematic being resolved. Acting on key factors that make a systemic difference, programme actors collaborate and progress toward the achievement of the strategic intent of the situation “n” desired.

In the organisational universe, at the strategies level, the “strategic intent” usually adopts a symmetrical form akin to institutional performance. In academia, it is at this level that the expression “strategic intent” was coined. As stated above, strategic intent implies the alignment of the vision; mission; core values; due consideration of the strategic environment; and, response to stakeholders’ expectations. These are all translated into SMART organisational goals, calibrated in organisational plans aiming at the optimum use of resources, and influence leveraging by means of strategic alliances and partnerships.

At the policies level, one expects sagacity, prudence, practical wisdom and shrewdness, consensus building and expediency. The statement⁶ “Policy demands occasional compromise” infers the need to have broad-minded and open-system approaches. Stakeholders have their own universes of interest. To create a commonly understood and binding policy implies the overlap of these universes and finding the largest “consensus space of agreement” if the policy is to be sustainable and adhered to. Policies are the expression of definite courses of action adopted and pursued by governing bodies and administrations, whether public, non-governmental and private. Policies express “strategic intents” aiming at achieving a common good and improved situation for stakeholders. Noteworthy are the Millennium Declaration and the MDGs which are milestones for mankind. They are bold, inspirational, measurable strategic intents expressing wellbeing targets articulated for the first time at a global level with commitment from all nations. Quite an impressive achievement in themselves!

Strategic intent is foundational to country-led monitoring and evaluation systems

A strategy is an alternative chosen to make a desired future happen, such as achievement of a goal or solution to a problem. Management is the organization and coordination of the activities of an

6 Quote from the *Webster’s Encyclopedic Unabridged Dictionary of the English Language*, Portland House, New York, 1997.

enterprise in accordance with certain policies and in achievement of clearly defined objectives. Monitoring is the supervising of activities in progress to ensure they are on-course and on schedule in meeting the objectives and performance targets. Evaluation is the rigorous analysis of completed or ongoing activities to determine the extent to which intended and unintended results are being achieved. Evaluation supports evidence-based decision making and management accountability by examining rationale; relevancy; effectiveness; efficiency; coherence; sustainability; and, connectedness. These definitions⁷ point toward the evidence that all starts with a clear enunciation of the strategic intent.

The first requirement for the soundness of any country-led monitoring and evaluation system is its alignment with the strategic intent of the intervention⁸. A requisite for any relevant statement of strategic intent is the evidence of a sound diagnosis of the existing situation and identification of the key factors at play and the SMART articulation of the intended changes sought. A proper monitoring framework will translate the strategic intent with its implementation goals into a coherent set of performance measures covering both the internal logic and the externalities of the systemic approach pursued.

The main challenge of any country-led monitoring system is to be simple and manageable. The current propensity is the facile approach of identifying many (too many) performance indicators, too many of them. This leads to confusion concerning what is important and even the possible erroneous belief that the achievement of indicators leads to the fulfilment of the strategic intent. A strategic intent implies substantive thinking about what and how it is to be achieved. Ideally the scope of a monitoring system ought to be reduced to cover only the essential factors affecting the successful and effective implementation of the process that will yield the strategic intent of the intervention.

Monitoring systems often are too complex because many development actors are involved in the achievement of a strategic intent of an intervention. They often reflect the pressures from development actors to trace their respective attribution or contribution. This leads to an aggregation of indicators having to be tracked and reported on, instead of providing a systematic and systemic reporting system.

7 Definitions adapted from those found on the website of BusinessDictionary.com

8 Intervention means either a policy, strategy, institutional strengthening, programme, project, activity, task, product, service.

A country-led monitoring system should start from a sound diagnosis of the initial situation and track performance indicators that measure change viewed from a wholesome national perspective. The purpose of a country-led M&E system is to assess the extent to which there is evidence of a change of situation or behaviour. The focus is on the outcomes and impacts and processes producing them. Traditional externally supply-driven monitoring systems focus more on the outputs and attribution of particular funding sources. The new paradigm shift requires monitoring and evaluation systems to pass the fulcrum from the supply to the demand side. From a country perspective, one should be able to understand the strategic intents of the interventions together with their performance scorecards⁹ enabling easy tracking of progress and providing evidence for evaluation.

There is also a need of a paradigm shift concerning evaluation. A country-led evaluation system will first address the strategic intent of intervention, their rationale and relevance to improving the common good in conformity with national values and objectives. Country-led evaluation will look at external support as a contribution to national capacity strengthening. Evaluation will serve the purpose of assessing positive and negative effects and support rational decision-making. It will emphasize the complementarities of stakeholders' actions rather than crediting singular contributors. Evaluation will provide evidence to exercise an overall judgement of the worthiness of interventions and if possible, their opportunity costs.

9

In his article *The First Scorecard* of August 2006, Arthur Schneiderman demonstrates that it was developed in 1987 at Analog Devices. Robert Kaplan and David Norton publicized the scorecard approach in 1996, when they published *The Balanced Scorecard*, Harvard Press, Boston. Balanced scorecard is a tool to execute and monitor the organizational strategy by using a combination of financial and non financial measures. It is designed to translate vision and strategy into objectives and measures across four balanced perspectives: financial, customers, internal business process and learning and growth. It gives a framework ensuring that the strategy is translated into a coherent set of performance measures. Kaplan and Norton further articulated the implementation of the organizational strategy in their book *Strategy Maps*, published by Harvard Business School Publishing, Boston, 2004. A strategy map is a diagram that shows how an organisation creates value by connecting strategic objectives in explicit cause-and-effect relationship with each other in the four BSC objectives (financial, customer, processes, learning and growth).

Conclusion

The strategic intent implies an anticipated result that guides the planned actions. It requires concentration, commitment and stamina to see it through. It's all about thinking, living and acting intentionally. Intention and attention are inextricably linked. Clarifying the strategic intent focuses attention on what really matters to you. Desired changes begin at this point. Managing change is key to success, adapting to externalities and appropriating opportunities to propel forth the strategic scenario maximising achievement and minimising efforts.

In reading many academic writings, it has become clear that even scholars have difficulty in capturing in words the fullness of the concepts of strategic intent and what happens in real life. At the risk of being as guilty of the same over-simplifications, I dare summarise by saying that the driver steps¹⁰ of successful achievement of strategic intents are:

- translate the strategic intent into SMART strategies;
- align the strategic drivers of the operational plans with the strategic intent;
- simplify the monitoring system to the track only the nevralgic factors;
- make sure that the achievement of the strategic intent mobilises everyone;
- make the realisation of the strategic intent an unfolding continual process;
- build on opportunities and promote change through leadership;
- when facing uncertainties, evaluate and learn-by-doing, and most of all
- believe and live the strategic intent.

An effective monitoring system for assessing the achievement of the strategic intent will entail these essential features:

- the intentionality of tracking strategic results;
- the systematic translation of the strategic intent and its delivery process into key performance indicators;

10 Adaptation from Robert Kaplan and David Norton *The Strategy Focussed Organisation*, Harvard Business School Publishing Corporation, Boston, 2001.

- a discernment for the nevralgic factors which influence systemically the logic model of the intervention;
- data collection and metrics of the performance of those key indicators.

An evaluation system focussed on the strategic intent will enable a judgement on the intended and unintended, positive and negative effects of the results achieved. Its prime contribution is to provide feedback and learning about the rationale, relevancy and effectiveness. It will avoid being blurred by detailed process considerations. Such an evaluation system views the intervention from a Merlin perspective and starts with the end-result as the starting point.

Hamel & Prahalad wrote: "If the strategic architecture is the brain, the strategic intent is the heart. It should convey a sense of stretch – current resources and capabilities are not sufficient for the task." Like the old sayings: "When there's a will, there is way." and "Nothing is difficult if you love what you do." In other words, the strategic drivers are purpose and passion.

When you are clear about the way to be, and living in tune with your intentions, not only will your leadership be better, but you will experience a greater sense of wellbeing. In the context of a country-led monitoring and evaluation system, it helps to adopt an indigenous perspective of reality when assessing the nevralgic effects of external support to development.

References

- FRANK, Greif and Associates. *Strategic Intent.com*. Available at: <http://www.strategicintent.com/home/>
- GARDINER, W. L., and, AVOLIO B. J. (1998). *The charismatic relationship : a dramaturgical perspective*. In *Academy of Management Review*.
- HAMEL, G. and PRAHALAD, C.K.. *Competing for the Future*, Harvard Business School Press, Boston.
- JARZABKOWSKI, Paula. (2005). *Strategy as Practice: An Activity-Based View*, Sage, London.
- KAPLAN, Robert and NORTON, David. (2004). *Strategy Maps*. Harvard Business School Publishing, Boston.
- KAPLAN, Robert and NORTON, David. (1996). *The Balanced Scorecard*. Harvard Press, Boston.
- LOVELL, Pamela and KELLY, Julie. *Strategic Intent Com Australia*. Available at: <http://www.strategicintent.com.au/content/index.php/site/home/>

- MALI, Paul. (1972). *Managing by Objectives*, John Wiley & Sons, New York.
- MANTERE, Saku and Sillince, John. (2007). *Strategic Intent as a rhetorical device*. In Scandinavian Journal of Management. Elsevier.
- MANTERE, Saku and Sillince, John A.A. (2006). *The Social Construction of Strategic Intent*. In www.tuta.hut.fi/library/working_paper/pdf/mantere-sillince-strategic-intent.pdf
- MELISSA, Kelly-McCabe. (2007). *Clear Intent Strategy Inc*. Available at: <http://www.clearintentstrategy.com/index.htm>
- ODIORNE, George S. (1981). *MBO and Strategic Planning*. In Management Handbook. USA.
- OECD. *Evaluation and Results Based Management* OECD Publications, France.
- PRAHALAD, C.K. and Hamel, G. (1989). *Strategic Intent*. In Harvard Business Review, Boston.
- ROGERS, Everett M. (2003). *Diffusion of innovations*, New York.
- SEARLE, John R. (1994). *Intentionality: an essay in the philosophy of mind*, Cambridge University Press.
- SENGE, Peter. (1990). *The Fifth Discipline*, Bantam Doubleday Dell Publishing Group, Inc.
- SMITH, Charles E. (1994). *The Merlin Factor: Leadership and Strategic Intent*. In Business Strategy Review. Oxford University Press.
- SPANY, Andrew. *What Really Matters*. In Spany International (Website)
- SPANY, Andrew. (2003). *Strategic Achievements*. In Industrial Engineer.
- WHITE, T. H. (1958). *The Once and Future King*. UK.

SUPPORTING PARTNER COUNTRY OWNERSHIP AND CAPACITY IN DEVELOPMENT EVALUATION.

THE OECD DAC EVALUATION NETWORK

*Hans Lundgren, Head of Evaluation Section,
Development Co-operation Directorate, OECD*
Megan Kennedy, Consultant, OECD

Introduction

In the context of ongoing implementation of the Paris Declaration on Aid Effectiveness and a growing desire to improve development outcomes through better aid management and mutual accountability for results, donors and partners are working together to cultivate partner-led development evaluation. The OECD's Development Assistance Committee Network on Development Evaluation is a leading international forum where evaluation managers and specialists from donor nations and multilateral organisations come together to co-ordinate and improve the evaluation of international development assistance. Their efforts take place in a context where more emphasis has been placed on what works in development, what doesn't and why, and on appropriate methods to assess results and impact. This article provides an overview of the Network's efforts to enhance partner country ownership of development evaluation.

Evaluation refers to the process of determining the worth or significance of an intervention. "Development evaluation" is the systematic and objective assessment of an on-going or completed development project, programme or policy, its design, implementation and results. In this article the term is used primarily for evaluation of activities classified as official development assistance (ODA), and can include programmes and projects implemented by non-governmental organizations, partner governments or external partners in developing countries.

Evaluation of international development co-operation should facilitate learning, inform decision-making processes of both recipients and donors and increase accountability for the results of aid.

Evaluation can be carried out throughout the programme lifecycle. It includes, but is not limited to: ex-post; process; formative; summary; participatory; theory-based; and, impact evaluation. The ultimate goal of development evaluation is to contribute to improved development outcomes.

Evaluation is a cross-cutting capacity that reaches beyond the public sector. An evaluation system includes not just the production of evaluation reports, but also, policies, agenda setting, and the use and dissemination of results for accountability and/or learning purposes. It involves a diverse group of stakeholders: partner and donor governments; beneficiaries; civil society; implementing partners; programme staff; the general public and others.

The OECD DAC Network on Development Evaluation

The Development Assistance Committee (DAC) is the principal OECD body through which its member countries deal with development co-operation. Within the DAC, the Network on Development Evaluation brings together evaluation managers from development agencies and ministries of 23 OECD DAC members and 7 multilateral organisations. Its mission is to increase the effectiveness of development policies and programmes by supporting high quality, independent evaluation of aid. The efforts of the DAC Network on Development Evaluation provide an apt framework for considering current donor efforts to facilitate partner-led evaluation systems.

Supported by a small secretariat based in Paris, the Network focuses on improving the quality and co-ordination of development evaluation. To this end, the Network develops evaluation guidance for practical use, facilitates donor co-ordination, supports evaluation capacity development, and improves knowledge sharing through an online evaluation resource centre called DEREc – which presents member evaluation reports and other development evaluation resources.¹

In the context of new assistance strategies, political commitments to scale-up aid and the push for improved aid effectiveness based on mutual accountability for results, donors are working to strengthen their own evaluation functions. At the same time, they are recognising the pressing need to strengthen the evaluation

1 Visit the Development Evaluation Resource Centre DEREc at: <http://www.oecd.org/dac/evaluationnetwork/derec>

function in partner countries. Efforts to promote partner-led evaluation are intensifying. These efforts are building on an emerging consensus regarding the need for partner-led development contained in the commitments of the Paris Declaration.

Why country-owned evaluation is needed

Though often subsumed within monitoring under public management, development evaluation has multiple functions. In a context where questions remain about the best ways to achieve development goals, evaluation provides valuable information to improve development programmes. Evaluation also serves a dual accountability function: by holding implementing partners accountable to funders for the use of development assistance and by holding donors and implementers accountable to the intended beneficiaries (and the wider global community), for development results. High quality evaluation can support the push for better results-focused management to achieve development goals, such as the Millennium Development Goals (MDGs).

Unfortunately, development evaluation and monitoring often take place only to satisfy external requirements. Such “donor-centric” evaluation perpetuates a control-focused view of the role of evaluation and tends to de-motivate those involved from the partner side. The resulting evaluations may be of little use to local decision-makers, staff and beneficiaries because the evaluation is designed to meet external funder needs. These needs may neglect key questions or accountability important to other stakeholders. Low partner buy-in can also result in limited use of findings. Partners and beneficiaries can often provide relevant and useful information and perspectives including on which programmes or projects need to be evaluated and what core evaluation questions need to be asked.

Furthermore, partner ownership is critical to build the sustainability of evaluation systems, and can ensure that the evaluation agenda meets locally defined evaluation needs. Finally, independent, high quality evaluation is important beyond international development co-operation programmes since there are accountability and information needs to be met throughout the public sector.

Evaluation of the aid effectiveness agenda

Development evaluation has evolved along with changes in aid modalities and the development environment. Assessments of aid

have become more participatory (involving local stakeholders in donor evaluations), and are now increasingly joint and sometimes partner-led. The “aid effectiveness agenda” challenges donors and partners to improve the results of development co-operation. The Paris Declaration on Aid Effectiveness, endorsed in March 2005, by over one hundred ministers, heads of agencies and other senior officials, lays out an action-orientated roadmap intended to improve the quality of aid and its impact on development.² Each of the five pillars of the Paris Declaration – ownership, harmonisation, alignment, managing for development results, and mutual accountability – has important implications for the field of development evaluation. New forms of development assistance (such as basket funds, general budget support, regional programmes, etc) rely more on partner country systems – highlighting partner evaluation needs and capacity issues.

Moving beyond beneficiary and partner *participation* in donor-led evaluations is key. True ownership means beneficiary and partner *initiation* and decision-making power over evaluation agendas, processes and outputs. The push for partner-led evaluation has grown in the context of more aligned development co-operation approaches. In response, over the past two decades, the World Bank, the UN, the OECD DAC, and some donor and partner governments have been developing approaches to encourage partner-led evaluations. Donor headquarters are increasingly open to methodological and organisational changes in evaluation. This openness provides an opportunity to continue towards country-driven, co-ordinated and coherent evaluation that is useful both for country policy formulation and for accountability.

Strengthening partner evaluation capacity. The work of DAC donors

While partner country capacity is not synonymous with ownership, the two must go hand in hand. Capacity is now recognised as a “critical missing factor in current efforts to meet the MDGs,” and there is growing awareness of the critical link between partner evaluation capacity and the successful management of inter-

2 *Paris Declaration on Aid Effectiveness: Ownership, Harmonisation, Alignment, Results and Mutual Accountability*. March 2005, High-Level Forum on Aid Effectiveness. The Paris Declaration builds on agreements made at the International Conference on Financing for Development in Monterrey, Mexico, 2002 and the Managing for Development Results: Second international-Roundtable on Results, in Marrakech, February 2004.

national development programmes. As participatory approaches to development evaluation have become more common, capacity issues in beneficiary communities and partner countries have come to the fore.³ Capacity development is a key part of donor support for enhanced country ownership of evaluation.

Evaluation capacity is the ability of people and organisations to define and achieve their evaluation objectives.⁴ Capacity involves three interdependent levels: individual, organisational and the enabling environment. Evaluation capacity development (ECD) is understood as the process of unleashing, strengthening and maintaining evaluation capacity. ECD is a long-term change process, targeted in the context of strengthening capacity in related systems of management, accountability and learning. Demands for improved results have drawn attention to capacity gaps in donor and partner development agencies – leading to an explosion of interest in ECD.

ECD is a core element of the DAC Evaluation Network's work programme. A series of regional seminars were held in Africa, Asia and Latin America and the Caribbean in the 1990s. These joint efforts of the OECD DAC and the multilateral development banks, including the Inter-American Development Bank, aimed at promoting and strengthening evaluation capabilities in developing countries. Though there was wide commitment to improving capacity, and a good deal of consensus among partners, the resulting action plans gained little traction and did not result in significant improvements. These efforts, though unsuccessful in stimulating sustained capacity in development programmes, did raise awareness and demonstrated a growing consensus on the importance of ECD and the need for strategic prioritization of efforts. In this way, they laid the groundwork for later efforts.

The Schaumburg-Müller study on donor support to, and experiences with, ECD found extensive efforts underway in donor agencies. At a workshop on joint-evaluation, held in Nairobi in April 2005 in collaboration with developing country partners, the issue of capacity was raised in the context of enabling developing country stakeholders to take on a fuller role in joint-evaluations. One of the key recommendations from the workshop was that "developing country governments should be supported to build their institutional capacity

3 See for example, proceedings from the 6th and 7th Meetings of the DAC Network on Development Evaluation. Can be found under Meeting Documents on www.oecd.org/dac/evaluationnetwork.

4 Definitions used in this paragraph are drawn from OECD 2006.

for initiating and leading joint-evaluations [and]... all partners need to look at innovative ways of providing funding for aid recipients to build their evaluation capacity." Donors committed to continue expanding their ECD efforts.

At the Third International Roundtable on Managing for Development Results, held in Hanoi in February 2007, capacity issues were a key dimension in the discussions, underlining the importance of renewed and focused attention to the matter. A 2006 fact-finding study led by Japan for the DAC Evaluation Network, found that extensive ECD work continues. The study included 26 agencies, including 21 bilateral and 5 multilateral.

The agencies reported a total of 88 separate ECD interventions. Different modalities of support included training and scholarships (37); workshops (31); technical support to projects/programmes (18); financial support (18); joint-evaluations (22); dialogue at policy levels (10); and, other types (8). Interventions range from training parliamentarians how to effectively read and respond to an evaluation report; IT infrastructure for data collection systems; empowering beneficiaries to participate actively in assessing programme outcomes; to training programme managers to draft quality terms of reference. The diversity of interventions in this area is characteristic of both the multi-dimensional nature of capacity development work, and of the lack of a clear definition of what exactly constitutes capacity development (which leads to variation in donor reporting).

Many donors support international and in-country evaluation training programmes, such as IPDET which was created out of recognition of the lack of suitable training opportunities for development evaluators. The Shanghai International Programme on Development Evaluation Training (SHIPDET) was inaugurated in April 2007 and has also been supported by several donors.⁵

Several donors, in particular the regional development banks, have made support for evaluation organisations a priority in their capacity development work⁶. From a donor perspective, the recent growth in evaluation associations (such as IOCE, AfrEA, IDEAS and national

5 Over a 3-year period, SHIPDET will be held semi-annually with the spring program focusing on Chinese participants and the autumn program focusing on international participants from the Asia and Pacific region. The program is jointly sponsored by the Ministry of Finance of the People's Republic of China, the World Bank, the Asian Development Bank and the Asian Pacific Development and Finance Centre. IPDET website: "IPDET Worldwide." Accessed July 2008. <http://www.ipdet.org>

6 For more on the role of evaluation organisations see Segone M. and Ocampo A. (2006), "Creating and Developing Evaluation Organisations – Lessons learned from Africa, Americas, Asia, Australasia and Europe", IOCE.

organisations) is a positive step that brings hope for sounder, increasingly partner-led evaluations of development activities in the future. Experience has shown that evaluation associations play a critical role in strengthening and sustaining monitoring and evaluation capacity, providing opportunities for useful dialogue, interaction and learning⁷. National evaluation organisations can serve as learning hubs, offering training and resources, and supporting communities of individuals committed to evaluation and accountable governance. They can also help donor agencies identify potential evaluation partners in developing countries and beneficiary communities. Professional associations contribute to building an enabling environment for an evaluation culture.

Collaborating with evaluation associations. Support to the African Evaluation Association (AfrEA)

AfrEA was founded in 1999 in response to a growing demand for information sharing, advocacy and advanced capacity building in evaluation in Africa. Since the initial phase of the association, 33 local and international organisations have supported its activities, including 6 member countries of the DAC Network on Development Evaluation as well as the Network itself. At the AfrEA Conference in 2004, 25 local and international organisations provided financial and/or in-kind support and coordinated and hosted Conference sessions and strands. Most recently, at the 2007 Conference, the group placed growing emphasis on the evaluation capacity gaps in partner countries and the role of international partners in helping build sustainable capacity.

Source: AfrEA <http://www.afrea.org/> Adapted from MFA Japan and OECD "Fact-finding survey on evaluation capacity development (ECD) in partner countries." (2006)

Evaluation Capacity Development: lessons learned

An array of key lessons has emerged from ongoing donor ECD efforts. The 2006 ECD study compiled donor observations about what works well and what does not, providing a useful synthesis of experience based knowledge regarding ECD strategies. Donor assessments provide information on what factors contribute to successful (or less successful) evaluation capacity development. Many of these reports have been confirmed by the capacity development literature and independent evaluations of ECD activities.

7 See for instance, presentation "Evaluation networks contributions to the Impact Evaluation Initiatives," by Oumou Bah Tall, President International Organisation for Co-operation in Evaluation (IOCE) at the MES-IDEAS Workshop. Kuala Lumpur, 4 April 2008.

Experience has clearly demonstrated that a one “size-fits-all” approach is not appropriate in evaluation capacity development. It goes without saying that the institutional, organisational and individual capacities of developing country partners vary widely. ECD approaches must be tailored to fit the individual and institutional context at hand. Imported “standard” capacity packages (such as generic evaluation training manuals) may not meet the needs most relevant to stakeholders in a particular context. Strategic, locally developed, carefully tailored interventions are more likely to have a significant, sustainable impact. This is one reason why the availability of evaluation training opportunities *in-country* has been cited as being a significant factor contributing to the success of ECD activities.⁸ To ensure relevance, initiatives should be led by beneficiaries *from the outset*. Partners should take the “driving seat,” not just in needs assessment, but throughout the programme lifecycle, including identifying priorities, developing plans, and monitoring and evaluating ECD initiatives.

Donor and partner stakeholders have observed that the focus should not just be on doing *more* but doing *better* capacity development work. This means co-ordinated approaches which are partner-led, beneficiary owned, and address all three levels of capacity (enabling environment, individual and organisation). Dimensions of the evaluation system beyond individual skill building (in particular the demand for and use of evaluations), and the accountability environment in which evaluation takes place, require further attention.⁹ Co-ordination of ECD efforts is vital. It adds coherence and improves efficacy, especially when beneficiary and partner stakeholders actively shape the joint approach.

The use of a multi-layered approach which provides a strategic package of various interventions targeting the three capacity levels is particularly constructive. Such a strategic approach should involve both direct evaluation skill building and the necessary support systems to boost demand and use of evaluation. Partnerships with

-
- 8 Ministry of Foreign Affairs of Japan for the OECD DAC Network on Development Evaluation, “Fact-finding survey on evaluation capacity development (ECD) in partner countries.” (2006)
 - 9 For example, a 2004 evaluation of the International Program for Development Evaluation Training (IPDET) found that many participants met strong resistance from within their own agencies and institutions when they attempted to put into practice the evaluation training they had received out of country. The political and “cultural” dimensions of institutions were unaffected by trainings targeted at individuals, resulting in frustration and failure to use capacity that had been created. Jua, Management consulting Services: “Evaluation of the International Program for Development Evaluation Training.”(2004)

different agencies can be a particularly useful way to build such a strategic approach capable of addressing various points within the evaluation system simultaneously.

Donors and partners report that a high level of commitment to evaluation and understanding of the benefits of monitoring and evaluation, especially among top levels within the partner government, helps ensure that capacities are employed appropriately. Individual or organisational “champions” with a high level of commitment and position of power can be critical in generating momentum towards change.

The benefits of evaluation must be clear to convince staff and decision-makers of its usefulness and to shore-up commitment. Such buy-in also helps ensure that useful evaluation outputs are produced which will impact policy and programming decisions.

An early and visible “success”, such as a high quality evaluation which has a major policy impact perceived by stakeholders as meaningful, can be critical in building support in and around evaluation systems. Such successes raise the positive incentives for individuals to participate in evaluation and can increase individual demand for training and for other capacity development activities. The visibility of evaluation outputs helps improve the accountability environment making it more likely that quality evaluations will be produced *and used* consistently.

In short, donors have identified direct support for evaluation capacity development as a useful way to contribute to improved partner ownership of development evaluation. ECD remains a priority concern and an area for further learning. The ways donors choose to evaluate their own assistance programmes, and the support they provide for partner-directed and joint-evaluation efforts, also support partner ownership, and will be discussed in the following section.

Learning by doing. Monitoring and evaluation capacity development in Vietnam

The partnership of Vietnam and Australia in M&E capacity development in Vietnam provides some valuable illustrations of a successful, joint capacity development process. This bilateral partnership takes place in the context of a joint effort to harmonise bi- and multi-lateral donor work in Vietnam and align with the government's own policies and plans.

Joint reflection on evaluation training in Vietnam reveals several lessons, primarily, the importance of local stakeholder leadership and commitment. The most successful strategy is based on a "learning-by-doing" approach to adult education which builds individual skills and teamwork through actual field visits, data collection exercises and other hands-on evaluation activities. This process is rigorously monitored and new competencies tracked, to ensure a high level of skill attainment and long term flexibility to meet changing needs. Participants also highlighted the need to identify and support evaluation "champions," individuals who become promoters of the new evaluation culture, skills and tools they acquire. Communications technology, government ownership and institutional support, compliment individual and team skill building. Lastly, the Vietnam case reveals that externally supported ECD can have positive spill-over effects into other government departments beyond those involved directly in aid management.

Source: Cuong, Cao Manh and John Fargher, "Evaluation capacity development in Vietnam," room document for the OECD DAC Network on Development Evaluation, 6th meeting. (Paris, 27 – 28 June 2007) and Vietnam Australia Monitoring and Evaluation Strengthening Project (Phase II): "Case study of M&E capacity building in Vietnam." (December 2006)

Facilitating ownership through joint and partner-led evaluation approaches

A "joint-evaluation" is an evaluation conducted collaboratively by more than one agency. Joint-evaluation has been on the international development agenda since the early 1990s. Such collaborative approaches, be they between multiple donors, multiple partners or some combination of the two, are increasingly useful at a time when the international community is prioritising mutual responsibility for development outcomes and joint approaches to managing aid (such as basket funds, SWAPs, direct general budget, sector and programme support). Joint-evaluations can strengthen joint programme planning and implementation. Experience has shown that joint approaches can lead to greater understanding of *overall* cumulative impacts of various international development efforts. More inclusive evaluation processes can have direct capacity strengthening effects for participants

from both donor and partner agencies.¹⁰ The push for more joint-evaluation is also motivated by the desire to reduce the sometimes onerous burden on partner countries, in-country staff and beneficiaries of multiple single donor evaluation field visits, data requests, etc.

An example of the value added of joint-evaluation approaches is the 2006 multi-donor, multi-partner joint-evaluation of general budget support (GBS), which involved 24 aid agencies and covered support to 7 countries during a ten year period for an amount of nearly \$4 billion. Its purpose was to assess to what extent and under what circumstances GBS is relevant, efficient and effective for achieving sustainable impacts on poverty reduction and growth. The findings contributed significantly to the review of donor policy and operational guidance in this area. Part of the reason the evaluation had so much influence is that, in addition to being of high quality, it was carried out jointly, giving its findings more legitimacy and weight.

Joint-evaluations are also increasingly used as a means to promote partner ownership. The term was once used to refer almost exclusively to multi-donor evaluations, but joint-evaluations have become more inclusive over the past decade and involve a growing number of non-governmental and developing country partners. Joint approaches facilitate the matching of complimentary capacity, initiative and resources of local and external partners. Participants in joint-evaluations report that they can be useful in building individual skills as well as cultivating working relationship between and within agencies. Working together can help create shared understandings and strengthen learning to help create more relevant programmes and policies. Still, careful attention must be paid to evaluation agenda setting in joint contexts to ensure that evaluation programmes are not skewed towards donor priorities exclusively.

One example of an evaluation that nurtured meaningful partner ownership of the evaluation process is the recent Netherlands and China joint country-led evaluation of the Development and Environment Related Export Transactions (ORET/MILIEV) programme in China. The evaluation was based on a strong donor-recipient partnership. It was motivated by the shared recognition that the majority of evaluations of development aid programmes are led by donors and are carried out to meet donors' requirements and that more evaluations from the perspective of the partner country are needed.

10 Presentations and discussion at the DAC Network on Development Evaluation often highlight examples where useful learning took place in the context of a joint evaluation project, or underline areas where learning could have been facilitated better.

The two agencies set out to establish an appropriate governance structure to ensure joint responsibility throughout the entire evaluation process. The intention was to have the partner in the lead with the donor playing a support role.¹¹

As described in their joint presentation of lessons learned, the donor's role in this "first generation" country-led evaluation was one of "nurturing the country's demand and facilitating evaluation activities."

Participants felt that the biggest challenges came from differences in evaluation cultures and systems which required negotiation and sometimes time consuming co-ordination.¹² As should be expected in this type of experimental evaluation, the partners were faced with institutional and capacity limitations. Some of these were addressed as part of the process, through integrated ECD measures. By working together the partners were able to produce a high quality evaluation that contributed to learning and informed efforts to improve the programme's efficiency. The final report served as the basis for a dialogue between the governments on better targeting the programme to meet core development goals such as improving the situation of women, protecting the environment and targeting western China.

***Key messages from the Netherlands.
China Joint Evaluation of the ORET/MILIEV programme***

Joint reflection on the evaluation exercise and previous experience concluded that in order to improve partner ownership of joint evaluation work, it is important to:

- Ensure that the evaluation meets the needs of both countries.
- Build consensus to design and conduct evaluation (such as: identifying the key issues in writing TOR, choice of field study cases, etc.).
- Engage the key stakeholders in the design, conduct, and interpretation of evaluation.
- Develop capacity of partner countries to do *and* use evaluation.
- Ensure that evaluation is used appropriately.
- Enable partner countries to evaluate themselves rather than having evaluation "done to them."

Source: Joint presentation by the Netherlands and China to the 6th Meeting of the DAC Network on Development Evaluation. Paris, France. (June, 2007)

11 Joint presentation by the Netherlands and China to the 6th Meeting of the DAC Network on Development Evaluation. Paris, France, June, 2007.

12 These barriers have been confronted in multi-donor evaluations as well. See: "Joint Evaluations: Recent experiences, lessons learned and options for the future." (OECD DAC, 2005)

Concrete suggestions for building partner country ownership

The following practical suggestions on encouraging partner participation in and ownership of joint-evaluations have emerged from the experiences of network members. These and other suggestions on identifying partners for joint-evaluations and conducting joint work are outlined in detail in the “Guidance on managing joint-evaluations,” (OECD, 2006) produced by the DAC Network on Development Evaluation.

- Systematic, effective communication on planned evaluations is needed between partner countries, donor country offices, and donor headquarters evaluation units. This means sharing evaluation plans well in advance and being open to joint programming in the planning stages.
- Partner country authorities should be enabled to take ownership of co-ordinating the advance planning for joint-evaluations. This is an area where co-ordination within and between donor agencies (harmonisation) can assist partner stakeholders in assuming a leadership role.
- Donor agency managers should consider, *on a systematic basis*, whether each evaluation can be undertaken with partner country participation and efforts made to maximize participation when appropriate. Assessments of partner capacity should be based on evidence, not assumptions, and build on experience and working relationships.
- Partners should be brought on board before decisions are made on the ground rules, the terms of reference (TOR), and the selection of the evaluation team.
- The possibility of joint-evaluation and the capacity needs to achieve evaluation goals should be considered in the design stage of new projects and programmes. This facilitates timely start-up of the evaluation, and gathering of baseline data. Ownership involves more than participation of partners and beneficiaries in needs assessments or as informants for impact evaluations. Ownership must be encouraged and reinforced *throughout* the programme lifecycle.
- When several developing country partners are involved, efforts should be made to facilitate co-ordination of their inputs. Opportunities for south-south learning in particular should be

identified, and, whenever possible, facilitated and supported by donors as part of the joint-evaluation experience.

- Steering Committees should be encouraged to meet in partner countries to ease partner participation. Capacity enhancing benefits of visiting and meeting at other agencies should also be considered.
- Genuine ownership of joint-evaluations often remains with the donors rather than with the country partners because of the financing. Donor managers sometimes feel that because their agency is financing the evaluation they will be accountable for its quality and should therefore retain tight control over the process. To redress this imbalance donors and partner countries should develop and fund partner government budget needs for evaluations. Partner countries should be facilitated to contract at least some of the consultant evaluation team.

Donor support for partner evaluation systems must go beyond funding technical capacity building activities. Specifically, undertaking joint-evaluations can compliment ECD efforts, build more collaborative and transparent relationships, and encourage partner leadership in evaluation of aid. Ongoing work by the members of the OECD DAC Evaluation Network, and others, continue to improve and expand joint approaches through learning based on evaluation experience.

Partner-led joint evaluation in South Africa

The International Development Co-operation (IDC) directorate in the National Treasury has established a system of joint evaluations for assessments of the relevance, impact and success of different programmes of support. The aims are to ensure transparency, embed accountability, and deepen the knowledge development process to contribute to improving programmes of development support. The findings of the evaluations are used to inform Country Strategic Frameworks agreed between the IDC and the donors. A *Development Co-operation Report*, published in 2000, reviewed the effectiveness and impacts of development co-operation from 1994-1999 and gave recommendations for the future. New joint evaluation modalities with bilateral donors were developed. South Africa provides one of the more interesting examples of partner initiated evaluation of development co-operation.

Source: Adapted from: OECD DAC Network on Development Evaluation, "Guidance for Managing Joint Evaluations." (2006)

Alignment and harmonisation: the role of improved co-ordination

In addition to facilitating ECD and undertaking joint and partner-led evaluations, donors' efforts to align and harmonise development assistance also make a contribution to strengthening ownership. In the Paris Declaration, donors and partners committed to synchronise development co-operation (including evaluation) with the development plans and strategies of partner countries. This includes efforts to direct more development assistance through partner systems, rather than creating parallel management structures. Better planning of evaluations, and involvement of partners and beneficiaries early on in evaluation programming, are needed to reach the goal of better alignment. However, this is an area where progress towards meeting commitments has been slow.

Harmonisation of donor evaluation works (meaning co-ordination of the various efforts of different external partners) can reduce the evaluation burden on developing country partners and facilitate alignment. Considerable progress is being made among DAC network members in this area. This has been achieved through more joint work and sharing of advance evaluation plans. The goal of sharing evaluation plans is to maximize opportunities for shared learning and coordination and minimize repetition of evaluation work. Harmonisation must be done carefully and paired with alignment to ensure that co-ordinated donors don't overwhelm the evaluation agenda to the detriment of partners or beneficiaries.

Producing international evaluation standards and resources

The DAC Evaluation Network produces and disseminates evaluation tools, guidance and standards as part of its regular work programme. Establishing international standards for development evaluation helps to create a shared basis for joint work. The norms and standards produced by the network also serve as a form of direct capacity development providing partners with resources to build their evaluation knowledge and take a more active role in setting and carrying out evaluations.

For example, the draft DAC Evaluation quality standards (OECD DAC, 2006) were formed and agreed upon through a participatory process that engaged partner country evaluators, members and

non-network members from a variety of development agencies.¹³ They therefore represent an emerging international consensus on key standards for evaluation of development co-operation. This short document outlines core elements of a quality evaluation process and product, such as the criteria to be used in evaluation and the format evaluation products should take. Other examples of internationally distributed evaluation resources include the DAC Evaluation Principles and guidance on joint, humanitarian, conflict prevention and peace-building and country programme evaluations.¹⁴

The DAC Glossary of key terms in evaluation and results based management (OECD, 2002) was first printed in English, French and Spanish is now available in thirteen languages. The high demand for this document demonstrates the strong demand for evaluation and management resources coming both directly from partner countries and from donor staff engaged in joint work.

Issues to consider: challenges and opportunities for improved ownership of development evaluation systems

This article has explored partner ownership from a donor development evaluation perspective, highlighting the links with evaluation capacity, and the roles of partner-led and joint-evaluations, alignment and harmonisation and the development of international norms and standards in increasing partner ownership. Several issues regarding ownership, capacity and the aid relationship merit further discussion.

Simultaneously meeting donors', beneficiaries' and partners' evaluation needs remains a challenge. Joint approaches and the transition to partner-led development evaluation raise the question of how to meet, most effectively, the sometimes divergent accountability and learning needs of donors, beneficiaries and partners. Partners must own development processes, including evaluations of development co-operation. Yet external partners and developing country governments also have evaluation needs when it comes to understanding and assessing the results of ODA. Evaluation needs

13 The draft standards are currently being applied for a test phase of three years and will be finalised in 2009.

14 For a complete list of documents and guidance pieces from the OECD DAC Network on Development Evaluation visit "Publications, Documents and Guidance" at: www.oecd.org/dac/evaluationnetwork

vary both across and among these groups. To meet these multiple needs, with: the least evaluation burden; lowest co-ordination cost; greatest contribution to development knowledge; highest levels of mutual accountability to funders and beneficiaries; and, maximum capacity building effects is a challenge. More experience needs to be acquired and explored through practical experience with joint-evaluations, and intensified efforts to follow through on commitments to ownership and mutual accountability.

Co-ordinated donor efforts need to link better with partner priorities and information needs. Joint and co-ordinated evaluation work needs to be mindful and take into account its effects on local evaluation systems and on evaluation capacity. Evaluations should, when feasible, look at relevance and impact not only in terms of donor requirements but also be based on the partner country's priorities and beneficiary interests.

Partner's monitoring and evaluation systems must serve purposes beyond aid evaluation. Capacity development and institution building efforts need to keep the wider partner governance context in mind. The institutional position of aid evaluation should balance independence and learning and be integrated into partner governance and management systems as much as possible. Development evaluation should also take into consideration stakeholders (especially civil society and the intended beneficiaries of development assistance) outside the government. The goal is an evaluation system that meets the needs of the partner, not one that is effective only in assessing the use of donor funds. To achieve this, evaluation system development must be led by partners, but donors can play a facilitating and supportive role by mobilising resources and co-ordinating their own work to increase capacity, to strengthen organisations and, to improve the accountability environment.

Citizen voice and accountability are still limited. Citizens of donor countries rarely see, and almost never directly experience the results of the development co-operation they fund. At the same time, citizens of developing nations, who directly experience the results (or lack thereof) of development spending often have minimal say in the allocation and programming of external funds. Weak or opaque governance systems can compound this "principle-agent" problem and highlight the importance of using evaluation to provide relevant, reliable information to all stakeholders. It also highlights the need to look beyond official evaluation units or divisions, to the *overall* governance and accountability systems of donor and

beneficiary countries. Even where the capacity to carry out quality evaluation is high, there will be little incentive to employ those capacities if participation and accountability remain weak. Though evaluation is just one piece of the development co-operation puzzle it might serve as a “hook” or focus point for strengthening governance and local ownership of development processes.

Conclusion

Partner-led evaluation can contribute to improving development results. High quality, independent evaluation reinforces accountability systems within and between donor and partner countries. Evaluation of development processes involves a cross cutting set of skills and enabling factors – from the individual and organisational level to the accountability environment. Efforts to support partner leadership in development evaluation should focus on strategic capacity development and co-ordinated, joint approaches to evaluation of development co-operation programmes.

True ownership will in most cases require not only much stronger capacity on both sides, but also a shift in the balance of evaluation power. A way to support such a shift is to enable more systematic and critical partner assessments of donor contributions to development goals, as set by partners themselves. A quality, independent, partner-owned evaluation system is an indication of the relative success of overall efforts to increase ownership of aid management and improve transparency and public accountability.

References

Breier, Horst (2005), *Joint-evaluations: recent experiences, lessons learned and options for the future*. OECD DAC Network on Development Evaluation.

Chinese National Centre for Science and Technology Evaluation and the Policy and Operations Evaluation Department of the Netherlands Ministry of Foreign Affairs. (2006), *Country-led Joint-evaluation of ORET/MILIEV Programme in China*.

Hauge, Arild O. (2003), *The development of monitoring and evaluation capacities to improve government performance in Uganda*, Evaluation Capacity Development Working Paper Series, no. 10, World Bank.

IADB and OECD DAC. (1993), *Regional Seminar on Monitoring and Evaluation in Latin America and the Caribbean: strengthening evaluation capabilities for sustainable development*.

Liverani, Andrea and Lundgren Hans. (2007), *Evaluation systems in development aid agencies. An analysis of DAC peer review 1996-2004*, In: Evaluation, vol. 13(2): 241-256.

OECD. (2006), *The Challenge of Capacity Development. Working towards good practice.*

OECD DAC Network on Development Evaluation. (2006), *Guidance for Managing Joint Evaluations.*

OECD DAC Network on Development Evaluation. (2005), *Joint Evaluations: Recent experiences, lessons learned and options for the future.*

OECD DAC Network on Development Evaluation. (2002), *Glossary of Key Terms in Evaluation and Results Based Management.*

Schaumburg-Müller, Henrik. (1996), *Evaluation Capacity Building – Donor support and experiences.* OECD DAC Expert Group on Aid Evaluation.

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

Wood, B. Kabell, D. Sagasti, F. and Muwanga, N. (2008), *Synthesis report on the first phase of the evaluation of the implementation of the Paris Declaration.* Copenhagen.

COUNTRY-LED EVALUATION. LEARNING FROM EXPERIENCE¹

*Oswaldo Feinstein, Professor at the Master in Evaluation,
Complutense University, Madrid. Former manager,
Operations Evaluation Department, the World Bank*

This chapter starts from a country-led evaluation (CLE) experience, continues with a discussion on the approach, and proposes a wider approach, shifting the focus from a specific type of evaluation to “country based evaluation systems” (CLES) which generate country-led evaluation as products. It shows that this latter approach has already been fruitful.

Experience in Mozambique with a CLE

At the end of the 1990's, and inspired by Robert Picciotto (at that time Director-General, Operations Evaluation Department, the World Bank), efforts were made to design and carry out country-led evaluations (CLE). The evaluation department of the World Bank, jointly with UNDP's office, with the support of the evaluation department of the Dutch Ministry of Foreign Affairs (IOB), discussed an approach to CLE at the Working Party on Aid Evaluation of the OECD. It was believed that CLEs would promote ownership by partner or “recipient” countries, and therefore greater use of the evaluations, which would thus enhance the value of evaluations.

The proposed approach was to launch a mission to Mozambique with representatives from the three organizations mentioned above, so that they would discuss with the government of Mozambique, and eventually with representatives from civil society, the possibility and interest of a CLE in Mozambique.

Thus, a CLE mission was launched and the Mozambican counterparts appeared to be very receptive to the idea. Given UNDP's Evaluation Office interest in relying on the UNDP office in Mozambique

1 This chapter is written from an “emic” (insider) perspective, given the involvement of its author in CLE work, and thus complements the presentation by Adrien and Jobin (2008). It develops and updates a presentation made by Oswaldo Feinstein at the workshop organized by the International Development Evaluation Association (IDEAS) in Prague, 2006.

as the focal point for the CLE initiative in Mozambique, the mission proceeded accordingly, and a senior Mozambican UNDP official became the key counterpart of the mission. It is to be noted that both the Netherlands's Embassy and the World Bank representative in Mozambique were also very supportive of the CLE.

The identification of a government or civil society "champion" to play a leading role in the CLE is strategic. In the case of Mozambique, the CLE mission identified a Ministry that was expected to play that role, but it turned out that the Ministry had great difficulty in mobilizing other government units which could have a solid interest in a CLE. Nevertheless, the mission gained interest and support in Mozambique for an evaluation workshop in which the CLE concept would be presented, offering a platform to elicit interest in CLEs from government and civil society representatives.

That workshop was held in Maputo, in 2002. The Minister of Health delivered a keynote speech and expressed interest at that workshop in a CLE of health programs in Mozambique. The seminar was also used as an opportunity to promote evaluation capacity development in Mozambique and to facilitate internal country evaluation networking. The workshop was followed-up with monitoring and evaluation diagnostic work and effort, supported by the World Bank, to develop a country based monitoring and evaluation system for the poverty reduction strategy (PRSP).²

Rationale for the CLE and a vision

The rationale for the approach was developed in a note drafted by the World Bank and jointly submitted by Dutch Ministry of Foreign Affairs, World Bank and UNDP in March 2003 after the formal session of the DAC Working Party on Aid Evaluation.

The argument developed in that note was as follows:

The fact that most evaluations of development aid have been led by donors and were done to satisfy donors' requirements had at least two significant consequences: lack of country ownership of these evaluations and a proliferation of donor evaluations leading to high transaction costs for the countries.

2 Several aspects of this experience have been presented by one of the resource persons of that workshop, Aderito Sánchez, in <http://unpan1.un.org/intradoc/groups/public/documents/CLAD/clad0043712.pdf> (the other two resource persons were Rogerio Pinto and Osvaldo Feinstein, the latter being the team leader)

On the other hand, as development assistance is moving towards a policy-oriented programmatic, country-led approach, it is also worthwhile promoting country-led evaluations which will assess the new modalities of development aid and also increase country ownership (and therefore usefulness), of evaluations, reducing at the same time the countries' transaction costs associated with evaluations.

However, so far experiences with CLEs have been mixed if not disappointing. IOB, OED/WB and EO/UNDP offered to support independent country-led evaluations in a number of partner countries. A link to the PRSP process has been explored in 2001 with a selection of partner countries. However, these countries gave priority to monitoring.

The mixed results can perhaps be explained through discussion of various aspects. One element is that the drive towards ownership is partly supply-driven, as is the case with PRSPs in general. A second element is that evaluation as an instrument of learning in current management theories (as in Results Based Management) is often downplayed vis-à-vis monitoring. This is visible in most PRSPs. A third element may be the perceived risk on the side of partner countries that independent evaluations of donor support may have political and financial consequences. A heavy aid dependency could translate into a reluctance to evaluate the role of donors independently. A fourth and perhaps crucial element is that the offer of support was not integrated into the policy cycles of PRSPs, Consultative Groups, Round Tables and other regular mechanisms of interactions between donors and partner countries. A fifth element had to do with the time frame: starting up a process towards a country-led evaluation requires much more time than expected because of the necessary internal negotiations between ministries, actors, evaluators and so on.

The challenge for the future is to focus attention on the crucial role of independent evaluation in development for learning purposes and to provide a basis for accountability. This role of evaluation has been recognized in donor policies and programs and is enshrined in the DAC Principles on Evaluation of Development Assistance. There is no similar recognition in, for example, the PRSP framework and in current discussions on results based management in development. This recognition may provide a more solid basis to overcome the obstacles as mentioned in the previous point. The next challenge is then for country-led evaluations to be incorporated in these policy processes.

Furthermore, CLEs require evaluation capacities at the country level. At the same time, a crucial way to develop these capacities is through “learning by doing”. Suitable training and technical assistance can serve as catalysts in the process of developing evaluation capacities. However, actual opportunities to use these capacities, such as those that can be provided by a CLE, play a crucial role both in mobilizing these capacities and in ensuring their sustainability. Involving nationals (mobilizing existing national capacity) in the evaluation of external assistance projects is one of the ways to start off the process of learning by doing.

In addition, it should be noted that CLEs are “country” led, i.e., not led by the donors, nor exclusively by government. Also civil society can lead the CLE process and/or it could play a key role in evaluating the performance of the public services through different means which can allow them to articulate their voice. The donors could still play a role, particularly in the “first generation” of CLEs, by nurturing the country’s demand for this type of evaluation (for example, through brainstorming sessions and/or workshops and also by asking for mutual evaluation under the ownership of the country concerned).

Countries could lead the evaluation by determining which evaluations will be done, steering and managing them. In some cases the evaluations could be contracted out by a governmental and/or civil society organization. Some donors may be able and willing to contribute to set up a fund that countries could use to pay for these evaluations (a “country-led evaluation fund”, CLEF).

The CLEs could range from evaluations at the project level to sector and country level evaluations. The latter would evaluate development aid in the country from the country’s perspective. It could be preceded by evaluations at sector level (country sector evaluations), which could use project evaluations as building blocks, promoting also the development of self evaluation by public agencies.

The note concluded with the formulation of a vision that could guide the CLEs: to develop a Country-led Evaluation System (CLES) that at a later stage will be able to produce evaluations useful for the country and the donors, based on evaluation capacities developed at the country level, with high country ownership of the evaluations and with low transactions costs for the countries and for the donors. This system could also play a key role in the evaluation of all national development efforts, whichever the source of their funds. Donors could periodically assess the quality of country-led evalu-

ations and could use CLE results as an important source for their own evaluation needs³.

An initial reaction to this approach was that it implied a contradiction, as it was in a way a donor-induced country-led evaluation approach. However, the argument was made that, in an initial phase, there was a need for a sort of demand induced CLE which could establish a “proof of the concept”. Then, at a later stage, there would be no need for such an inducement. However, in an initial phase the inducement could be needed in order to “awaken” the “latent demand” for CLEs.

Opportunities, achievements and lessons learned

It should be noted that though the emphasis was initially on country-led evaluations, for some of those that were involved in this experience as well as in other evaluation ventures, it became clear that it makes more sense to focus at a higher level, moving from the level of single evaluations to evaluation systems (see above,). The difficulty in fully grasping the importance of this shift becomes apparent in its neglect in a recent note based on a set of CLE regional workshops, where no reference is made to system level. This is despite it being the focus of one of the keynote presentations of the Prague workshop⁴ (which is quoted several times in that contribution).

Furthermore, it is worth observing that the focus on country-led evaluation systems is fully compatible with the Paris Declaration emphasis on country based systems⁵. Although generally difficult to find good examples of CLEs, there are some remarkable cases of CLES, particularly in Latin America, where three country cases can be highlighted: Chile, Colombia and México. In these countries the CLES have yielded multiple CLEs (the Chilean case has been considered a “factory of evaluations”)⁶.

3 As will be seen below, this “vision” has started to become a reality. See, for example, Rojas et.al. (2005) and Cunill Grau & Ospina Bozzi (2008).

4 See Adrien & Jobin (2008).

5 The text of the Paris Declaration can be found in <http://www.oecd.org/dataoecd/11/41/34428351.pdf>, whereas an evaluation of its implementation is provided in Wood et. al. (2008)

6 These cases have been documented for example in Cunill Grau & Ospina Bozzi (2008), Rojas et. al. (2005) y Feinstein & Hernández (2008)

Lessons from the CLE experience

Finally, the following lessons can be drawn from the CLE experience:

- The “entry point” for CLEs & CLES efforts is crucial. This can be:
 - a) country selection
 - b) selection of institutions
- Focus on systems that promote the generation and use of country-led evaluation systems (CLES) rather than on conducting CLEs
- Take into account CLEs and CLES experiences in different regions
- CLES (and CLEs) require a time frame of several years.

References

Adrien, Marie-Helene and Jobin, Dennis (2008), *Country-Led Evaluation*. In Segone (2008) Bridging the gap. The role of monitoring and evaluation in evidence-based policy making. UNICEF CEE/CIS Regional Office, Switzerland.

Cunill Grau, Nuria & Ospina Bozzi, Sonia (2008), *Fortalecimiento de los sistemas de monitoreo y evaluación (M&E) en América Latina*, Caracas: Banco Mundial, CLAD.

Feinstein, Osvaldo & Hernández Licona, Gonzalo (2008), *The Role of Evaluation in Mexico: Achievement, Challenges and Opportunities*, Mexico: SHCP & World Bank <http://siteresources.worldbank.org/MEXICOEXTN/Resources/MXNewsletter-QualityofPubExpenditure-No2.pdf>

Hyden, Goran (2008), *After the Paris Declaration: Taking on the Issue of Power* Development Policy Review, 2008, 26 (3): 259-274.

Rojas, Fernando; Mackay, Keith; Matsuda, Yasuhiko; Shepherd, Geoffrey; del Villar, Azul; Zaltsman, Ariel & Krause, Phillip (2005), *Chile: Study of Evaluation Program*, Washington DC: World Bank.

Segone, Marco, ed., (2008), *Bridging the gap. The role of monitoring and evaluation in evidence-based policy making*. UNICEF CEE/CIS Regional Office, Switzerland.

Wood, Bernard ; Kabell, Dorte; Muwanga, Nansozi & Sagasti, Francisco (2008), *Evaluation of the Implementation of the Paris Declaration (Phase One: Synthesis Report)*, <http://www.oecd.org/dataoecd/19/9/40888983.pdf>

COUNTRY-LED IMPACT EVALUATION. A SURVEY OF DEVELOPMENT PRACTITIONERS

*Marie-Hélène Adrien, President, Universalia and
former President, IDEAS*

*Denis Jobin, Vice President, IDEAS, and Manager, Evaluation Unit,
National Crime Prevention Center, Public Safety, Canada*

Introduction

The International Development Evaluation Association (IDEAS) is dedicated to harmonizing and improving the ways in which development evaluation is conducted, including developing a common understanding of the concepts and methods which underpin the practice. As an association of development professionals, drawn largely from developing and emerging countries, IDEAS is committed to seeking the best ways to further its three-fold corporate mission of knowledge sharing, networking and capacity building.

On April 4, 2008, IDEAS held a workshop on impact evaluation and aid effectiveness in Kuala Lumpur, Malaysia, which was co-hosted by IDEAS and the Malaysian Evaluation Society (MES). The theme of the conference, "Evaluation under a Managing-for-Development Results Environment," and the topics discussed resonated with IDEAS' corporate mission and served several of its objectives. IDEAS presented the results of a survey of practitioners on impact evaluation at the workshop.

In the present paper, we will discuss the context of country-led evaluation (CLE), the concept of CLE and quality, IDEAS survey and conclusion.

Country-led evaluation: related concepts

The CLE took its shape within the context of paradigms change of aid delivery. Indeed, as reflect in the Monterey consensus, Millennium Development Goals and Paris Declaration, the role of developing countries move from recipient of aid to developing partners, which after demonstrating *good governance* capacity, is fully responsible of their development. The concept of *Good governance*

is increasingly used as more donors base their aid on the conditions and reforms that lead to it¹. In this context what does *good governance* mean? For the World Bank *Good governance* means "the manner in which power is exercised in the management of a country's economic and social resources for development"². One of the ideas that lead to *Good governance* is the fact that partnering developing countries enhances the ownership of their development, thus becoming country-led.

Good governance has many desirable characteristics: it is participatory, consensus-oriented, accountable, transparent, responsive, effective and efficient, equitable and inclusive, and follows the rule of law. It also has many benefits – it minimizes corruption, gives voice to the most vulnerable, and ensures that the views of all are taken into account in decision making. It is responsive to the present and future needs of society.³ But more importantly, *good governance* reduces a country's transaction costs.⁴ Several decades ago, North (1986) demonstrated the importance of transaction costs (TCs) in any economy and suggested that a country's successful economic performance can be attributed to an institutional structure that keeps its TCs low (North 1990). What are transaction costs? Transaction costs are sometimes referred to as the costs of distrust or the indirect costs of making an agreement. These are costs related with searching a partner, negotiating the terms of agreement with that partner (before the exchange) and enforcing or renegotiating a given agreement over time (after the agreement). From an institutional economic perspective, one can deduce that governance embrace all forms of economic organisation – from network to hierarchy - and a single purpose of governance is the minimisation of transaction costs (Williamson, 1975, 1985, 1991a & b); thus good in this context precisely means low transaction costs. *Good governance* is the one who keep transaction costs low. Then what are the links between *good governance* and CLE? This is the subject of next section.

1 Santiso, Carlos Good Governance and Aid Effectiveness: The World Bank and Conditionality The Georgetown Public Policy Review Volume 7 Number 1 Fall 2001, pp.1-22.

2 World Bank (1992) Governance and Development Washington, DC: The World Bank.

3 <http://www.unescap.org/pdd/prs/ProjectActivities/Ongoing/gg/governance.asp>

4 This annotated bibliography supports the evidence of causal links between governance and development. World Bank (2000) Reforming Public Institutions and Strengthening Governance: A World Bank Strategy, P. 179-185.

The Link between Country-led evaluation and Good governance

As we pointed out before, the field of development evaluation evolved considerably as demonstrated by the paradigm changes that occur over the last decades⁵. Indeed, the international development arena has contributed to broadening the scope and design of evaluation – from an earlier, narrower focus on projects to broader assessments that encompass policy, partnerships, and institutions, and the development of evaluation methodologies that deal with challenges faced in development aid.

At the same time and in parallel to these developments, there has been increasing pressure to make evaluation central to a country's own development process. The field of evaluation is continuously being reshaped by the evolving context of international aid, and particularly by the continuing recognition that effective development assistance requires country leadership and the capacity to exercise it.⁶ The Paris Declaration and Millennium Development Goals favour the development of national country-led evaluation practices in emphasizing the importance of ownership, alignment, harmonisation, managing for results, mutual accountability, and, *good governance* – which is perhaps the most important.

So what is the relationship between *good governance* and Country-Led Evaluation? Kaufmann distinguished six key dimensions of *good governance*⁷:

- Voice and accountability
- Government effectiveness
- The lack of regulatory burden
- The rule of law (which includes protection of property rights)
- Independence of the judiciary, and
- Control of corruption.

5 See Marie-Hélène Adrien; Jobin Denis 2007 'Country-Led Evaluation: Lessons Learned from Regions' in Bridging the gap: The Role of Monitoring & Evaluation in Evidence-based Policy making Ed. Segone, Marco, UNICEF , http://www.unicef.org/ceecis/evidence_based_policy_making.pdf

6 (...) donor agencies should "respect partner country leadership and help strengthen their capacity to exercise it." *Paris Declaration on Aid Effectiveness*, High Level Forum, Paris, February 28-29 2005, p. 2

7 Kaufmann, Kraay and Zoido-Lobaton (1999)

CLE directly impacts three of these six dimensions of good governance: voice, accountability, control of corruption.

Voice: CLE, which are consensus-oriented, provide the voice of partnering country recipients and respective beneficiaries of development efforts. CLE provide a powerful consultation tool in modern public management, as the process is participatory and the beneficiaries and users of CLE are consulted. By contributing to voice out the stakeholders and beneficiaries preferences, CLE enhance trust and transparency toward public institution which in turn reduce transaction costs.

Accountability: CLE contributes to transparent, responsive, and equitable governance by giving voice to the opinions and views of stakeholders who support a project, program or policy. It allows partnering countries to become more accountable for the performance of development interventions by generating knowledge about *what works* and *what does not work*, and proposing solutions to improve the delivery system, which in turn feed into better policy making. The evaluative information generated through CLE supports learning and improve decision-making, which is essential for more effective governments. Again, a positive impact in reducing transaction costs.

Corruption: CLE is a deterrent to corruption, as projects and programs under scrutiny are more likely to detect corruption than those that are not; thus improving performance and reducing transaction costs.

The relationship between CLE and *good governance* is clear: CLE reduces transaction costs, fosters trust in public institutions, a deterrent to corruption, and improves government effectiveness.

Impact Evaluations

What do we mean by impact evaluation?

The Development Assistance Committee (DAC) of the Organization for Economic Cooperation and Development (OECD) defines an impact as:

A positive or negative, primary or secondary long-term effect produced by a development intervention, directly or indirectly, intended or unintended.

The World Bank describes impact evaluation in the following way:

An impact evaluation assesses changes in the well-being of individuals, households, communities or firms that can be attributed to a particular project, program 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 program?”. Since we cannot observe this group both with and without the intervention, the key challenge is to develop a counterfactual – that is, a group that 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 program while removing confounding factors.

Impact evaluation is aimed at providing feedback to help improve the design of programs and policies. In addition to providing for improved accountability, impact evaluations are a tool for dynamic learning, allowing policymakers to improve ongoing programs and ultimately better allocate funds across programs. There are other types of program assessments including organizational reviews and process monitoring, but these do not estimate the magnitude of effects with clear causation. Such a causal analysis is essential for understanding the relative role of alternative interventions in reducing poverty.⁸

What do we mean by Quality?

It was noted at the IDEAS/MES workshop that despite a significant body of shared lessons learned and recent debates on impact evaluation, a fundamental question remains about the quality of impact evaluations. Unfortunately, there is no agreed definition of quality in this context. In the field of evaluation, quality is usually considered as the degree of compliance with evaluation standards.⁹ However, most evaluation standards are process-oriented, while a definition should be method free: does not favor a method but focus on the results produced by any given methods. For instance, one study¹⁰ defines evaluation quality as a way to minimize bias, of which there are four sources:

8 <http://go.worldbank.org/2DHMCRFFT2>

9 Schwartz R., Mayne J., eds, *Quality Matters, Seeking Confidence in Evaluating, Auditing and Performance Reporting*, Transaction Publishers, Rutgers, New Jersey.

10 David P. Farrington, *Methodological Quality Standards for Evaluation Research*, 2003

1. **Statistical Conclusion Validity** establishes whether the cause and effect variables are related. With this type of validity, one must ensure adequate sampling procedures, appropriate statistical tests, and reliable measurement procedures.
2. **Internal Validity** establishes whether the intervention was the reason for the outcome or whether the outcome would have occurred anyway.
3. **Construct Validity** establishes whether the theoretical assumptions behind a given intervention are sound and evidence based.
4. **External Validity** establishes whether there was a generalization of causal relationships across different persons, places, and times, and the operational definitions of interventions and outcomes. According to this definition, a quality impact evaluation must deal with, among other things, counterfactuals. This is essentially what is required by the external validity criterion, which has the effect of limiting the range of approaches or methods to those which are controlled by reference to comparison groups or through hypothetical comparisons (e.g. theory-based evaluations or longitudinal analysis)¹¹.

The authors would like to propose a definition of quality that focuses on the results of an evaluation rather than its methods. An evaluation of quality is determined by *“the joint ability that an evaluator will a) assess and b) report on the performance of an institutional arrangement by the product of its competence (ability to assess) and the product of its independence (ability of revealing)”*¹². This definition has the advantage of being method-free; what matters is the ability to assess an institutional arrangement, based on the competence and skills of the evaluator. Furthermore, since impact evaluation plays an accountability role, policy makers and ultimately the taxpayers would want to know what happened with the public monies committed to those programs and projects. The ability of evaluators to report without hindrance on the effectiveness (impact) – or lack of it – is the key to assessing evaluation quality matters.

11 Another paradigm is reflected in the work of Pawson and Tilley, who suggest that the “Realistic Evaluation” or the context, mechanism and outcomes (CMO) approach, should focus on the context of an intervention by asking, “What works, for whom and why?” .The CMO approach relaxes the requirement for external validity; and therefore provide an alternative and competing vision of what constitute (or not) good quality evaluation.

12 Adapted from DeAngelo, L., 1981, Auditor Size and Audit Quality, Journal of Accounting and Economics, 3. In Jobin, Denis, A Performance audit based approach to evaluation: An Agency theory perspective (Forthcoming).

The context for Impact evaluation

With the recent growing demand from development agencies and developing country governments to demonstrate the effectiveness of development expenditures, there is increased scrutiny of methodologies employed by the evaluation community when conducting impact evaluations. The debate centers on the problem of a selection bias that can often occur as a result of the evaluation's design. Generally speaking, it is the authors' view the evaluation community has not welcomed this debate and has instead been extremely protective in its initial reaction. However, there is still an opportunity for the evaluation community to play a role in shaping international initiatives in support of impact evaluation that are still being formed, such as Network of Networks for Impact Evaluation (NONIE) and International Initiative for Impact Evaluation (3ie).¹³ The effectiveness of impact evaluations can likely be enhanced if the development and evaluation communities move past technical deficiencies in methodologies and focus on quality impact evaluations and development policy.

Amid a growing demand for better evidence of development effectiveness, the Center for Global Development (CGD) organized a working group on closing the evaluation gap. The group's report, "*When will we ever learn? Improving lives through Impact evaluation*," noted:

For decades, development agencies have disbursed billions of dollars for programs aimed at improving living conditions and reducing poverty; developing countries themselves have spent hundreds of billions more. Yet the shocking fact is that we have relatively little knowledge about the net impact of most of these programs. In the absence of good evidence about what works, political influences dominate, and decisions about the level and type of spending are hard to challenge.

The report generated many responses in the evaluation community, including: a) development efforts have focused almost exclusively on the use of randomized control trials, with little recognition of their limitations; b) little has been done to recognize alternative methods or develop new methodologies better suited to the evaluation of complex interventions within complex systems; c) questions about the meaning of impact evaluation and its quality.

13 Howard White, "Making Impact Evaluation Matter" (April 2008)

The CGD report also provided a catalyst for several new initiatives on impact evaluation, including:

- **NONIE – Network of Networks for Impact Evaluation**

NONIE, a collaborative initiative formed in November 2006, is a network of networks comprising the DAC Evaluation Network, the United Nations Evaluation Group (UNEG), the Evaluation Cooperation Group (ECG), and a fourth network drawn from the regional evaluation associations. Its purpose is to foster a program of impact evaluation activities based on a common understanding of the meaning of impact evaluation and approaches to conducting impact evaluations. NONIE's objective is "to enhance development effectiveness by promoting useful, relevant and high quality IE."

- **3iE – International Initiative for Impact Evaluation**

3iE's aim is "encouraging the production and use of evidence from rigorous impact evaluations for policy decisions that improve social and economic development programs." 3iE complements NONIE's efforts by improving the impact evaluation of development programs.

IDEAS survey on Country-led impact evaluation practitioners

In this context, the authors conducted a web-based survey between mid-March and April 1, 2008. The objectives of the survey were: a) to understand the position of IDEAS' members with respect to impact evaluation issues, and b) to understand the evaluation community's position with respect to impact evaluation issues.

While the authors do not claim that the survey was scientific, which would have permitted the generalizing of findings with a comfortable degree of confidence, it nevertheless provided valuable insights into what evaluation practitioners think of the important issues. Indeed, several evaluation groups were surveyed and reached through discussion groups, including: Evaltalk, Xc-eval, IDEAS discussion group, the Mandé & News theory-based evaluation group, the Afrea discussion group, and MES members.¹⁴

14 <http://bama.ua.edu/archives/evaltalk.html> ; <http://groups.yahoo.com/group/IDEAS-Int/> ; http://groups.yahoo.com/group/Theory-Based_Evaluation/; <http://groups.yahoo.com/group/AfrEA>; <http://groups.yahoo.com/group/MandENEWS>; <http://groups.yahoo.com/group/XCeval>.

Survey results

More than 100 IDEAS members responded to the survey (a response rate of over 20 percent) and 246 non-IDEAS members of other evaluation groups responded. The survey provided interesting results, as it demonstrated the heterogeneous character of impact evaluation practices. While few significant differences between IDEAS and non-IDEAS members¹⁵ were found to exist, the wide range of methods and approaches used translate into differences of views with respect to impact evaluation (Table 1 and 2).

With respect to the range of significant and recurring obstacles encountered when conducting an impact evaluation, for both groups measurability came first. This is probably the main challenge associated with impact evaluations, and all the more so in the context of Country-led Impact Evaluations.

Table 1: Which evaluation methods are you most familiar with?

<i>Answers</i>	<i>IDEAS Members</i>	<i>Non-IDEAS members</i>
Performance Indicators / Performance Measurement	75.5%	69.1%
The Logical Framework Approach	83.3%	68.3%
Theory-based Evaluation	35.3%	31.3%
Formal Surveys	46.1%	59.3%
Rapid-Appraisal Methods	39.2%	29.7%
Participatory Methods	66.7%	58.9%
Public Expenditure Tracking Surveys	7.8%	4.1%
Cost-Benefit and Cost-Effectiveness Analysis	21.6%	17.1%
Other	10.8%	11.0%

15 AEA, Afrea, CES, EES, MES, are the main others sources of membership.

Table 2: What kinds of methods have you mostly used to conduct impact evaluations?

<i>Answers</i>	<i>IDEAS Members</i>	<i>Non-IDEAS members</i>
Two-Group Experimental Designs (experimental design)	16.5%	14.0%
Classifying Experimental Designs (experimental design)	5.2%	3.8%
Factorial Designs (experimental design)	7.2%	3.8%
Randomized Block Designs (experimental design)	9.3%	8.9%
Co-variance Designs (experimental design)	3.1%	4.3%
Hybrid Experimental Designs (experimental design)	5.2%	3.4%
The Non-equivalent Groups Design (experimental design)	8.2%	11.5%
The Regression-Discontinuity Design (quasi-experimental design)	5.2%	6.4%
The Proxy Pre-test Design (quasi-experimental design)	6.2%	7.7%
The Separate Pre-Post Samples Design (quasi-experimental design)	21.6%	33.6%
The Double Pre-test Design (quasi-experimental design)	2.1%	5.1%
The Switching Replications Design (quasi-experimental design)	2.1%	1.7%
The Non-equivalent Dependent Variables (NEDV) Design (quasi-experimental design)	3.1%	2.6%
The Regression Point Displacement (RPD) Design (quasi-experimental design)	3.1%	0.9%
Case Study (non-experimental design)	56.7%	53.2%
Qualitative Impact Evaluation Approach (physical causality; no counterfactual used)	55.7%	50.6%
Theory-based Evaluation	27.8%	23.0%

<i>Answers</i>	<i>IDEAS Members</i>	<i>Non-IDEAS members</i>
Not applicable	18.6%	17.0%
Other (please specify)	8.2%	8.5%

Table 3: What are the major obstacles you have encountered in conducting impact evaluations?

<i>Answers</i>	<i>IDEAS Members</i>	<i>Non-IDEAS members</i>
Technical issues (availability of respondents; translations; etc.)	28.0%	32.4%
Content issues (sampling; questionnaire design; codification; data analysis; data reliability; etc.)	23.0%	29.8%
Measurability issues (data accessibility, etc.)	53.0%	48.3%
Challenge in finding the appropriate set of skills for such assignments (statistical analysis, use of control groups, etc.)	25.0%	23.9%
Ethical issues	9.0%	11.8%
Cost limitations	42.0%	42.9%
Time limitations	50.0%	41.6%
Threats to independence (i.e. challenges with dissemination of the impact evaluation results)	18.0%	16.0%
Not applicable	16.0%	13.9%
Other	15.0%	13.9%

Table 4: Summary of Selected Survey Responses

<i>Questions</i>	<i>IDEAS Members</i>	<i>Non-IDEAS Members</i>
Familiar with both quantitative and qualitative approaches	65%	68.6%
Had experience with impact evaluations	75%	79%
Had never conducted any impact evaluation	20%	18%

Questions	IDEAS Members	Non-IDEAS Members
In terms of the evaluation gap, indicated that there is indeed a gap between the desired number of impact evaluations and the actual number that are carried out	57%	67%
With respect to what constitutes a good impact evaluation, indicated that counterfactuals were not essential in conducting a good impact evaluation	41%	32.6%
With respect to what constitutes a good impact evaluation, indicated that counterfactuals were essential	32%	33.1%
With respect to what constitutes a good impact evaluation, had no opinion	26%	34.3%

Country-led impact evaluations: some challenges

When it comes to conducting quality impact evaluations, the case studies presented in Kuala Lumpur (available on IDEAS web site at: www.IDEAS-Int.org) revealed that the challenges for development practitioners in developing countries are consistent with those generally associated with conducting impact evaluation¹⁶ and with those revealed in our survey, such as measurability problems and finding the right skills.

Country-led impact evaluation

A sample of the impact evaluations presented at the IDEAS/MES workshop provided country cases that are also good examples of country-led evaluations (CLE). Indeed, a CLE is considered an 'evaluation in which the country leads the evaluation by determining which evaluations will be done, and is responsible for steering and managing them.'¹⁷ Thus, impact evaluations carried out in this context are, as a matter of course, a type of CLE, which in this paper are referred to as Country-led Impact Evaluation (CLIE).

16 Bamberger, M., Rugh, J., Church, M., & Fort, L. (2003), Shoestring Evaluation: Designing Impact Evaluations under Time, Budget and Data Constraints. *American Journal of Evaluation* 2004; 25: 5-37.

17 Country-led evaluations. A discussion note prepared by WB/OED, UNDP/EO and IOB. March 2003

The cases presented were from developing or transition countries, including Azerbaijan, Romania, Trinidad and Tobago, Uganda, and Vietnam, although their shared experiences exhibited considerable variance between them. Indeed, with the Azerbaijan case, an evaluation was conducted in the irrigation sector using longitudinal data from an annual survey adopting quasi-experimental approaches. In Romania, the impact evaluation needed to clarify the context of the intervention as well as the relationship between the impact and the process. In Trinidad and Tobago, the authors understand that the prerequisites for rigorous impact evaluation have not yet been achieved (such as the incentive to use performance information, which has not been collected either nationally or from evaluations commissioned by donors). In Uganda, evaluating the National Agricultural Advisory Services using mixed methods created some challenges, such as the importance of external factors and institutional arrangements. Finally, for Vietnam, one main challenge of the community-based project impact evaluation was measuring the changes using both quantitative and qualitative indicators without any baseline data. The following cases illustrated the type of challenges one doing CLIE is dealing with.

One example comes from a 2007 independent evaluation of a four-year community based rural development project in northern **Vietnam's** Phu Tho province. The project employed a community-based approach to improve hygiene and nutrition, boost agricultural production, and enhance the capacities of local authorities and communities with a view to empowering them. Although the project had a logical framework, output and impact indicators were not clearly defined, and baseline data were not structured into a monitoring system with indicators.

To overcome the measurability challenge, the evaluation developed an innovative approach to appraising impact without any previously-established indicators. The methodology included: an assessment of beneficiary and stakeholder "perceptions of change" in livelihoods and the environment; a review of secondary sources (provincial statistical reports), project history and monitoring reports; and the collection of primary data through key informant interviews, community focus groups, and household surveys. Two important aspects of the design were translating qualitative perceptions into quantitative frequency analysis and funneling the quantitative results into proxies for impact assessment, and triangulating the results to compare perceptions from different groups.

While the evaluation was constrained by limited time and resources, and the need for skilled analysts, the evaluation process was rapid and cost-effective. The mix of methodologies was a practical solution for measuring impacts through quantitative proxies as well as qualitative analysis.

In concluding their presentation, the evaluators shared some project-related issues under discussion at the UNDP, including the question of the link between development effectiveness and impact evaluation, considerations of impact assessments on development, as well as concerns about accountability and how unintended consequences were being treated.

Romania is faced with numerous challenges in developing a national evaluation culture. The country's public administration needs to increase its capacity in results-based management and build monitoring and evaluation (M&E) systems. While there is a national evaluation strategy applied to structural funds, evaluation is at an embryonic stage in other policy areas. Interest in evaluation is growing, but demand is still low, there is confusion about its use as a management tool, and there are few experienced evaluators or professional networks. The Evaluation Facility – a project of the Evaluation Central Unit of the Romanian Ministry of Finance and Economy – is encouraging policy and decision makers to commission evaluations and support good management of evaluation exercises for developing a national evaluation culture.

The Interim Evaluation of the Strategy for the Decentralization of Pre-university Education in Romania examined the implementation of the strategy in a pilot group of three schools in three counties, and gathered counterfactual data from a control group of three other schools. It used a mix of formative and summative approaches. As a process evaluation, it assessed the implementation of the decentralization strategy in pilot schools, and was also intended to contribute to building a functional M&E system in the Ministry of Education. For this reason, it was suggested that it would have been more effective if the evaluation had been combined with an institutional/organizational evaluation.¹⁸ As an impact evaluation, it assessed the expected and unexpected effects, both positive and negative, of the decentralization. At the time of the presentation, the findings of the evaluation were still being consolidated.

18 Roxana Mihalache, "'Learnings' of impact evaluation in education policies in a developing evaluation culture – case of Romania" (April 2008).

The major lesson learned from the Romanian impact evaluation was that in the early stages of developing an evaluation culture and capacity, it is important to nurture the demand for evaluation, rather than insist on an ideal design that does not meet the expectations of the beneficiary. In a developing evaluation culture, such as Romania, impact evaluations cannot be addressed in the absence of a process evaluation.

Conclusion

While the recent attention and urgent debates on impact evaluation could either unify or divide the evaluation community, several key issues remain to be addressed, including:

- What do we mean by impact evaluation?
- Why carry out country-led impact evaluations (CLIE) in a context that suffers from a lack of capacity and funding?
- How to carry out quality CLIEs? What incentives are there in the development context?
- How to measure impacts in a context of poor data availability?

For the evaluation practitioner, these questions have several implications. While most stakeholders feel that more impact evaluations should be carried out, it is important to be able to directly attribute impact to an intervention, which requires both baseline information and implementation monitoring. Nevertheless, opinions are mixed on whether direct attribution requires counterfactuals.

This is an indication of the need for better understanding of what constitutes an impact evaluation and further capacity development in this area, which is aligned with the authors' findings with respect to Country-led evaluations in developing countries.

The IDEAS survey indicates that many evaluation practitioners agree with the CGD report on the evaluation gap. However, regardless of how interesting the ongoing debates on methods and approaches may be, they should not get in the way of other important discussions about setting standards and the need for M&E specialists to organize themselves to ensure and maximize the quality and credibility of their work.

Reference

Adrien, Marie-Hélène, and Denis, Jobin. (2008) "*Country-led evaluation: lessons learned from regions*". In: Segone, Marco: Bridging the gap. The rôle of monitoring & evaluation in evidence-based policy making. UNICEF.

Available at: http://www.unicef.org/ceecis/evidence_based_policy_making.pdf

DeAngelo, L. (1981) "*Auditor size and audit quality*", In: Journal of accounting and economics, Volume 3.

Farrington, David P. (2003) "*Methodological quality standards for evaluation research*," In: The Annals of the American Academy of Political and Social Science, 587.

Jalandhar Pradhan. (2008) *Challenges of monitoring and evaluating maternal and child health programme in developing countries*.

Available at: http://www.ideas-int.org/Documents/Jalandhar_Pradhan.doc

Jobin, Denis, *A Performance audit based approach to evaluation: An Agency theory perspective* (Forthcoming).

Khan, Alexa (2008) *Impact evaluation and development effectiveness: Are we jumping the gun in developing countries?*

Available at: http://www.ideas-int.org/Documents/Alexa_Khan.doc

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge: Cambridge University Press.

Mihalache, Roxana. (2008) '*Learnings*' of impact evaluation in education policies in a developing evaluation culture – case of Romania.

Available at: http://www.ideas-int.org/Documents/Roxana_Mihalache.doc

Mukhtov, Murad. (2008). *Impact evaluation of rehabilitation and completion of irrigation and drainage infrastructure project*.

Available at: http://www.ideas-int.org/Documents/Murad_Mukhtarov.doc

Santiso, Carlos. (2001) "*Good governance and aid effectiveness: The World Bank and conditionality*". In: The Georgetown public policy review, Volume 7, Number 1. Fall. pp.1-22.

Schwartz R and J.Mayne (eds.), *Quality matters, seeking confidence in evaluating, auditing and performance reporting*, Transaction Publishers, Rutgers, New Jersey.

Wallis, J., and D. C. North (1986). *Measuring the transactions sector in the american economy*, In Long term factors in American economic growth. Engerman, S. and Gallman, R. eds. University of Chicago Press.

Wellington Didibhuku Thwala, *Evaluation of public works employment creation programmes and projects in South Africa: thirty years of learning*.

Available at: http://www.ideas-int.org/Documents/Wellington_Didibhuku_Thwala.doc

White, Howard (2008). *Making impact evaluation matter*. Available at: http://www.ideas-int.org/Documents/Howard_White.doc

Williamson, O. E. (1991a), "*Strategizing, economizing and economic organization*". In Strategic management journal, vol:12. pp.75-94.

Williamson, O. E. (1991b), *Comparative economic organization: The analysis of discrete structural alternatives*. In *Administrative science quarterly*, 36, pp. 269-296.

Williamson, Oliver E. (1985), *The economic institutions of capitalism. firms, markets, relational contracting*. Free Press.

Williamson, Oliver E. (1975), *Markets and hierarchies: analysis and antitrust implications. A study in the economics of internal organization*. Free Press.

WB/OED, UNDP/EO and IOB (2003). *Country-Led Evaluations*. A discussion note.

THE ROLE OF NATIONAL, REGIONAL AND INTERNATIONAL EVALUATION ORGANIZATIONS IN STRENGTHENING COUNTRY-LED MONITORING AND EVALUATION SYSTEMS

*Oumoul Khayri Ba Tall, President,
International Organization for Cooperation in Evaluation, IOCE*

Introduction

The number of evaluation organizations (associations, societies, networks) has greatly increased in recent years, from 6 in 1997 to about 70 currently. While this reflects a growing interest in evaluation worldwide, it becomes crucial to analyze what value-added evaluation networks bring to the role evaluation is expected to play in improving development results. The statement that “Development is something that must be done by a country, and not to a country” is at the heart of Country-led approaches (CLA). The CLA concept was introduced in the mid 90s, and recently complemented by that of Country-led Extended Monitoring and Evaluation (CLE), which we believe is intimately related.

Embedded onto the CLA and CLE spirit alike, is ownership, which needs capacity to express its full potential and value, and both contribute, in principle, to a “virtuous cycle” of better public policy results.

This article builds on others, on the subject of evaluation organizations and evaluation capacity development (ECD), in earlier issues of this UNICEF series. It discusses further the comparative advantage of national, regional and international organizations, as well as the challenges in strengthening national monitoring and evaluation systems designed as CLE. Hence, we will not expand on a complete SWOT¹ analysis of evaluation networks to justify what the specific weaknesses and threats may be. Instead, we will refer to pertinent articles whenever appropriate.

1 SWOT: Strengths, Weaknesses, Opportunities and Threats.

There is generally a shortage of evaluation capacity, as current market needs (evaluation demand) develops rather quicker than the market response (supply) worldwide. The shortages are, in particular, of suitable skills, a suitable environment, and of an adequate evaluation culture. The gap to be filled is even wider and more complex in developing countries and in development evaluation. Most of our arguments will be illustrated from the developing country context. However, we also assume that we are dealing with universal principles and values which are not, and should not, remain developing country specific. Context matters everywhere and the call for Country-led methodologies should not overshadow the fact that monitoring and evaluation systems should remain “country-led” everywhere. CLE is called on to revise inappropriate country priorities and processes to produce more pertinent, coherent and sustainable results to improve peoples’ lives.

In general, evaluation will contribute to strengthening country-led-monitoring and evaluation systems in the following ways: (i) build awareness and evaluation culture; (ii) encourage a domestic evaluation demand, which will ultimately: (iii) extend the scope beyond Aid; and, (iv) improve the supply side through ECD strategies.

Evaluation networks worldwide: a brief overview

There is an unprecedented growth in evaluation organizations in response demand from international development agencies, bilateral and multilateral cooperation agencies, development banks and funds, governments, non governmental organizations and public sector. Quesnel provided a fairly complete picture of the different groups and accounted for more than 60 groupings in 2005. A current (2008) listing on the International Organization for Cooperation in Evaluation (IOCE) website contains 73 national and regional evaluation network references, mostly located in developing countries.

If we look into the evaluation network members, we find that older and more mature networks are more professional in nature and, very much like other professional sectors, seeking to gather various individuals with common concerns and interest in an emerging and growing profession. In contrast, newer networks tend to be more diverse and inclusive in their membership and in their interest. Kriel identified two groups emerging from her analysis of the 14 IOCE case studies: (i) organisations formed to organize and provide structure for an existing but fragmented community of evaluation stakeholders – mainly prac-

tioners, academician and researchers, and (ii) organisation formed to raise awareness and, in effect, build a community of evaluation stakeholders. As the practice evolves, evaluation organisations tend to carry characteristics of both groups, and their preoccupation tends to broaden to what a participant referred to, in a 2004 African Evaluation Association (AfrEA) conference workshop, as: "Associations should not serve as a trade union for evaluators, but as a dialogue space for evaluation stakeholders to shape the relationship between evaluators and the larger community".

In Africa, we usually find professional evaluators. For instance, consultants and evaluation officers in development agencies, projects and NGOs; academicians and university researchers; and, as government staff. This variety of profiles is often challenging even in terms of organisational setting, but it has the unique value of enabling a wide dialogue among evaluation stakeholders in the country and sometimes involves international links. In reality, however, not all parties are equally active, and in my experience, government people tend to be the least active. While this variety of members is a good thing, it needs a sound strategy and much personal involvement and the sustained efforts of a "core group of champions" to make it work, and to result in good state (practice and use) of evaluation in the country.

Almost all evaluation networks claim to be working on ECD. The networking features provide important opportunities to produce and share knowledge through "cross fertilization" of information and ideas. They help to advance the evaluation agenda in many ways. In this era of knowledge, evaluation capacity is increasingly recognized as a key factor in systems performance. Evaluation networks are playing a growing role in sustainable ECD, in particular in the developing world.

Different functions of evaluation: a brief reminder

Evaluation is about "extracting the true value of an action" in order to determine what benefits were made to the lives of those affected by the action. This simple statement does not, however, elude the inherent complexities and diverse realities of the concept of evaluation. If we agree on this broad definition, we still need to define what action is to be evaluated (the object), how to do it (methodology), why we do it (purpose), and who should do it (actors). Depending

on the answers, we have different realities which translate differing visions and interest of the same concept. A basic question would be why those different “things” are still called evaluation.

Traditional functions of evaluation stress the managerial and accountability features. Emerging approaches put the governance and policy dialogue dimensions forward. As a management tool, evaluation serves for evidenced-based decision making. The evaluation manual of the French Cooperation summarises this function as follows: “to gain greater knowledge, to better appreciate the value of an action, and make better decisions”. Evaluation is used as an accountability mechanism, fostering greater transparency, enhances governance and democracy, and the voice of civil society. Evaluation serves the knowledge generation and information sharing on public policies at different levels, and for different stakeholders, as a way to construct the policy dialogue and enlighten public policy processes. In countries where the policy dialogue is lacking, evaluation is seen as a way to “allow individuals to have a voice in their destiny”. This is the sense of a recent book authored by Ukaga and Maser. These functions reflect different types of evaluation, each requiring methodologies based on a combination of one or more basic approaches (formative versus summative, etc) depending on the object and context of the evaluation.

Evaluation networks play a key role in the evaluation arena as the functions evolve and “actions” being evaluated become more and more complex. To accommodate this evolving diversity and increasing complexity, evaluation networks are deeply engaged in critical thinking, knowledge generation and sharing, which makes an important value they bring in, to help advance the theory, practice and usefulness of evaluation. More importantly, evaluation networks are becoming an important actor in various development initiatives, at the national, regional and international level, in which they seek to support, but also influence, the processes so that the different voices they represent are heard and acknowledged. Recent impact evaluation (IE) initiatives led to the formation of the Network of Networks on Impact Evaluation (NONIE) in November 2006, by the three agency evaluation networks: The United Nations Evaluation Group (UNEG), OECD/Development Assistance Committee (DAC), and the OECD/Evaluation Capacity Group (ECG), with the aim to develop guidance on IE and set up a strategy to promote its use. Because the need to involve developing country perspectives was acknowledged, NONIE was then expanded to include developing country representatives identified through the global and regional

networks, led by IOCE, who form the 4th network (IOCE, the International Development Evaluation Association (IDEAS), AfrEA and other regional networks were each invited). The expectations will be met if the networks succeed in actively constructing the dialogue on the theme of IE to reflect the perspectives of developing countries in the processes defined. Evaluation networks will be better prepared to fulfil their mission in NONIE as they will meet the challenge of strengthening CLE systems. This is the sense of their call that ECD be considered an integral part of the NONIE supported strategies. In addition, evaluation networks, and IOCE in particular, will seek to reflect the basic values laid out in its foundation, “cultural diversity, inclusiveness and bringing together different evaluation traditions in ways which respect this diversity”.

Evaluation capacity

A debate on capacity and evaluation capacity is essential to understand how evaluation can actually contribute to better policy design, implementation and end results that are genuinely owned by the country, which is what CLE is about.

Capacity includes different realities from individual to institutional level. It is usually defined as “the power of something to perform or to produce. It is a continuing process of learning and change management. Capacities exist at different levels and several dimension”. Different levels of capacity range from the people, the unit/organization, the institutional infrastructure, and the policy environment.

Capacity is defined by the United Nations 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 is also the ability to perform and implement. Evaluation guidelines, principles, and ethical code of conducts are a key tool for capacity. Evaluation organizations are deeply engaged in the development of such tools. Most of them are inspired by the American Evaluation Association (AEA)’s *Guiding Principles for Evaluators* and the *Evaluation Standards of the US Joint Committee on Standards for Educational Evaluation*. AfrEA adopted the Joint Committee Standards and adapted them to the African context. Of the major changes made, new sections on participatory approaches were introduced, and the “African Evaluation guidelines” were adopted in 2002. In September 2006, a team of 30 evaluation practitioners representing all AfrEA member organ-

izations gathered in Niamey to produce an updated version. This exercise by itself was a major attempt to explore and scrutinize the evaluation practice in the continent to provide better guidance to evaluation stakeholders. In recognition of the diversity of the membership structure, participants to the working group and the final workshop were carefully selected to reflect a wide range of evaluation stakeholders from government, academicians, development partners, civil society organization and the private sector.

Previous publications in this UNICEF series dedicated a number of articles to ECD and the role of evaluation organizations. It is no doubt in the literature that evaluation capacity is strongly linked to evaluation organizations, in such a way that evaluation organizations are cited in many places as an element of evaluation capacity. While we strongly agree that evaluation organizations and evaluation capacity are intimately related, we do not believe in any simple cause and effect relationship and we need to analyze the criteria that make evaluation organizations successful in building capacity in a country.

Apart from the older evaluation associations, there is not yet enough evidence of convergence between good evaluation capacity and strong evaluation networks, in the middle income and developing countries. We would like to see that happen, as suggested by the conceptual framework of evaluation development theories. In fact, it might well be an apparent dilemma that illustrates the difficulties of defining evaluation capacity in a single, simple, static and linear way. Another dimension of capacity is the time frame. Capacity is not a short term business, and neither is development. It is rather a process that captures gradually the knowledge input to construct the ability to intervene in a favourable environment.

Evaluation culture

Beyond the technical and institutional aspects, the first challenge to developing evaluation capacity is the notion of evaluation culture. The concept of evaluation culture is not easy to define precisely. It is, however, intuitively easier to identify certain criteria that set out why evaluation is more likely to be successful in some environments than in others. I will call this the “Evaluation readiness”. An organization with an evaluation culture is one that: (i) refers to a known, shared policy about evaluation within the organization, meaning that: (ii) all members accept the use of evaluation, and: (iii) all members understand why the organization uses evaluation;

(iv) all can design or get advice on design of necessary evaluations; and, (v) all use evaluation, particularly to support change and development. Evaluation culture is important because it is fundamental to supporting the expression of the full potential of evaluation and to lead to the effective use of evaluation as a development mechanism. Further analysis of organization dynamics show that one of the most important elements, and the one that is the most often reported as missing, is the use of evaluation findings. Organizations usually have policies and perform evaluations as a technical and routine process, but then make no use of evaluation results to foster change.

The evaluation culture is affected by the values and rules of the organization or society (i.e. the organisation and society's culture), in particular with regard to information and power. To define the evaluation culture of an organization, Murphy poses the following questions: "Who does the evaluation? Who gets and uses the knowledge? How much institutional power do these people have? What is the culture of communication in the organization?" He concludes that virtually anyone in the organization could do evaluation, the results may be used properly or not, with or without consultation, and the combination of different responses will give as many evaluation cultures. Using this framework, we can tell that the evaluation culture under the traditional donor-led approaches is externally driven, and the CLE calls for the development of a national evaluation culture.

To strengthen the evaluation culture, evaluation organizations need to understand the rationale as above, as well as the framework for its use. Building the sort of evaluation culture we would like to see will usually require change in the individual as well as organizational culture (bureaucratic, hierarchy, leadership, goal-oriented, loose opinionated groups). What matters is the capacity to manage this change. Many evaluation organizations claim that they aim to build an evaluation culture, but a clear and thorough strategy for that is yet to be defined. Again, we have identified several points as possible guidelines from experience in specific fields such as education: fighting the stigmatism that threatens evaluation use, such as prior bad experience of instrumental use, unethical use, and un-useful evaluations which are a waste of time and resources.

Evaluation guidelines, principles, and ethical codes of conduct are a key vehicle to improve evaluation acceptability and credibility in the community. Kriel suggested that locally initiated and executed

“best practice” evaluations and high quality monitoring and evaluation methods be actively sought out, encouraged and rewarded, as a way to enhance evaluation culture. Networks are already actively engaged in some of these “applied research” methods and practices, mainly through their regular meetings, workshops and conferences.

Basle, in his preface for the white book on monitoring and evaluation and public action, revisits the evolving functions of Evaluation, from the “expert knowledge” times to the current functions of democratic debates around the worth and value of public policies, where all stakeholders have vested interest. In such settings, the role of the evaluator is also evolving more towards a facilitator of the evaluation design and process. Basle calls this the era of “monitoring and evaluation” where the capacity needs are those for “self-evaluation”. In some cases, the push for evaluation may come from the official institutions such as in France and in many European countries via the European funds (in the 1990’s). However, the need will arise gradually, then the network usually follows to support and strengthen the emerging evaluation culture (the French Evaluation Society or Société Française d’Evaluation -SFE- was created in 1999).

Strategies to strengthen country-led monitoring and evaluation systems

A growing form of knowledge organizations rooted into the national context are the communities of practice (CoP) that are developing around related themes such as the Asian and the African CoP on Management for Development Results (MfDR). The Asian Development Bank (ADB) website defines the Community of Practice (CoP) as “an informal network, a group of people who share a common sense of purpose and desire to exchange knowledge and experiences in an area of shared interest”. Through mutual learning and sharing of information, a CoP can develop and strengthen core competencies by developing and spreading good practices, connecting “islands of knowledge” into self-organizing networks of professionals, and fostering cross-functional collaboration.

The following paragraphs explore the role of evaluation networks in each of these dimensions, but we will also bear in mind that the CLE in return, will contribute to strengthening the networks, a sort of the “chicken and egg” paradigm. When the systems are in place

and working, networks play the key role of dissemination and sharing of knowledge, “cross-fertilization of ideas”, empowering evaluation stakeholders, strengthening the role of civil society, and sustaining all these achievements. On the other hand, networks are expected to contribute to building the system, through their advocacy and ECD roles.

Coherent strategies are needed to find suitable and sustainable mechanisms to address challenges facing the rapid growth of the evaluation sector, and the additional concerns for developing countries are summarised in the following three dimensions: (i) to develop an endogenous evaluation demand; and, (ii) to improve the quality of evaluation services on offer; and, (iii) to extend the scope of evaluation to policy level and development strategies.

Create a domestic evaluation demand

One major limit to CLE is the lack of domestic evaluation demand. Evaluation in developing countries is usually the domain of international development partners, who commission and conduct most evaluations. Of course, they do this in the light of their own concern of getting information on how well they are doing to assist the country, and not necessarily on how well the country is doing, which is quite different.

Quesnel identified three conditions for success or failure of ECD: (i) awareness and appreciation, at the government decision making levels, of the importance and necessity of evaluation. In other words, the existence of a demand for evaluation; (ii) the institutionalization and meaningful integration of the various evaluation functions in the government machinery at national, sectoral, programme/project and sub-statal levels; and, (iii) the development of human and financial resources to support a professional, dedicated, and effective cadre of evaluators and evaluation managers.

Under the context of CLE, it is the national actors who should have the primary responsibility to commission and undertake or oversee the implementation of the evaluation project. This does not happen naturally, as we said, it is the role of evaluation networks to create awareness of the benefits of evaluation at the national level. Their action complements the role of the government and its international partners, the latter in most cases have been the entering point of evaluation in developing countries.

Two main features of networks enable them to play such a role: their broad constituency, and their international linkages almost

everywhere. Usually all development stakeholders at the country level are members of the national association or network (government, international partners, civil society, private sector), which creates a great opportunity for dialogue on policies, and a vehicle to foster alignment and harmonisation. Many networks divide themselves in sub-groups, along geographic or thematic lines (Réseau nigérien de suivi et évaluation (RÉNSE) in Niger, Société Française de l'Évaluation (SFE) in France) which provides a greater anchorage into the fundamental needs and real-life issues facing policies.

Awareness building actions usually target actual or potential users, populations, and the public opinion. Awareness building activities for potential clients and users (development stakeholders) is part of the construction of an evaluation culture respectful of higher standards of good evaluation practice and use. Because it deals with public action, one may say that all parties involved in a public action have a vested interest in its evaluation.

Awareness can be build through the dissemination of information to target audiences, specific training to explain and illustrate the benefits of evaluation through workshops; debates; press articles; and, invitation to evaluation events. In developing context, where traditional ways of communication may be of limited access, evaluation networks will have to come up with innovative influential strategies to access diverse and non conventional development actors. A number of capacity building activities are specifically designed for parliamentarians and grass roots populations (participatory evaluation).

A useful way to create demand and domestic capacity is of course institutionalisation, which is the responsibility of policy makers. However, rules and regulations alone can not do it. I like to cite the case of Niger, where the "evaluation sensitivity" rose to a point where the government created a dedicated ministry, but the attempt failed shortly after, and the ministry was not included in the next government. Several questions could be asked and lessons learned from this case: is the creation of a ministry a good strategy; are there pre-requisites to that such as the existence of enough support in higher levels of the country; was this ministry a result of the national evaluation networks' action in the country or, a requirement of the donors or development partners?

It is important for the national government to keep leadership of the CLE mechanism, so as to balance power relationships with other partners and for legitimacy. However, we assume the political will,

governance requirements, and sufficient knowledge will be there. If not, then the virtuous cycle of the evaluation process would be the perfect vehicle to make change happen. It means that other development stakeholders will have equal interest to demand the evaluation they deem important for the country, in particular parliament, but also community-based and Civil Society Organisation (CSO).

Extend the evaluation object and scope beyond aid

Ownership is the key factor to reverse the development trends where poverty remains despite significant economic growth recorded in African countries. What it implies is the need to allow countries to decide, by themselves, how they would like to make use of their financial resources (domestic as well as foreign aid resources), and how they will manage its use to produce results. In other words, this is about ownership of development and development evaluation. It took donors and the development machinery so long to understand what seems rather obvious: that aid money should be managed from inside, and not from outside, to actually serve the development needs.

If development policies are owned, then the monitoring and evaluation system in place is more likely to be owned, which means that it is designed for the sole purpose of informing the client on how the policy performed, what results were observed and, what benefits obtained. Ownership of the process means accountability not only to donors, as is the case when the policy is solely from donors perspective. Even when the program is 100% from donors money, it is more likely to result in positive outcomes for the beneficiary if the programme is made accountable not only to donors but to clients as well. This new paradigm is what is needed to achieve ownership of policies supported through Aid resources. Additionally, and more importantly, it is the overall policy and its pertinence and coherence that need to be looked at, regardless of the sources of finance, so that the response to the questions of efficiency and effectiveness make sense with regard to development objectives. Basically development stakeholders, including beneficiaries, should have a common understanding of the objectives to be reached, and the way they will monitor and evaluate the implementation and the results.

Country-led systems should allow this in-depth and global approach in development policies. Evaluation networks understand the need to evaluate policies beyond aid. They are mobilizing their resources to advance the theory and practice, to face the methodological challenges caused by the complexities in development evalua-

tion (IDEAS run a number of workshops in various regions to learn more about CLE experiences). This complements what institutional development agencies are already doing. Due to their diverse membership base, and the reach that it allows, these networks are in a better position to detect innovative experience and practice where those exist, and to give the voice to non-traditional development actors, which increases their chances to design suitable and accepted solutions.

Finally, one major benefit of owned strategies is that they have a better chance of resulting in effective buy-in and use. This is as valid in monitoring and evaluation solutions as it is in development policies.

Improve the supply side through evaluation capacity development

The development of human and financial resources to support the professional, dedicated, and effective cadre of evaluators and evaluation managers is the third of three conditions of success identified by Quesnel.

Despite the dynamism observed among the evaluation community, we have not yet reached the state where evaluation is considered a profession. The evaluation community is being challenged by a poor record of practice. This includes failure to meet certain quality requirements and the number of evaluations proving not to be useful. In particular, in development evaluation, recent debates following the publication of the report from the Center for Global Development (CGD): “When will we ever learn”, claimed that more impact evaluations should be undertaken, to increase the effectiveness of development interventions. The report says that the majority of evaluations undertaken have failed to demonstrate the impact of development actions, and therefore have had limited usefulness. Behind this call for “more rigorous evaluations”, many practitioners have seen a call for higher quality evaluations in all phases from design to the final report, possibly up to dissemination and actual use of evaluation findings and recommendations.

The solution for higher evaluation quality is partly in education and training of evaluation practitioners as well as commissioners, both formally and informally and in the development of professional norms. Networks have been instrumental in developing information on competencies, standards and norms, and ethical codes for evaluation. But the biggest challenge is in the use of these “norms” as

guides to serve the actual purpose of quality and useful evaluations. Perrin listed various forms of evaluation professional development training or events which many evaluation networks offer to their members and other interested public (Canadian Evaluation Society (CES) with the Essential Skills Series introductory course, European Evaluation Society (EES) with their residential summer school, etc.). He also advocates the necessity to develop the whole range of skills needed for any evaluation to reach its goals, including soft skills, which may well go beyond the capacity of single individuals or entities, and single training programmes.

Learning is obviously one major way to enhance the quality of evaluation and it happens in the classrooms as much as outside, and recently in the internet. Training by doing is one major cost-effective capacity building method that is getting increased attention. Evaluators from the south are integrated into larger and more experienced teams conducted by lead-evaluators from the north, usually selected by the donor agency. This strategy proves to be effective under certain conditions. The partner from the south should be actually integrated in the team, and given substantial task from the beginning, not just the administrative and organisational aspects of the field visits, or the summary of literature reviews, as it is often the case. Again networks play a crucial role in organizing the supply and demand of evaluation consultancy services.

Most of those who practice evaluation in their professional life have never received a formal education in evaluation as a separate self standing discipline. Usually, they have taken evaluation courses as part of their curricula in traditional disciplines such as education, medicine and social science, or have been trained later in the many professional development events existing. In development evaluation, the International Programme Development Evaluation Training (IPDET), organized by the World Bank in collaboration with Carleton University in Ottawa, is the first formal comprehensive training we know of. IPDET graduates comprise the largest number of IDEAS members, a worldwide evaluation organisation dedicated to promote international development evaluation. In this case, the training programme has been intertwined with a networking mechanism, with the obvious aim of providing the evaluation community more opportunities to continue the learning process.

There is a growing consensus within the evaluation community that the time has come for professionalisation, to command more respect and trust from the public, yet some concerns still exist. One

major project the CES is currently working on, the “Professional Designations Project”, was presented in their last Conference in Quebec in May 2008. It attracts much interest and attention, and some apprehensions as to what the final product will look like. This project will surely offer the opportunity to clarify the questions of certification or licensing (of professionals) and accreditations (of training programmes and schools). In the meantime other initiatives are on the way, to develop certification trainings (UNEG, IPDET), which shows that the demand is there for some sort of recognition.

The evaluation community is trying to attract greater interest from the academic world and specialised training institutions and to increase opportunities to engage them in developing evaluation curricula. The example of IPDET with Ottawa may be seen as a good practice worth replicating. Such initiatives are taking place in other parts of the world such as Latin America (with UNICEF and the ReLAC partnering with a number of universities to offer training), and in English speaking Africa (recently planned). A prospective target group is students who are offered special rates to attend conferences or workshops as a way to encourage more interest into the field.

As Bamberger puts it in describing the role of IOCE, “national, regional and international networks can mobilize experience, documentation and resource persons to provide support in many areas of ECD”. Thus far, in addition to professional development workshops run during the conferences, networks are offering fundamental resources through their websites and list-servers, newsletters, magazines, journals and other publications.

Conclusion

To summarize, the key role of evaluation associations and networks is to improve evaluation theory, practice and utility while “serving as a dialogue space for evaluation stakeholders to shape the relationship between evaluators and the larger community”². What value evaluation brings into countries and how this will happen transcends national boundaries³, however this should be deeply rooted in countries first, to be effective. Evaluation is deeply embedded into the major development initiatives that have been taking place in recent years, such as the Millennium Development Goals (MDGs) and the Paris Declaration.

-
- 2 Elliot Stern unpublished notes from a session in AfrEA Conference 2004.
 - 3 Russon and Russon, 2005: the “Quality of evaluation is an issue that transcends regional and national boundaries”.

MDGs is a call for better development providing precise indicators on what needs to be achieved to uplift the world living standard to a more acceptable level, given the status of the world's wealth and knowledge. It is a call to make use of human intelligence to win the battle against poverty. Paris Declaration reminds us that the willingness to speed up and sustain development has to come from the countries themselves, in particular those most in need; it takes a national effort from all development stakeholders to win this battle, as well as international solidarity to complement the resources needed. Both are grounded in the principles and values of CLE, which is the missing link to activate the virtuous cycle of the development process through evidenced-based policy design.

I wish to make it clear that I am not assuming that effective evaluation networks will lead automatically to good evaluation standing in a given country, there are examples of countries with no strong evaluation network which are making sensible progress towards good evaluation policies and practices. Ghana is a good example. In general, it is observed that evaluation networks because of their work on the ground, tend to be an effective way to build capacity in a given country, and even beyond the country, as in the case of the American Evaluation Association (AEA), Canadian Evaluation Society (CES), IDEAS and IOCE.

To be effective in strengthening CLE systems, evaluation associations and networks must play this role of organising the national dialogue amongst all development stakeholders in the country, and make the bridge to the international community of evaluation. Of course, organisations must be operational, well organised, based on a supportive and efficient governance structure, and evolve in an enabling environment, to be able to play such a fundamental role.

References

- American Evaluation Association (AEA). *Guiding Principles for Evaluators and the Evaluation Standards of the US Joint Committee on Standards for Educational Evaluation*.
- Bamberger, Michael. (2006) *Evaluation Capacity Building*. In Segone, M. and Ocampo, A. *Creating and Developing Evaluation Organizations: Lessons learned from Africa, Americas, Asia, Australasia and Europe*. IOCE.
- Center for Global Development (CGD). (2006). *When will we ever learn. Improving Lives through Impact Evaluation*, USA.
- Kriel, Lise. (2006) *How to Build Evaluation Associations and Networks: Learning from the Pioneers*. In Segone, M. and Ocampo A. *Creating and Developing Evaluation Organizations: Lessons learned from Africa, Americas, Asia, Australasia and Europe*. IOCE.

Ministère des Affaires Étrangères / Directorate General for Development and International Cooperation (MAE/DGCID). (2007). *Evaluation guide 2007*. Available at: http://www.diplomatie.gouv.fr/en/IMG/pdf/399__Int_Guide_eval_EN.pdf

Perrin, Burt. (2005) *How can information about the competencies required for Evaluation be useful*. In Canadian Journal of Program Evaluation, Canada.

Quesnel, Jean Serge. (2006) *The Importance of Evaluation Associations and Networks*. In Segone, M. and Ocampo, A. *Creating and Developing Evaluation Organizations: Lessons learned from Africa, Americas, Asia, Australasia and Europe*. IOCE.

Russon and Russon. (2005). *Quality of evaluation is an issue that transcends regional and national boundaries*, USA.

Segone, Marco. (2006) *National and Regional Evaluation Organizations as a Sustainable Strategy for Evaluation Capacity Development*. In Segone, M. and Ocampo, A. *Creating and Developing Evaluation Organizations: Lessons learned from Africa, Americas, Asia, Australasia and Europe*. IOCE.

Ukaga. Okechukwu and Maser, Chris (2004). *Evaluating Sustainable Development: Giving People a Voice in Their Destiny*.

BRINGING STATISTICS TO CITIZENS: A "MUST" TO BUILD DEMOCRACY IN THE XXI CENTURY

Enrico Giovannini, Chief Statistician, OECD

Introduction

The fundamental role of statistics in modern societies has been underlined many times. In some countries, the role of statistics as "public good" has been described in the constitution. So, how is the revolution coming from the "information society" and the availability of new information and communication technologies changing the role of statistics? How does this change relate to the functioning of a democracy in the "information age"?

This paper identifies some key challenges for official statistics in terms of relevance, legitimacy and, therefore, their role in modern societies. Moreover, it investigates how citizens see and evaluate official statistics and the role that media play in this respect, using empirical evidence concerning several OECD countries. Some conclusions are drawn about the need to transform statistical offices from "information providers" to "knowledge builders" for the sake of democracy and good policy.

The value added of official statistics: where does it come from?

Economic statisticians, and especially national accountants, have developed guidelines on how to measure the value added of each and every economic activity, but very little effort has been put into the measurement of the output and the value added associated with the work of national statistical offices (NSOs) and of international organisations producing statistics. A recent survey carried out on 28 countries¹ indicated that the most frequently used output indicators include: number of publications (or number of releases); number of publication copies sent to subscribers; number of visits to the Internet page; number of indicators accessible in the Internet databases; number of tables viewed in the Internet databases;

1 See <http://www.unece.org/stats/documents/ece/ces/bur/2008/25.e.pdf>.

number of presentations at conferences and seminars; and, number of media quotations. Many NSOs also try to measure the quality of output with quantitative indicators (punctuality of releases, number of errors discovered in published information, revisions in statistical database, etc.), or user satisfaction surveys. Of course, all these measures are very important to monitor the implementation of the work programme and the usage of statistics. However, can we really say that they are good measures of output and/or value added of official statistics? In the following we will try to develop a "model" to measure the value added of official statistics using the statistical standards developed to measure economic activities.

According to the International Standard Industry Classification (ISIC Rev.1), the production of official statistics is a non-market service. It is part of Section L, Division 75 "Public Administration and Defence", Group 7511 "Administration of the State and the economic and social policy of the community", which includes "administration and operation of overall economic and social planning and statistical services at the various levels of government".

According to the System of National Accounts, services are the result of a production activity that changes the conditions of the consuming units. In particular:

"The changes that consumers of services engage the producers to bring about can take a variety of different forms such as:

- (a) changes in the condition of the consumer's goods: the producer works directly on goods owned by the consumer by transporting, cleaning, repairing or otherwise transforming them;
- (b) changes in the physical condition of persons: the producer transports the persons, provides them with accommodation, provides them with medical or surgical treatments, improves their appearance, etc.
- (c) changes in the mental condition of persons: the producer provides education, information, advice, entertainment or similar services in a face to face manner"².

For statistics, the third case seems to be the relevant one. Therefore, the value added of a statistical service should be related to the change in the mental condition of the individual.

2 System of National Accounts 1993, page 123.

For market services the price paid by the consumer reflects, by definition, the value that she or he attributes to the fruition of the service, but for non-market services a different approach must be followed. According to Atkinson (2005), various methods can be followed to evaluate the value added of non-market services, but, as a general rule, methods aimed at measuring outputs should be preferred over those based on the measurement of inputs (salaries and intermediate costs). In particular, "the output of the government sector should in principle be measured in a way that is adjusted for quality, taking into account the attributable incremental contribution of the service to the outcome" (page 187).

What should be the final outcome of official statistics, considering what the SNA says? "Knowledge" seems to be the answer: knowledge of economic, social and environmental phenomena³. If a person knows nothing about a particular issue and looks at relevant statistics, should that person not become more knowledgeable (to a certain extent) about that subject? Of course, the "new" knowledge could eventually lead the person to particular behaviours, but for that to happen the person needs to combine the statistical information with other information (including their beliefs, ideology, opportunity cost considerations, etc.). Therefore, the immediate outcome of the consumption of statistics is not the behaviour, but the expansion of the information set used to make decisions.

We could then conclude that the value added of official statistics (VAS) is linked to what the actual (not the potential) users know about the facts that are relevant to them in making their decisions. Therefore, from a collective point of view, this value can change according to two factors: the size of the audience (i.e. the number of people who know official statistics, N); and, the quantity of official statistics (QS) actually included in the information sets relevant for each individual's decisions:

$$VAS = N * QS$$

If only a small group of people are aware of official statistics, the probability of society using them to make decisions is relatively small. On the other hand, if everybody knows about official figures, but individuals do not actually use them when making decisions, their value added will be minimal.

3 As reported by Wikipedia, the Oxford English Dictionary defines "knowledge" variously as: (i) expertise, and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject, (ii) what is known in a particular field or in total; facts and information or (iii) awareness or familiarity gained by experience of a fact or situation.

Globalisation, the information society and political reforms (that require individuals to take decisions that in the past were taken by the government – pensions, education, etc.), are making N bigger than ever, while QS can depend on several factors, such as:

- the total amount of official statistics that reaches a generic user (QSR). This amount depends on two elements:

$$QSR = QSA * MF$$

where QSA represents the total statistical information produced by the official source and the role played by media (MF), which can emphasise or reduce the actual amount of information communicated to the generic user;

- the relevance of the official statistics communicated to the user (RS);
- the trust that individuals have in official statistics (TS);
- the individuals' "numeracy" (i.e. the ability to reason with numbers and other mathematical concepts, NL).

We could then write the following expression:

$$VAS = N * [(QSA * MF) * RS * TS * NL]$$

Of course, it is extremely difficult to quantify the different elements that enter in the equation. However, some sparse evidence exists. For example, as described in Giovannini (2007):

- 69% of the European citizens believe that it is necessary to know key economic data (such as Gross Domestic Product (GDP); unemployment rate; inflation rate; etc.)⁴, but 53% of European citizens do not have even a vague idea of what the GDP growth rate is in their country and only 8% know the correct figure⁵;
- 45% of Europeans tend not to trust official statistics, while 46% tend to trust them;
- in the United States, the most common source of information on official figures is television (TV), (78%); followed by newspapers (58%); Internet (37%); radio (34%); family/working networks (34%); and, magazines (14%). The five main TV networks quite frequently report data on the unemployment rate (83% of cases on

4 These data were collected in 2007 by the European Commission (Eurobarometer) at the OECD's request in preparation for the second OECD World Forum on "Statistics, Knowledge and Policy" (www.oecd.org/oeecdworldforum).

5 Similar figures have been obtained by Curtin (2007) for the United States.

average), but much less frequently data on GDP growth (46%) or inflation rate (35%). Looking at the 27 most popular newspapers, on average they covered just 39% of the official reports on GDP, 53% of those concerning Consumer Price Index (CPI), and 52% of those announcing the official unemployment rate⁶;

- finally, when disseminating US economic data, Associated Press and United Press International (the most popular wire services) typically do not mention specific source agencies in their releases. This approach has a clear impact on the "brand name" of the source: 23% of Americans have never heard of official unemployment data or the source agency; the comparable figures are 34% for CPI and 40% for GDP.

This review underlines three key points for the following discussion: first, the way in which statistics are used/perceived by users (especially citizens) depends on several factors and some of them are not under the control of the original source; second, in several countries the situation is far from being satisfactory in terms of trust in, and communication of, official statistics; third, statisticians have to address these issues (measurement of their output and value added; relationships with media and final users; brand image; etc.) very seriously, especially if they wish to respond to the challenges coming from the "web 2.0 revolution".

Statistical information, citizenship and democracy

Information plays a great role not only in modern micro and macro-economic models. It is also important in "public choice" models, in the so called "positive political theory", which are based on rational choice modelling and on analytical conclusions reached by the economic theory. Downs (1957) first introduced rational models for the political choice of individuals, considering the election mechanism as a "market" in which politicians supply different political platforms which are demanded by voters, who have to decide whether and how to vote. To do that, the generic voter estimates a "party differential", i.e. the difference between the expected util-

6 "If we presume that the 27 papers with the largest circulations all had access to the wire reports, the lack of complete coverage would be an active decision of the newspaper to not carry the report. It was likely to reflect a judgement about the newsworthiness of the latest figures given their subscribers' interests. There was a tendency for newspapers to more frequently report the latest official figures when it represented an unfavourable development, which may reflect the greater importance people place on the information content of 'bad' news" (Curtin, 2007)

ity derived from the choice between various (normally two) parties' candidates. A voter whose differential between parties is non-zero subsequently takes into consideration the cost of voting. To vote, the cost of voting must be lower than the "discounted utility" of voting, calculated using the likelihood that his vote will make a difference in the election.

What is extremely important here is to note that one of the components of the voting cost is the cost of collecting information. Acquiring information about candidates and policies can be very expensive and the value derived from this search must be discounted by the fact that the individual has little impact on the final outcome of the elections. Thus, the citizen is viewed as a "rational ignorant" and the obvious impact of missing or limited information on political issues is that the percentage of informed voters in elections could be very low. This is not a good thing for democracy.

In other models based on "game theory", political elections are seen as incomplete contracts between a less informed principal (the voter) and an agent (the politician) who has to achieve the principal's goals in an incomplete information structure. If a representative democracy is a form of state in which people control the choice of government, through elections, voters have the opportunity to achieve four major objectives: aggregate their personal preferences, making clear to politicians their welfare function; aggregate dispersed information about the correct political decisions; solve an adverse selection problem by selecting the best candidates; mitigate moral hazard problems by holding elected officials accountable for their actions.

The major problem is that, contrary to the principal-agent link in a market, the principal (the voter) does not have a proper indicator at a reasonable cost (such as price), that can drive the politician's actions. The most politicians can commit is an input (public expenditure, tax rates, etc.), not an output (economic growth, low inflation, etc.). That is, a programme not a result. They can commit themselves on variables they control, but the promised results depend on the reliability of the commitment and the solidity of the theory used to identify instruments and evaluate expected results.

The sticks and carrots (i.e. the sanction of no re-election, the premium of being re-elected) mechanism only works if there is a proper measure of outputs/outcomes delivered by a certain policy. Of course, information plays a great role in this process. In fact, in a world of costly information, rational citizens will spend more time

informing themselves about their own private purchases than about public policies, for which their efforts will have little effect. Therefore, voters, like shareholders of a large firm, face the difficult task of monitoring the activities of large hierarchies staffed by people who have information and expertise that is unavailable to the average voter⁷.

If elections are seen as a particular kind of contract, politicians use elections as a way to gather individual preferences in a social welfare function, trying to maximise it in order to be re-elected in the future. In contrast, other voters observe political outputs/outcomes and decide if their objectives have been achieved, and re-elect the good politicians or change their preferences. However, voters are in a weaker position, because at the beginning of the process they cannot discriminate between good and bad politicians, especially in a majority system of elections where political platforms are very similar. Moreover, when elections have taken place, politicians use their information advantage to maximise their "rent", without accomplishing the goals preferred by citizens.

In economic terms we have here both an "adverse selection" and a "moral hazard" mechanism. The first could be mitigated through a mechanism by which good politicians, through high-cost actions, do their best to demonstrate that they are superior to the relatively bad politicians in terms of better achieving citizens' goals. The second, instead, could be addressed with an incentive mechanism, by which the politicians who do not attain voters' goals are punished with no re-election. To do this at least one performance indicator is needed to evaluate if voters' goals have been reached. Of course, voters should be able to constantly monitor such an indicator. Following Swank and Wisser (2003), a higher probability of observing the policy outcomes narrows welfare losses. This gives the right incentives to the incumbent politicians for examining projects, and enlarges the range of examined policies. This suggests that it is in the interest of the citizens to improve the likelihood of observing politicians actions. Elections are not an appropriate "stick and carrots" mechanism to enforce an effective political process. It is instead, information, which plays the main role. As long as indicators about concrete actions and achieved results are a right measure of policy, and properly publicised, they may help society to achieve better goals with less resources.

7 A similar relationship exists between politicians and bureaucrats (see Niskanen, 1971 and Holmstrom, 1979).

Knowing and using statistics to make decisions

As discussed above, the importance of statistical information for democratic processes has been underlined by “public choice” models. The recent literature on the relationships between public opinion, political choices and the functioning of modern democracies argues that there are big differences between what the general public and specialists, such as economists, think about key issues. Increasing attention is given to public opinion, even when it is poorly informed. For example, Blendon et al. (1997) looked at the results of national surveys which compared the public and economists’ evaluations of current and past economic performance, their expectations for the economy and their perceptions of why the economy is not doing better. They found that a large proportion of citizens (especially those without a college degree) believed that the economy is performing worse than official data show. Moreover, their results indicate a substantial gap between how the public and economists see the economy.

These findings have been extended by other researchers. For example, Caplan (2002), examining the results of the Survey of Americans and Economists on the Economy, finds that beliefs about the economy differ systematically with ideological preferences. Kirchgassner (2005), looking at data on various countries, concludes that the gap between economists and the rest of society is wider in Continental Europe than in Anglo-Saxon countries.

Blinder and Krueger (2004) present some evidence about what U.S. citizens actually know about key economic facts. They found that a significant number of Americans do not know very much about the country’s economic situation. They also tested a range of factors that might explain how people’s beliefs are shaped. They found that ideology was the most important determinant in shaping the public’s opinion. Self-interest was the least important, and economic knowledge was in between. Therefore, their findings seem consistent with an idea from political science: people often use ideology as a short cut for deciding what position to take, especially when it is difficult to properly inform oneself. They conclude that “there is room for hope that greater knowledge will improve decision making, even though it appears from our survey that efforts in this direction have shown less than impressive results to date”.

Following this example, the OECD has promoted the first co-ordinated international survey on what citizens know about key economic statistics (see www.oecd.org/oecdworldforum). The survey, carried out by Eurobarometer was aimed at measuring what citizens know about key official statistics and their confidence in these figures. It was conducted between 10 April and 15 May 2007 in the 27 EU countries, plus Turkey and Croatia. Around 1000 people in each country were interviewed. A first set of questions concerned the extent to which European citizens are aware of key economic figures, such as the GDP growth rate, the unemployment rate and the rate of inflation. Other questions were aimed at assessing whether citizens think that it is important to know these figures, believe that these figures are used to take political decisions, and trust official statistics.

On average, 69% of the respondents believe that it is necessary to know these key economic data, but the variance is extremely high across countries. Cyprus, France, Spain and Portugal are the countries with the highest percentage of citizens (more than 80%) who have this conviction. In Slovenia, Lithuania, Bulgaria and the Netherlands, on the other hand, only 50% to 60% of people believe that it is important to know these figures.

Unfortunately, believing that it is very important to know key economic indicators is not the same as having a good knowledge of them. The survey also asked questions relating to what citizens know about statistics on GDP growth, unemployment rate and inflation rate. The answers are quite discouraging. On average, 53% of European citizens do not have even a vague idea of what the GDP growth rate is and only 8% know the correct figure. The corresponding percentages when it comes to unemployment rates are 48% and 11%, while for the inflation rate they are 28% and 6%. This is not just a European problem, as similar figures have been obtained by Curtin (2007) for the United States.

Figure 1: Importance of knowing key macroeconomic indicators

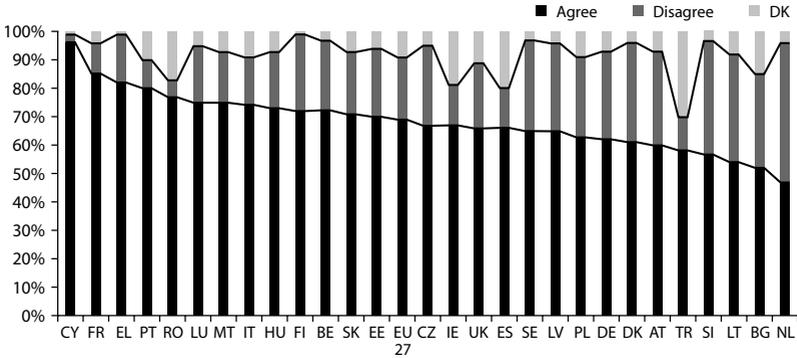
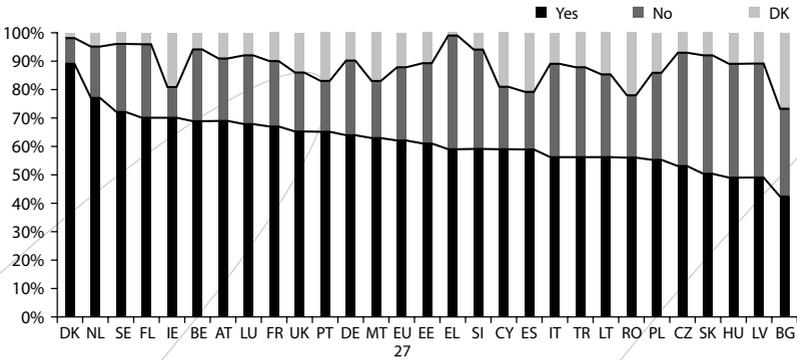


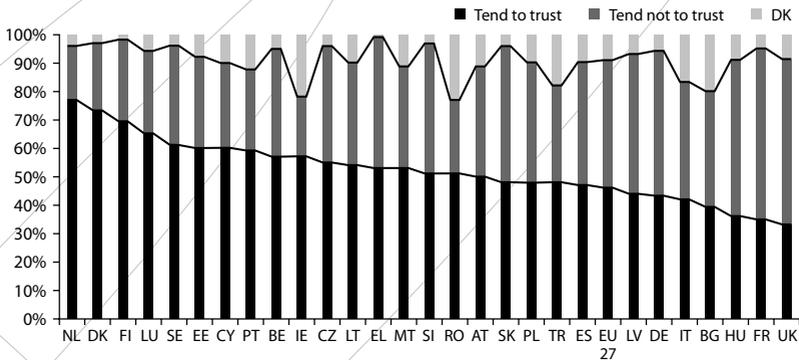
Figure 2: Use of statistical information to take political decisions



The main conclusion that emerges from these data is that people would like to know more about what is going on in their country, but their actual knowledge of key data is very limited. Is this because they pay no attention to official data? Is it because they do not trust them? To investigate this issue, a second question concerning the use of statistics for policy making was included in the survey: *“Some people say that statistical information plays an important role in business, public and political decision making. Personally, do you think that, in your country, political decisions are made on the basis of statistical information?”* On average, 62% of the respond-

ents consider that, in their respective countries, political decisions are made on the basis of statistical information. Here, again, the variance is quite significant. In general, Scandinavian countries have the highest shares of "yes" answers: for example, 89% of Danish respondents answered in this way, as did 77% of respondents from the Netherlands. On the other hand, several former communist countries have the lowest percentages of citizens who believe that political decisions are taken on the basis of statistics.

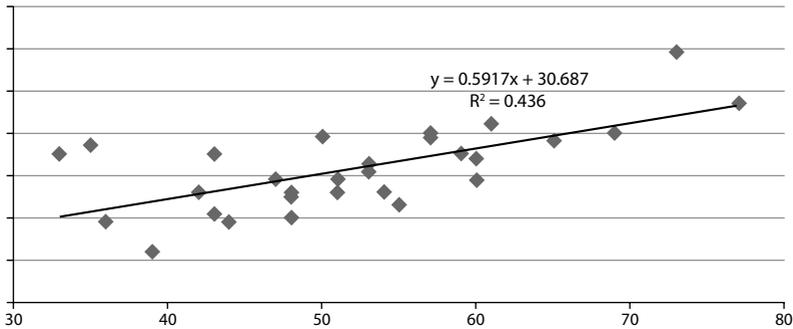
Figure 3: Trust in official statistics



Lastly, trust in official statistics was evaluated. 45% of European citizens tend not to trust official statistics and 46% tend to trust them. Here, too, the highest percentage of trust is shown in some northern European countries (the Netherlands, Denmark and Finland), while the United Kingdom, France and Hungary show the lowest trust in official statistics.

In summary, these results confirm both the existence of a general demand for economic data as part of the global knowledge that people should have in order to better understand what is going on in their country, and the fact that a large majority of citizens are not aware of them. The results also confirm the serious issue of trust that official statistics face today. The strong correlation between the belief that statistical information is used for policy making and the trust in official statistics also shows that the way in which they are perceived by citizens also depends on the way in which policy-makers use statistics, and vice versa.

Figure 4: Belief that statistics are used to make political decisions (Y axis) and trust in official statistics (X axis)



There is also large and well-established literature that analyses the way people use information to make choices. Much of the most influential work takes a psychological or behavioural perspective. Specifically, H. Simon, J. March and R. Cyert, all working at Carnegie Mellon University, have made pioneering contributions to the study of the cognitive processes underlying the way people make (rational) decisions. Their research has been extended by D. Kahneman, P. Slovic and A. Tversky, amongst others, whose work looks at the rules that people use to guide their decisions, when decisions are complex and they do not have perfect information.

Recent work relates more directly to statistics and their dissemination. Carroll (2003) tests a model of how empirical expectations are formed. His approach takes the news as the key provider of information on macroeconomic variables. He adds to this, firstly, the idea that people do not update their expectations and personal forecasts continuously but probabilistically. In addition, he looks at the role professional forecasters play in informing the media. Specifically, Carroll's model offers a way to relate the public's forecasts to those aired by the media, which in turn originate from professional forecasters. In his empirical analysis, he uses data on the expectations of professionals from the Survey of Professional Forecasters (SPF) as an input to this model. He finds the model is quite good at explaining the public's expectations for general inflation and unemployment measured by the Michigan Survey of Consumers.

Empirical work by Doms and Morin (2004) supplements Carroll's analysis. These authors elaborate the role of the media. In particu-

lar, they establish three important ways through which the media affects the public’s views on the state of the economy:

- by conveying economic data and expert opinions;
- by sending a signal based on the tone of the economic report and the volume of reporting (e.g. number of articles); and
- by the volume of reporting, which influences the likelihood of people updating their expectations (this adds to the signal value of the amount of reporting).

What can we conclude from this brief overview? The first conclusion is that, notwithstanding the efforts made by statisticians to produce reliable statistics, by the media to disseminate them to citizens, and the general improvement of education, the “statistics, knowledge and policy” chain is far from well-established. The second, policy-oriented, conclusion is that since the “chain” is not working to its maximum “capacity”, something can and should be done to reinforce the links between statistical evidence and its use by individuals, in taking their own decisions, and via democratic decision-making processes.

Globalisation and the dissemination of information

This evidence makes it clear that, as Einstein said: “information is not knowledge”. Of course, trust in the source of information plays an important part in the way people use the available data to make their decisions. Therefore, what people know must not be confused with the amount of information they receive every day and absorb from the most disparate sources. Instead, knowledge refers to a complex and dynamic process involving cognitive mechanisms whose effect is not reducible to what is known by the subject at a given point in time. Therefore, as the value added of official statistics depends on its contribution to building societal knowledge, it is necessary to understand how information, and at a higher level knowledge, is spread through the population in a globalised world.

Of course, knowledge and information are strongly related to each other, but in order for a body of information to “become” knowledge, cognitive mechanisms (usually referred to as processes of codification and de-codification), are required. Several models have been developed to explain how these mechanisms work. One which is particularly relevant to this discussion is the model based

on the so-called “epidemiologic” approach. Originally developed for cognition and culture by Dan Sperber (a French cognitive scientist), this approach seeks to explain the relation between human mental faculties and social cultural phenomena. Sperber argues that there are two kinds of representations: mental and public. The former depend on the functioning of each individual’s brain, while the latter are phenomena belonging to an environment of people who perceive and represent them in a certain way. The thrust of the epidemiological approach consists in relating the two representations to each other. In fact, individuals are used to representing mentally the contents derived from their own experience of life as well as from communication with others, with the effect of creating mental representations that, in turn, end up being shared through language and further communication.

In a nutshell, the epidemiologic approach states that information is spread in a society like a virus. At the beginning only a few people catch it, but then each “infected” person transmits it to others, and so on. However, every time there is a transmission the information changes a little, as viruses do. In this context, three points require special attention:

- the amount of news released by the media plays a key role in affecting what people know. Since their exposure to the media varies for many reasons, it seems inconsistent to assume that the same amount of information is available to everyone at the beginning of the process;
- the quality of media and their way of presenting information, which can make a huge difference to people’s capacity to grasp the sense of the what is communicated (a few seconds of a speaker to inform about the GDP growth in the last quarter or, 30 minutes of a debate among experts about the economic situation of the country, clearly can have very different impacts);
- the degree of exposure to the media is not sufficient for a person to be properly informed and to process the news so as to show actual knowledge of the subject at issue. For example, some people are likely to be more interested in economic information than others, and also the capacity to fully understand and effectively process that information varies considerably from individual to individual.

Like the spread of a disease through the population, the news penetrates through to the agents in various degrees. Moreover, the news

to which people are exposed can come from a variety of sources, such as a community of experts, opinion leaders, friends, etc.

What does this mean for official statistics? If information is spread across society as a virus, which evolves with every passage, it would be fundamental for NSOs to reach as many people as possible at the beginning of the chain, to "vaccinate" them against the "ignorance disease". In this way, both the "brand image" of the statistical office would be transmitted together with the data, and the message itself would be as accurate as it could be. But this is not what NSOs normally try to do. Instead, they rely heavily on mass media, such as newspapers, radio, television, etc., who are delegated to present data to people⁸.

To maximise the impact on the "conventional" media, a large number of initiatives have been launched by NSOs, including training courses for journalists. The timing of data releases is also chosen to maximise their impact on the media. But how effective is this approach? Unfortunately, there are few case studies available to shed light on this issue (see Curtin, 2007). The results of Curtin's study "suggests that people's lack of knowledge can be in part attributed to the inadequate communication of that information by the mass media. It was true that news on unemployment was more frequently reported in the media, and people's knowledge of the unemployment rate was more accurate in the survey. The coincidence is suggestive but does not prove causation".

What is undisputable is that, in very rough terms, only 50% of key data concerning the US economy is actually passed on to citizens by TV or newspapers. This means that the overall value added of statistics is considerably reduced by the mass media, which filter data released by official sources depending on their corporate policies or political interests. Perhaps this is the only case of a public service whose final outcome is decided by the private sector!

Of course, the functions of wire services have been supplanted in recent years by the simultaneous Internet releases of the official statistics. In this way, people from around the globe can access the same data the instant it is released via the Internet. According to data provided by BLS, the full release of the unemployment rate was seen (on 4 May 2007) by 8,243 people, while the release for the CPI (on May 15, 2007) was opened 11,959 times (about 1% of all the visits to their Internet sites on those days). These figures

8 Of course, Internet also plays a crucial and growing role in reaching important but smaller audiences (academic experts, consultants, etc.).

show that, although these alternative communication channels are growing, they cannot replace the most classical ones.

The Web 2.0 revolution

Statistical data providers are aware of these problems and have heavily invested resources to improve their communication tools, especially the use of Internet. But new Information and Communications Technology (ICT) tools and the success of Internet are also profoundly changing the way in which people, especially new generations, look for and find data. For example:

- according to Internet experts, 95% of those who use Google do not go beyond the first page of occurrences. Once they reach a particular site, a similar percentage of users do not click more than three times to find what they want. If after three clicks they have not found what they are looking for, they quit the site;
- the way in which “discovery metadata” are structured is fundamental to their placement in the first page of Google’s results, but these metadata have nothing to do with the intrinsic quality of the information provided. Therefore, sources able to structure their “discovery metadata” well, can appear higher than those which have better quality information but do not invest in this kind of metadata.

Everybody is aware of the most popular tools and success stories developed by the Internet community over the last few years. Maybe, less people are aware of the deep changes that the web 2.0 is producing in the way in which “collective knowledge” is generated today using “wikis” and how this is affecting the “digital native” generation’s thinking⁹. Why is this so important for our discussion? The main reason is that this approach tends to transform the “consumer” of a particular information/service provided via Internet into a “prosumer”, i.e. a person who is simultaneously

9 *Web 2.0* refers to a perceived second generation of Web-based communities and hosted services – such as social networking sites, wikis and folksonomies – which aim to facilitate collaboration and sharing by users. The main difference between the first and the second generation of Internet platforms is that the former are mainly “repositories of information”, while the latter are “marketplaces” where people exchange and share information, meet people, discuss ideas, etc. A *digital native* is a person who has grown up with digital technology such as computers, the Internet, mobile phones and MP3. A *wiki* is a medium which can be edited by anyone with access to it, and provides an easy method for linking from one page to another. Wikis are typically collaborative websites, though there are now also single-user offline implementations.

a consumer and a producer of the information/service. Wikipedia is the most popular example of this approach, but there are many other platforms that use “collective intelligence” to develop innovative services.¹⁰

Of course, reliable statistics cannot be generated using “collective intelligence”, but this does not mean that this approach does not have a huge impact on the way in which statistics are perceived or used. If, for example, an authoritative member of a “community” spreads the information that a particular official figure (let’s say about inflation) is unreliable, it would be extremely difficult to change community members’ mind using the arguments usually used in statistical circles. Of course, the system also works to underline the validity of figures or sources. Just to highlight how this approach is typical of new Internet platforms, the developers of Wikipedia have recently proposed to build a discovery system based on “trusted user feedback from a community of users acting together in an open, transparent, public way”. In other words, the proposal is to replace Google discovery algorithms with a system based on the “recommendations” provided by users. This would represent a great challenge, but also a key opportunity, for statistical data providers, who should develop a new communication strategy to convince the whole Internet community to recommend official statistics instead of other sources.

The real question here is: are official data providers ready to engage themselves in this “new world” and therefore to invest significant resources in new forms of communication? For example, if web 2.0 platforms are a marketplace for discussion, and not just a repository of information, should not statistical institutions create discussion sites about the quality of data used in the public domain, including that of their own data? Of course, this could open a “Pandora’s box” and give ground to those who criticise official data. On the other hand it would allow statistical offices to be perceived as transparent institutions, as well as to express their criticisms on unreliable data produced by other sources. As stated by one of the Fundamental Principles of Official Statistics adopted by United Nations: Princi-

10 According to Wikipedia, “collective intelligence is a form of intelligence that emerges from collaboration and competition by many individuals” and it can be applied to several fields, such as cognition (market judgments, prediction of future economic and social events, etc.), co-ordination (collective actions, communities interactions, etc.) and co-operation (open source development, etc.). The study of collective intelligence may properly be considered a subfield of sociology, business, computer science and of mass behaviour, a field that studies collective behaviour from the level of quarks to the level of bacterial, plant, animal and human societies.

ple 4: “The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics”. This proactive approach would be certainly consistent with the idea of making the statistical agency a “knowledge builder” for the whole society, putting its unique technical capabilities at the service of the whole society, helping it to discriminate between good and bad information and thus gaining a stronger legitimacy.

OECD recent experiences

Over the last two years, the OECD has decided to experiment with new tools to make its statistics more accessible and re-usable by users, as well as to test new approaches to communicate statistics and engage people in exploring data and sharing their findings. Listed below are the actions which have been undertaken.

- In 2006 the OECD Council endorsed a new policy for dissemination of statistics, which involve the re-organisation of statistical products in three broad categories: *OECD Facts and Figures*: a series of simple tables, with commentary, aimed at non-specialists and specialists, to be freely available to all; *OECD Core Data*: up to 1000 ready-made tables, with metadata, drawn from all OECD databases, aimed at students, informed and specialist audiences, to be freely available to all; *OECD Statistics*: a portal giving access to all complete OECD databases, to be available on subscription using the free-at-the-point model¹¹. In this context, in December 2007 the OECD data warehouse OECD.Stat was made available to all users for free on the Organisation’s Statistics Portal (www.oecd.org/statistics). In May 2008 it registered half-million clicks on the “view data” button.
- The OECD is piloting the use of Adobe Flex to display statistical data graphically online. In order to ensure the portability of developments to the greater statistical community, this development is based on content in the Statistical Data and Metadata Exchange (SDMX) ISO standard¹².
- In 2007 the OECD made available the data published in its “Factbook” (a selection of more than 200 economic, social and

11 A key point of this strategy is that all statistical data and metadata need to be made available for easy reuse and reinterpretation by others, including the web 2.0 community.

12 The OECD is working with the European Central Bank (ECB) to create a Flex application that can interrogate SDMX data structure definitions and allow the user to view SDMX-ML data graphically and in tabular format.

environmental indicators) on Swivel.com, a web 2.0 platform for uploading, exploring, sharing data and disseminating insights via email, web sites and blogs. To manage OECD data, Swivel created a special label "Official Source" to distinguish data uploaded by organisations like the OECD and by individuals. A similar arrangement was also established with ManyEyes.com, run by IBM.

- In co-operation with the Gapminder Foundation (www.gapminder.com), the OECD has uploaded the "2008 Factbook" data on Trendalyzer, the software originally developed by Hans Rösling and his team. The OECD is also planning to create video clips where analysts would present "stories" about countries performances, policy reforms, etc. based on Factbook data and the use of Trandalyzer and other dynamic visualisation tools.
- In March 2008 the OECD Development Centre has launched *Wikigender* (see www.wikigender.org), the first "wiki-based" OECD initiative whose aim is to facilitate the exchange and improve the knowledge about gender-related issues around the world. A special section is devoted to statistical evidence, where "official" and unofficial data can be easily recognised and evaluated by the audience. In this respect, Wikigender serves as a pilot for the proposed development of a "wiki-progress", in the context of the Global project on "Measuring the Progress of Societies" (see www.oecd.org/oecdworldforum). In the first two months, Wikigender has 70.000 visits and the number of registered authors increased from 90 to 300.
- In April 2007 the OECD organised in Rome a conference on new approaches to visualise statistics. Then in June 2007, the first International Exhibition on "innovative tools to transform statistics into knowledge" was held during the World Forum on "Statistics, Knowledge and Policy". Finally, in May 2008 a second conference was organised in Stockholm (see www.oecd.org/oecdworldforum). All these events demonstrate the growing number of tools available to visualise statistics and bridge the gap between data and the human brain, as well as the key difference between "disseminating" and "communicating" data. On the other hand, they also confirmed the need to invest resources not only on the technical work, but especially on "storytelling", i.e. the capacity of extracting interesting stories out of data and present them in a comprehensible way to non experts.

Conclusions

In this paper we argued that the value added of official statistics depends on its capacity for creating knowledge in the whole society, not only among policy-makers. In fact, as demonstrated by public choice models, because of the power of information in our societies all individuals need statistics more than ever to make their decisions, including decisions on how to vote. At the same time, the development of a culture of “evidence-based decision making”, together with the transfer of some decisions from the State to individuals and the growing opportunities created by globalisation, has stimulated an unprecedented increase in the demand for statistics by individuals¹³. Finally, monitoring policy outcomes through statistical indicators is a common practice in a growing number of countries and at international level. As a result, citizens need more high quality statistics than ever in order to exercise their democratic rights, participate in the public debate and select the best politicians.

The development of statistical methods and ICT have reduced the cost of producing statistics, fostering the presence of new “actors” in the market of statistical information, including NGOs, private companies, lobbies, etc. But the multiplicity of sources is producing a “cacophony” in our societies, where users feel bombarded by data and find it increasingly difficult to distinguish between high and low quality statistics. Mass media love “numbers” and quote them as much as possible, without paying attention to their quality. Unfortunately, the declining trust in governments, as well as the behaviour of media and policy-makers, can affect overall trust in official statistics. The concept of “official” itself is not the most popular amongst new generations and other parts of our societies.

New ICT tools and the success of Internet are profoundly changing the way in which people, especially new generations, look for and find data. As previously referenced, according to Internet experts, 95% of those who use Google do not go beyond the first page of occurrences. Once they reach a particular site, a similar percentage

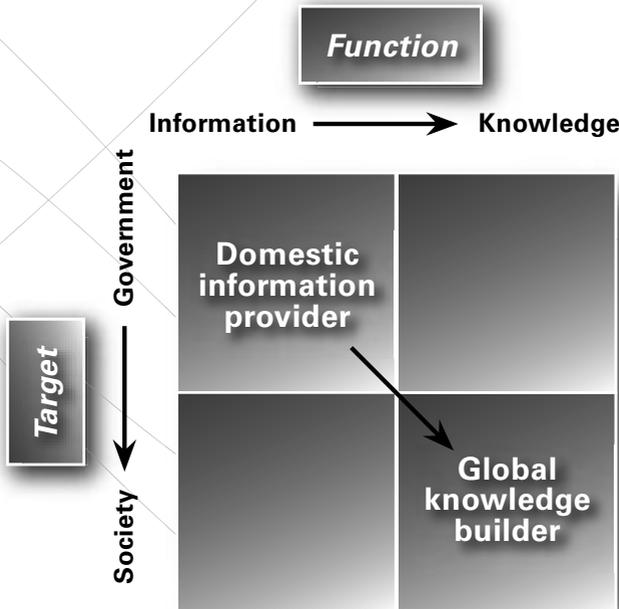
13 The seventh ISO Management Principle states that:

- *Effective decisions are based on the analysis of data and information.* The key benefits are: informed decisions, an increased ability to demonstrate the effectiveness of past decisions through reference to factual records, increased ability to review, challenge and change opinions and decisions.
- *Applying the principle of factual approach to decision making typically leads to:* ensuring that data and information are sufficiently accurate and reliable, making data accessible to those who need it, analysing data and information using valid methods, making decisions and taking action based on factual analysis, balanced with experience and intuition.

of users do not click more than three times to find what they want. If after three clicks they have not found what they are looking for, they quit the site.

The key message is that NSOs and international organisations have to become "knowledge builders" and not simply "information providers". The job of official statisticians should not be limited to producing and disseminating data, but should be about ensuring that statistics are actually used to build knowledge by all components of society, and therefore to be used in as many decision-making processes as possible. If the production of knowledge is a scale-free network (and there is some empirical evidence on this fact), where a growing number of nodes work together, NSOs should aim to be among the "big-connectors". Similarly, OECD and other international organisations should aim to be big connecting nodes at the global level. This requires innovative thinking, re-orientation of resources, alliances with new partners, revision of the skills needed to perform these new functions, changes in the legal and institutional set-ups, and better integration between national and international organisations.

Figure 5: Statistics offices from information providers towards knowledge builders



In this way, statistics can become more relevant than ever, maximising its value added in terms of the knowledge of citizens, businessmen and policy-makers. Instead of being seen as a technique, statistics could become a fundamental builder of societal knowledge, to improve decision-making at all levels. It could evolve from “statistics” (science of the state) towards “sociestics” (science of the society), to fully underpin the functioning of a democracy in the knowledge society.

References

Atkinson A. (2005), *Atkinson Review: Final Report. Measurement of Government Output and Productivity for the National Accounts*, Palgrave Macmillan, New York.

Blendon R., Benson J.M., Brodie M., Morin R., Altman D.E., Gitterman D., Brossard M. and James M. (1997), *Bridging the Gap Between the Public's and Economists Views of the Economy*, Journal of Economic Perspectives, 11 (3), pp. 105-118.

Blinder A. S. and Krueger A. B. (2004), *What Does the Public Know About Economic Policy and How Does It Know It?*, NBER Working Paper, n. 10787, September.

Caplan B. (2002), *Systematically Biased Beliefs about Economics: Robust Evidence of Judgemental Anomalies from the Survey of Americans and Economists on the Economy*, Economic Journal, 112 (April), pp. 1-26.

Carroll C.D. (2003), *Macroeconomic expectations of households and professional forecasters*, Quarterly Journal of Economics, 118 (1), pp. 269-298.

Curtin R. (2007), *What US Consumers Know About Economic Conditions*, paper presented at the OECD second World Forum on “Statistics, Knowledge and Policy”, www.oecd.org/oecdworldforum.

Doms M. and Morin N. (2004), *Consumer Sentiment, the Economy, and the News Media*, FRBSF Working Paper 2004-09, San Francisco: Federal Reserve Bank of San Francisco.

Downs A. (1957), *An Economic Theory of Democracy*, Harper & Row, New York.

Giovannini E. (2008), *The role of communication in transforming statistics into knowledge*, paper presented at the 2008 Statistics Conference of the European Central Bank, Frankfurt, May.

Giovannini E. (2007), *Statistics and Politics in a 'Knowledge Society'*, OECD Statistics Working Paper, www.oecd.org/statistics.

Holmstrom B. (1979), *Moral Hazard and Observability*, Bell Journal of Economics, 10, 74-91.

Kirchgassner G. (2005), *Why are economists different?* European Journal of Political Economy, 21 (3), pp. 543-562.

Marcuss R. D. (2007), *Question: Who Wants a Boring Statistics? Answer: Not Enough People to Keep the Statistics Healthy*, paper presented at the 56th Session of the International Statistical Institute, Lisbon.

Niskanen W. (1971), *Bureaucracy and Representative Government*, Atherton, Aldine.

Sperber D. (2005), *An Epidemiology of Representations. A Talk with Dan Sperber*, Edge, http://www.edge.org/3rd_culture/sperber05/sperber05_index.html.

Sperber D. (1996), *Explaining Culture. A Naturalistic Approach*, Oxford: Blackwell Publisher.

Van Tuinen H. (2007), *Innovative Statistics to Improve our Notion of Reality*, paper presented at the OECD second World Forum on "Statistics, Knowledge and Policy", www.oecd.org/oeecdworldforum.

PROACTIVE IS THE MAGIC WORD

*Petteri Baer, Regional Advisor, Statistical Division,
UN Economic Commission for Europe*

The Internet has made a significant contribution to improving the availability and accessibility of statistical information. Most national statistical agencies serve their users and the public by providing statistical information on-line. In the past, the main consumers of statistics were likely to be governments and ministries, but this is certainly not the case today. Statistical information is now available to anyone with access to the Internet.

Decades ago, print runs of statistical publications seldom exceeded 200 copies. For many countries, a distribution of more than thirty copies was considered to be high. Today, with the explosion of the Internet, national statisticians may have the feeling that “the whole world” is now their audience. In reality this is not the case. Efforts are still needed to achieve a significant increase in the number of users of statistical information. To put the information on the web is merely the starting point of a long process.

There is so much information out there

Publishing information on a website does not automatically equate to it being used. There are currently more than 500 million Internet hosts in the world¹. None of this guarantees that the information published on-line is actually made use of.

Even though visitors to a web site can be tracked, it is not possible to know who these visitors are. They may or may not be users of importance. Some visitors are just accidental and have opened the website by mistake. Some are not even people, but search engines, checking for new information to be indexed. In reality, although the number of viewers of a page may be high, a site could be reaching only a tiny share of potential users. There is no way of really knowing who are the users of information provided on the Internet.

1 <http://www.isc.org/index.pl?ops/ds/host-count-history.php>

Learn to know your users!

Information providers are often too quick to accept the present state of affairs. They may have made a big effort to create the website or renew its content and consider that the dissemination work is complete. It is not! Every information provider should ask themselves: "Do we know enough about our potential users, our potential customers?" and "Do we have enough information on our present users?"

If you do not know your users:

- you will not know how satisfied or dissatisfied they are
- you will not know about any unmet needs
- it will be difficult to develop quality services.

To address these issues it may be necessary to challenge the approach of the statistical agencies which focus primarily on production of statistics, not on effective use. Coverage, cost effectiveness and timeliness of production are often the most important issues for managers of statistical agencies. Also, much attention needs to be given to the methodological issues. Having put much effort into ensuring that high-quality information is produced, understanding if and how this output is used is often neglected. To some extent this is understandable. At the dissemination phase, an exhausted statistical producer may think: "I will put it out there and if people do not use it, they can only blame themselves". That kind of thinking is, however, unacceptable for a manager or director of a statistical agency.

Does the statistical agency have a role in decision-making?

It is the responsibility of top management of statistical agencies to know how decision makers perceive the value and importance of statistical services, be they in policy-making, business, research activities or education. It is short-sighted and even dangerous for a statistical agency not to invest in building and maintaining a good reputation.

Questions to be asked include:

- is building relationships with existing and potential users of statistical information an issue of strategic importance for us or, is building relations just one of many lower priority functions?

- is responsibility for public relations clearly assigned to an adequately resourced manager or group in the organization?

Proactive is the magic word

To develop better interaction with existing and new users it is vital to be proactive. Agencies must define potential user groups and describe their likely needs. The relative importance of each potential user group must be decided before developing a dissemination strategy. There is limited time and resources to provide services to all user groups and prioritization will be necessary.

Interaction with important users of services will provide valuable lessons. Through dialogue with users and analyzing feedback and customer behaviour, a better understanding of present and future needs of specific customers can be attained. By reaching a better understanding of customers' need structures it will be easier to serve them. This will allow the provision of better and more precise services to customers. A better understanding of customers will help develop a service-oriented culture and improve customer satisfaction.

By developing a customer service attitude, the value of the customer relationship will grow for both parties. The customer receives better service and the agency gets better value for time invested in building customer relations. When the relationship is mutually beneficial, customer loyalty will increase. This favours producer-customer dialogue and creates opportunities for analyzing feedback and observing customer behaviour, enabling a better understanding of present and forthcoming customer needs.

Encouraging open communication and having a learning attitude allows a wise service provider to view problems and set-backs as lessons, not failures. Lessons assist in modifying the service structures or targeting of potential customers, or both. For statistical information, customer needs are in many respects unlimited, so there is much to learn. It is important to find the right way to provide the information.

The continuous need for fresh information

When statistical services are responsive to user needs, they will be used repeatedly. Statistics are usually about observing changes over time, and in a changing world, the latest information is needed.

Therefore, it is especially wise to take good care of existing users and customers. They have probably grasped the value of statistical information for making decisions on their own activities and understand the importance of timely and fresh data. They should be served well, so that they will remain loyal customers.

A good starting point for efforts to improve statistical services is to analyze the behaviour of existing customers and find out more about their needs. It is impossible to customize or tailor services without this information. If customers are largely unknown, agencies may try to get some information on users through a pop-up questionnaire. Who has the time or the interest to reply to them? Practically nobody - or at least not too many users of importance. Up-to-date contact information is vital for communicating with customers and being able to respond to their needs with value-added services.

Basic statistical services are extremely important. They are indispensable for thousands of users who follow the main social and economic trends. However, statistical agencies can provide additional, value-added services and in doing so, it is easier to maintain and develop information on customer contacts.

Bonus services and other value-added services

Statistical agencies can and should provide additional, value-added services to accumulate contact information of their users. They should develop a mechanism for follow up and to discover which fields of statistics an individual customer shows an interest in. Such services may include analytical reports accompanying the latest statistical data and packaging different types of statistical information for particular user groups.

An additional service is to provide press releases to organizations, and to interested individuals other than just the press. A statistical agency produces press releases on numerous topics – why not make them available also for organizations and individuals outside the media? It will not harm the media – they have a much broader audience anyway. The advantages for the statistical agency are good: press releases are being reused and any responses can help establish a list of contacts with a real interest in statistical information. As a by-product the agency will accumulate information on the sphere(s) of interest of the registered contacts, information that can

be used for providing more details on other related services. Accumulating contact information also makes it possible to better target user surveys or invite customers to presentations and events.

Another relationship-building service is to send a publication catalogue, release calendar or, some other overview about forthcoming services, to customers with whom contacts have been established.

Chargeable services for more demanding clients

When providing customers with chargeable services, contact information is automatically received. It is needed for both the delivery of the service and for processing the payment. Following up on customer purchases will give the agency a better understanding of behaviour. It will identify the types of users that have more sophisticated statistical needs and are willing to pay for them. Also, information on the popularity of a specific service can be retrieved from purchase statistics on the agency's chargeable services, as can conclusions on the efficiency of related marketing campaigns. When statistical information is distributed free of charge, it is not so easy to measure the popularity of services, because many users remain anonymous.

Quite often, contact information stays in the files of individual staff members or divisions and may be maintained in very different and individual ways. Often the value of the contact information for the organization is not understood. Either no records on contacts are kept or the information is thrown away after the service has been provided.

A Customer Database brings efficiency into building relations

In the long run, it becomes necessary to bring contact information into a central database. Establishing a customer database will almost automatically improve the quality of the contact information. As structure and minimum content are defined, the information collected will be more complete and consistent. Duplication of contacts can be more easily avoided when all contacts are collected in one place. Updating can be better organized as the information is shared centrally. The value and usability of information grows

through the possibility of categorizing and grouping contacts based on needs and interests.

Organizations that maintain a customer database can do their contact building more systematically. Specific and precise targeting can be done based on categorization of the customers and potential contacts can be identified based on gaps in existing information. The agency can also enhance the coverage of contacts in different industries by comparing the contents of its customer database and its business register.

There is a wide range of software available for building a customer database or, to go one step further, for managing customer relations. In all cases the organization itself has to define which user groups and categorizations are important, be it institutional classification; classification of industries; size of customer organizations; records of purchase history or all of these. Outsiders cannot do this job— the categorization work has to be linked to the know-how of the present and planned services the organization provides.

This will involve work to be done on a number of strategically important issues, including: identification of user and customer groups; development of service concepts for the identified groups; developing good services based on these concepts; developing accessibility to and information on the services available; and, taking care that users are well informed. In other words, there is a need to proactively inform existing and potential customers about the existence of information services. This should be done in a systematic and efficient way.

To do this many activities are needed. Being service oriented will demand the organization make investments in:

- thinking
- learning
- developing
- experimenting
- testing
- new software
- equipment
- structuring and coordinating

“You will never learn to swim, if you don’t go into the water”

This work cannot be done in isolation in an office. Real contact with real customers, and users of statistical information, are needed. Otherwise the information is based on guesswork. Feedback systems and systematic research on the types and needs of users and potential users will also prove helpful. This work cannot be done without development costs, but in the long run these investments will be rewarded by growth in demand for statistical services and the growth in importance and authority of the statistical agency.

The art of turning critical feedback into improved services

Through chargeable services, the agency will receive more detailed and frequent feedback. When something is wrong, badly presented or just not good, paying customers are sure to react. With non-chargeable services that may not happen. Users of non-chargeable services in a way already know the response: “yes, our service should be better, but due to insufficient resources...” With chargeable services it is not easy to shift blame and there is greater pressure to improve performance.

More feedback will help statistical agencies to improve and develop their services. Interaction with critical customers may not always be easy, but it will certainly help in having a positive pressure to perform better.

To conclude, development of services, marketing and dissemination of statistical information are issues of strategic importance for any statistical institution. Understanding customers, marketing and building relationships are not just side functions or minor activities, they are closely linked with the reputation, future role and viability of statistical agencies.

References

- Baer, Petteri (2007), *Proactive is the Magic Word*. Presentation at the 56th ISI Conference in Lisbon 2007.
- Grönroos, Christian (2007), *Service Management and Marketing - Customer Management in Service Competition*. John Wiley & Sons.

Grönroos, Christian (2007), *In Search of a New Logic for Marketing*. John Wiley & Sons.

Gummesson, Evert (2002), *Total Relationship Marketing – From 4 Ps to 30 Rs*. Oxford, UK: Butterworth-Heinemann/The Chartered Institute of Marketing, 2nd revised edition.

Gummesson, Evert (2000), *Qualitative Methods in Management Research*, Thousand Oaks, CA: Sage, 2nd revised Edition.

Storbacka, Kaj and Lehtinen Jarmo (2001), *Customer relationship management - Creating competitive advantage through win-win relationship strategies*. Singapore: McGraw-Hill Companies.

Thygesen, Lars (1992), *Marketing Official Statistics without Selling its Soul*. Presentation at the 47th ISI Conference in Geneva.

Von Oppeln-Bronikowsky, Sibylle (1999), *Marketing and Pricing Policy*. Presentation at the 52nd ISI Conference in Helsinki.



Part 2

Good practices in country-led monitoring and evaluation systems

- Building monitoring and evaluation systems to improve government performance.
Keith Mackay, Evaluation Capacity Development Coordinator, Independent Evaluation Group, the World Bank 169
- Getting the logic right. How a strong theory of change supports programmes which work!
Jody Zall Kusek, Lead Coordinator of Global HIV/AIDS Monitoring and Evaluation Group, the World Bank
Ray C. Rist, Advisor, the World Bank, and President, International Development Evaluation Association (IDEAS) 188
- RealWorld Evaluation: conducting evaluations under budget, time, data and political constraints
Michael Bamberger, Independent consultant,
Jim Rugh, Independent international program evaluator 200
- Strengthening country data collection systems. The role of the Multiple Indicator Cluster Surveys
Marco Segone, Senior Regional Advisor, Monitoring and Evaluation UNICEF CEE/CIS
George Sakvarelidze, Monitoring and Evaluation Specialist UNICEF CEE/CIS
Daniel Vadnais, Data Dissemination Specialist UNICEF Headquarters 238

Strengthening country data dissemination systems.
Good practices in using DevInfo
Nicolas Pron, DevInfo Global Administrator, UNICEF Headquarters
Kris Oswalt, Executive Director, DevInfo Support Group
Marco Segone, Senior Regional Advisor, Monitoring and Evaluation,
UNICEF CEE/CIS
George Sakvarelidze, Monitoring and Evaluation Specialist,
UNICEF CEE/CIS.....252

Making data meaningful. Writing stories about numbers.
UNECE, Statistical dissemination and communication,
Conference of European Statisticians.....268

BUILDING MONITORING AND EVALUATION SYSTEMS TO IMPROVE GOVERNMENT PERFORMANCE

*Keith Mackay, Evaluation Capacity Development Coordinator,
Independent Evaluation Group, the World Bank*

Context

Country-led systems of monitoring and evaluation (M&E) are a concept whose time has come. A growing number of developing and transition countries and most if not all developed countries are devoting considerable attention and effort to their national M&E systems. Many do not label it as such – it may be called evidence-based policy-making, performance-based budgeting, or results-based management, for example – but at the core is an evidentiary system for public sector management that relies on the regular collection of monitoring information and the regular conduct of evaluations.

This paper first examines the various ways in which M&E systems can, and are, used to improve government performance. Key trends influencing developing countries to build or strengthen existing M&E systems are then reviewed. Next, the numerous lessons from international experience in building M&E systems are discussed, including the important role of incentives to conduct and especially to make use of M&E information. Ways to raise awareness of the usefulness of M&E, and to create incentives for the utilization of M&E, are listed. The use of such incentives can help to create demand for M&E. Finally, there is an examination of the importance of conducting a country diagnosis, to provide a shared understanding of the strengths and weaknesses of existing M&E, and, to foster a consensus around an action plan for the further strengthening of M&E.

This paper draws on a recent World Bank book written by the author that discusses all these issues in more depth. The book, *How to build monitoring and evaluation systems to support better government*, is available at:

http://www.worldbank.org/ieg/ecd/better_government.html

Use of monitoring and evaluation systems to improve government performance

M&E can measure the performance of all government policies, programmes, and projects. It can identify what works, what does not, and the reasons why. It also provides information about the performance of individual government ministries and agencies, and of managers and their staff. Additionally, it provides information on the performance of donors who support the work of governments.

The following are four main ways in which monitoring information and evaluation findings can be highly useful to government.

1. To support policy-making, especially budget decision-making (performance-based budgeting) and national planning. These processes focus on government priorities among competing demands from citizens and groups in society. M&E information can support government's deliberations by providing evidence about the most cost-effective types of government activity. Examples of this are different types of employment programmes, health interventions, or conditional cash transfer payments. M&E is widely viewed as a useful tool to help governments under fiscal stress reduce their total spending, by identifying programmes and activities which have relatively low cost-effectiveness. Performance budgeting also helps governments prioritize among competing spending proposals. In this way, it is a vehicle to help them achieve greater value for money from their spending.
2. To help government ministries in their policy development and policy analysis work, and in programme development.
3. To help government ministries and agencies manage activities at the sector, programme, and project levels. This includes government service delivery and the management of staff. M&E identifies the most efficient use of available resources; it can be used to identify implementation difficulties. For example, performance indicators can be used to make cost and performance comparisons (performance benchmarking) among different administrative units, regions, and districts. Comparisons can also be made over time which helps identify good, bad, and promising practices. This can prompt a search for the reasons for this level of performance. Evaluations or reviews are used to identify these reasons. This is the learning function of M&E, and it is often termed "results-based management".

4. To enhance transparency and support accountability relationships by revealing the extent to which government has attained its desired objectives. M&E provides the essential evidence necessary to underpin strong accountability relationships, such as of government to the Parliament or Congress, to civil society, and to donors. M&E also supports the accountability relationships within government, such as between sector ministries and central ministries, and between ministers, managers, and staff. Strong accountability, in turn, can provide powerful incentives to improve performance.

M&E is closely related to many other aspects of public sector management, as listed below.

- Budgetary tracking systems and financial reporting.
- Inter-governmental fiscal relations, including government decentralization, and the extent to which they encompass a focus on government performance.
- Accountability institutions such as national audit offices.
- Commercialization and private sector (profit and nonprofit) delivery of public services, for example, by contracting out government functions. Success in these activities requires a clear understanding of objectives and actual performance.
- Clarification and public reporting of programme goals, objectives, and the strategies necessary for achieving them.
- The setting of explicit customer service standards by service delivery agencies, and monitoring and publicizing the extent to which these are achieved. Civil service reform that focuses on personnel performance management and appraisal, including merit-based hiring, promotion, and firing. This approach recognizes the links between individual performance and project or programme performance.
- The quality of the civil service's policy advice and the extent to which this advice is evidence based (i.e. using M&E).
- Anti-corruption efforts. M&E can be used to identify the "leakage" of government funds by, for example, using public expenditure tracking surveys (PETS). Community monitoring of donor (or government) projects can also be an effective way to help curb corruption in the implementation of projects.

- Finally, M&E provides a vehicle to magnify the voice of civil society, and to put additional pressure on government to achieve higher levels of performance. Civil society (non-government organisations (NGOs), universities, research institutes, think tanks, and the media) can play a role in M&E in several ways, including both as a user and producer of M&E information.

Key trends influencing developing countries

The example of OECD countries is quite influential in the transition and developing countries. This influence extends to a number of areas of public sector management, such as customer service standards; results-based management; contracting out; privatization; performance pay; decentralization; and, performance budgeting. Most OECD governments place considerable emphasis on the four uses of M&E information: to support evidence-based policy-making (especially performance budgeting); policy development; management; and, accountability. OECD governments collectively possess a great deal of experience in this topic. There is a general understanding that for a government to improve its own performance it needs to devote substantial effort to measuring its performance. As Curristine (2005, pp. 88-89) has noted:

“Over the past 15 years, the majority of OECD governments have sought to shift the emphasis of budgeting and management away from inputs towards a focus on results, measured in the form of outputs and/or outcomes. While the content, pace, and method of implementation of these reforms varies across countries and over time, they share a renewed focus on measurable results.... In the majority of OECD countries, efforts to assess the performance of programmes and ministries are now an accepted normal part of government. Countries follow a variety of different methods to assess performance, including performance measures, evaluations, and benchmarking.”

In Latin America, the governments of at least 20 countries are currently working to strengthen their M&E systems. One influence on these governments is the demonstration effect provided by those countries with relatively advanced M&E systems, including Chile; Colombia; Mexico; and, Brazil. Related to this is a common set of economic and social pressures in Latin America. These pressures are the continuing macroeconomic and budgetary constraints; dis-

satisfaction that growth in government spending in the social sectors has not been matched by commensurate increases in the quality and quantity of services provided; continuing pressures to improve and extend government service delivery and income transfers; and, growing pressures for government accountability and for “social control” (i.e. clearer accountability of governments to ordinary citizens and to the congress).

In Eastern Europe an additional influence is seen. Countries which have joined the European Union or are candidate countries are required to strengthen their M&E systems. This is providing further impetus to the trend.

In poorer countries, initiatives of international donors such as the World Bank are also influential. The international debt relief initiative for heavily indebted poor countries has required, as a form of donor conditionality, the preparation of poverty reduction strategy papers (PRSPs) by the countries. These are to include an analysis of each country's M&E system, in particular, the adequacy of available performance indicators. PRSPs focus on the extent of the country's success in its poverty-reduction efforts to meet the Millennium Development Goals. However, most poor countries have found it difficult to strengthen their monitoring systems in terms of data production, and especially in terms of data utilization.

At the same time, there are strong accountability pressures on international donors themselves, to demonstrate results from the billions of dollars in aid spent each year, and to place more emphasis on M&E. For the World Bank, for example, these pressures have led to its results agenda. This results agenda requires that the Bank's country assistance strategies be focused firmly on the extent to which results are actually achieved, and on the Bank's contribution to them. Another donor trend is a somewhat changing emphasis in the loans made. This change is a move away from narrowly defined projects and toward programmatic lending. This entails provision of block funding, which is, in effect, broad budget support. The absence of clearly defined project activities, and outputs from such lending, also requires a focus on country results, or outcomes, of development assistance. This in turn requires a greater reliance on country systems for national statistics, and for M&E of government programmes.

Donors are working to share their experience, and that of developing countries, in the Managing for Development Results Initiative, which promotes better measurement, monitoring, and manage-

ment for results. This initiative has led to an ambitious programme of activities, including the preparation of a growing collection of resource materials and case studies, from developing countries, concerning the application of M&E and performance management at the national, sector, programme, and project levels.¹

Multilateral donors who are now heavily engaged in providing support at the country and regional levels to build government M&E systems include the African Development Bank; Asian Development Bank;² Inter-American Development Bank; and, the World Bank.³ A number of bilateral donors are also active in this area. One such is the United Kingdom's Department for International Development (DFID), which has had a particular focus on poverty monitoring systems and the use of performance information to support the budget process.

One final trend influencing the focus on M&E is the growth in the number and membership of national, regional, and global evaluation associations. In Africa, for example, there are now 16 national associations. There are also several regional associations, such as the International Programme Evaluation Network in the Commonwealth of Independent Countries (former Soviet Union countries); the African Evaluation Association; and, in Latin America, Preval and, the new regional association, ReLAC. At the global level there is the International Organisation for Cooperation in Evaluation, and the International Development Evaluation Association. These associations reflect, in part, the growing interest in M&E and the growing number of individuals working in this field. Such communities of practice have the potential to influence the quality of M&E work and thus to facilitate the efforts of governments to strengthen their M&E systems. Some national associations, such as the one in Niger (RenSE), have involved close collaboration among academics, consultants, government officials, and donor officials. This growth has the potential to spread awareness and knowledge of M&E among government officials, and so, to increase demand for it.

-
- 1 These materials are available at: <http://www.mfdr.org/>
 - 2 https://wpqp1.adb.org/QuickPlace/cop-mfdr/Main.nsf/h_Toc/8d074f8d6f17b0484825712b0028d2fb/?OpenDocument
 - 3 See for example <http://www.worldbank.org/ieg/ecd/>

Lessons from experience in building monitoring and evaluation systems

There is a growing literature on country experience in building government M&E systems (see, for example, Mackay (2007) and the references there). This literature confirms that there is broad agreement among experts in this area about the key lessons. These are as follows.

1. **Substantive demand from the government is a prerequisite to successful institutionalization.** An M&E system must produce monitoring information and evaluation findings which are judged valuable by key stakeholders; are used to improve government performance; and, which will ensure the funding and continuation of the M&E system. Achieving real demand for M&E is not easy. An important barrier can be a lack of knowledge about what M&E actually encompasses, particularly where the buy-in of key officials is necessary before a lot of effort is put into M&E.

The way around this conundrum is to increase awareness of M&E, in particular, its range of tools, methods, and techniques and, its potential uses. Demand can be increased once key stakeholders in a government begin to understand it better; are exposed to examples of highly cost-effective monitoring systems and evaluation reports; and, when they are made aware of other governments which have set up M&E systems which they value highly. It can also be persuasive to point to the growing evidence of very high returns to investment in M&E.

The supply side is also important including provision of M&E training, manuals, and procedures and the identification of good M&E consultants for example. M&E expertise is certainly necessary if reliable M&E information is to be produced. Those who view M&E in technocratic terms as a stand-alone technical activity tend to focus only on these issues. However, the supply side of producing M&E information is less important than demand. If demand for M&E is strong, then it can be relatively straightforward to improve supply in response, but the converse does not hold.

2. **Incentives are an important part of the demand side.** There need to be strong incentives for M&E to be done well and, in particular, for M&E information to be actually used. Simply having M&E information available does not guarantee use, whether by

programme managers, or by budget officials responsible for advising on spending options, or by a Congress responsible for accountability oversight. This underscores the dangers of a technocratic view which sees M&E as a set of tools with inherent value.

3. **Start with a diagnosis of what M&E functions currently exist** and their strengths and weaknesses, on both the demand and supply sides, when strengthening a government M&E system. The extent of actual utilization of M&E information must be identified, as well as the particular ways in which it is being used. Such diagnoses are themselves a form of evaluation. They are useful for the information and insights they provide, and also because they can be a vehicle for raising the awareness of the importance of M&E and the need to strengthen it.
4. **Find a powerful champion.** This can be a powerful minister or senior official who is able to lead the push to institutionalize M&E; to persuade colleagues about its priority; and, to devote significant resources to create an M&E system. A champion needs to have some understanding of M&E, in terms of tools and methods, and an appreciation of its potential usefulness for government. Government champions have played important roles in the creation of some of the more successful government M&E systems, such as those of Chile, Colombia, and Australia.
5. **Stewardship by a capable ministry.** This related feature of successful government M&E systems is stewardship to drive the design, development, and management of an M&E system. In many developed and upper middle-income countries this has meant the finance ministry. It certainly helps to have the institutional lead of an M&E system close to the center of government, for example, a president's office or a budget office (Bedi and others 2006).

In some countries, capable sector ministries have set up strong M&E systems. A notable example is in Mexico, where the Secretariat for Social Development (SEDESOL), a capable and respected ministry, manages an M&E system that emphasizes both qualitative and impact evaluations. These have included the well-known impact evaluations of the *Progresá* programme. Although expensive, these have been highly influential on the government. The programme now covers some 21 million beneficiaries, and the evaluation can be viewed as having been very cost-effective. Governments in other countries find such

examples of highly influential evaluations to be quite persuasive in relation to the potential usefulness of evaluation, and the merits of setting up a sound M&E system.

The success of M&E in SEDESOL has also helped persuade the powerful finance ministry and the comptroller's office to join the national evaluation council to create a whole-of-government M&E system. This indicates the powerful demonstration effect a successful sector agency can have.

6. A common mistake is to over-engineer an M&E system.

This is more readily evident with performance indicators. For example, Colombia's M&E system, SINERGIA, had accumulated 940 performance indicators by 2002. This number was unwieldy for the government's uses of the information for accountability purposes. It has subsequently been reduced to around 500. The appropriate number of performance indicators also depends on the number of government programmes and services and on the type of performance indicator. Senior officials would tend to make use of high-level strategic indicators such as outputs and outcomes. Line managers and their staff, in contrast, would tend to focus on a larger number of operational indicators that target processes and services.

7. The need to build reliable ministry data systems. A problem in African countries, and perhaps in some other regions, is that although sector ministries collect a range of performance information, the quality of data is often poor. Data are poor partly because they aren't being used; and they're not used partly because their quality is poor. In such countries there is too much data, not enough information. So, this lesson for the institutionalization of a government M&E system is to build reliable ministry data systems to help provide the raw data on which M&E systems depend. Data verification and credibility is partly a technical issue of accuracy, procedures, and quality control. Related to this issue of technical quality is the need for data to be potentially useful, for it to be available on a timely basis, easy to understand, consistent over time, and so forth.

8. Utilization is the measure of success of an M&E system.

The objective of government M&E systems is never to produce large volumes of performance information, or a large number of high-quality evaluations per se. This would reflect a supply-driven approach to an M&E system. Utilization is the measure of success.

9. **Provision of training in a range of M&E tools, methods, approaches, and concepts.** For an M&E system to perform well, it is necessary to have well-trained officials or consultants who are highly skilled in M&E. Thus, most capacity-building plans place considerable emphasis on provision of training in a range of M&E tools, methods, approaches, and concepts. Governments that contract out their evaluations also need to ensure that their officials are able to oversee and manage evaluations. They also need to understand the strengths and limitations (the relative cost-effectiveness) of various types of M&E.
10. **The structural arrangements of an M&E system are important** from a number of perspectives. One is to ensure the objectivity, credibility, and rigor of the M&E information produced by the system. On the data side, governments can rely on external audit committees to verify data. Some rely on the national audit office. Some rely principally on internal ministry audit units. However, some have no audit strategy. On the evaluation side, issues of objectivity and credibility are particularly important. Most Latin American countries deal with this by contracting-out evaluations to external bodies such as academic institutions and consulting firms. This achieves a certain 'distance' between the evaluators and the entities being evaluated, and this has advantages and disadvantages. In contrast, most OECD governments rely on sector ministries to conduct evaluations themselves, although this raises questions about the reliability of self-evaluations.
11. **Building an M&E system is a long-hall effort requiring patience and persistence.** This is the experience of countries that have built a government M&E system. It takes time to create or strengthen data systems; to train or recruit qualified staff; to plan, manage, and conduct evaluations; to build systems for sharing M&E information among relevant ministries; and, to train staff to use M&E information in their day-to-day work, whether that involves programme operations or policy analysis and advice. A handful of countries have been able to create well-functioning evaluation systems (in terms of the quality, number and utilization of the evaluations) within four or five years. In others it has taken more than a decade.
12. **Most countries with well-performing M&E systems have not developed them in a linear manner** according to a set plan. Instead, incremental and even piecemeal approaches seem to be common. One reason for this is the need to make mid-

course corrections as the progress, or lack of progress, with particular M&E initiatives becomes evident. External factors such as a change of government can alter the direction of an M&E system and also, lead to it being significantly strengthened or substantially run down or even abandoned.

13. **The value of regularly evaluating an M&E system.** The frequency of mid-course corrections as M&E systems are being built leads to this additional lesson from experience. Unsurprising, the objective of regular evaluation of the system is to find out what is working, what is not, and why. Such evaluations provide the opportunity to review both the demand and the supply sides of the equation, and to clarify the extent of actual utilization of M&E information, as well as the particular ways in which it is being used.

Incentives for conducting and using monitoring and evaluation. How to create demand

The importance of the demand side has already been noted. However, achieving strong demand within a country is not easy. Having examples of other countries (such as Chile, Colombia, and a number of OECD countries) which have invested the effort necessary to build a well-functioning M&E system, can be enormously influential in creating interest in M&E and building demand for it. Illustrating the cost-effectiveness of individual evaluations conducted in other countries can also persuade decision-makers about the merits of M&E. Some countries, such as Egypt, have developed a good understanding among key government ministers of the potential benefits of M&E. Yet efforts to institutionalize M&E in Egypt have been substantially frustrated by mid-level officials who did not buy into this vision of an M&E system.

The key issue here is the need to ensure there are sufficiently powerful incentives within a government to conduct M&E and to a good quality standard, and to use M&E information intensively. A public sector environment in which it is difficult for managers to perform to high standards and to perform consistently is hostile to M&E. Managers can do little more than focus on narrowly defined day-to-day management tasks. They are not willing to be held accountable for performance if they do not have some surety of the resources available to them or, if they do not have substantial control over the outputs of their activities. In this environment, M&E is understand-

ably seen by managers as probably unfair to them, and as a threat rather than an aid.

The nature of incentives for M&E also depends on how a country envisages using M&E information, whether for the learning function of M&E; or, primarily, for accountability purposes; or, as a tool for performance budgeting; or, if M&E is intended as a tool to support evidence-based policy formulation and analysis. While most countries would claim all these potential uses of M&E information to be important, it is usually the case that one or two predominate. Each of these intended uses of M&E involves different sets of stakeholders and thus incentives to drive the system.

Three types of incentive are presented in Box 1: carrots, sticks, and sermons. Many of these incentives have been used to help institutionalize M&E in developed and developing country governments. Carrots provide positive encouragement and rewards for conducting M&E and utilizing the findings. They include, for example, public recognition or financial incentives to ministries that conduct M&E. Sticks include prods or penalties for ministries or individual civil servants who fail to take performance and M&E seriously. These may include financial penalties for ministries which fail to implement agreed evaluation recommendations. Finally, sermons include high-level statements of endorsement and advocacy concerning the importance of M&E. They also include efforts to raise awareness of M&E and to explain to government officials what's in it for them.

Box 1: Incentives for conducting and using M&E: carrots, sticks, and sermons

Carrots	Sticks	Sermons
<ul style="list-style-type: none"> • Awards or prizes – high-level recognition of good or best practice evaluations or of managing for results • Provision of additional funding to ministries to conduct M&E 	<ul style="list-style-type: none"> • Enact laws, decrees, or regulations mandating the planning, conduct, and reporting of M&E • Highlight poor quality evaluation planning, data systems, performance indicators, M&E techniques, M&E reporting 	<ul style="list-style-type: none"> • High-level statements of endorsement by president, ministers, heads of ministries, deputies, and so forth • Awareness-raising seminars/workshops to demystify M&E, provide comfort about its feasibility, and to explain what's in it for participants

Carrots	Sticks	Sermons
<ul style="list-style-type: none"> • Conduct regular “How Are We Doing?” team meetings (managers and staff) to clarify objectives, review team performance, and identify ways to improve it • Assistance to programme areas in conduct of M&E – via help-desk advice, manuals, free training, etc. This makes it easier (reduces the cost) to do M&E and to use the findings • A government-wide network of officials working on M&E. This helps provide identity and support to evaluators (who often feel isolated within each ministry/entity) • Careful knowledge management of evaluation findings – e.g., providing easily understood executive summaries targeted to key audiences • Provision of budget-related incentives to ministries/agencies to improve performance • Greater management autonomy provided to programmes performing well 	<ul style="list-style-type: none"> • Withhold part of funding from ministries/agencies that fail to conduct M&E • Regularly publish information on all programs’ objectives, outputs, and service quality. Performance comparisons are particularly effective in highlighting good performers and embarrassing poor performers • Highlight adverse M&E information in reports to Parliament/Congress and disseminate widely. This can be politically sensitive and overly embarrassing to government • Set challenging but realistic performance targets – stretch targets – which each ministry, agency, and programme manager is required to meet • Require performance exception reporting where targets not met – requires programme areas to explain poor performance (Colombia) 	<ul style="list-style-type: none"> • Use of actual examples of influential M&E to demonstrate its utility and cost-effectiveness • Piloting of some rapid evaluations and impact evaluations to demonstrate their usefulness • Conferences/seminars on good practice M&E systems in particular ministries and in other countries to demonstrate what M&E systems can produce • Advocacy for government M&E on the part of multilateral and bilateral donors in their loans – this highlights and endorses M&E

Carrots	Sticks	Sermons
<ul style="list-style-type: none"> • Output- or outcome-based performance triggers in World Bank and other donor loans to governments • Performance contracts / pay for civil servants 	<ul style="list-style-type: none"> • Penalize non-compliance with agreed evaluation recommendations • Involve civil society in M&E of government performance, e.g. using citizen report cards, to stimulate better performance and accountability 	

The importance of country diagnosis

There is no single best approach to a national or sector M&E system. The particular approach a country should use depends on the actual or intended uses of the information such a system will produce. As discussed earlier, these uses range from assisting resource-allocation decisions in the budget process, to helping prepare national and sector planning, to aiding ongoing management and delivery of government services, to underpinning accountability relationships.

Efforts to build or strengthen government M&E systems clearly need to be tailored to the needs and priorities of each country. Conducting a diagnosis of M&E activities is desirable because it can guide the identification of opportunities for institutionalizing M&E. A formal diagnosis helps identify a country's current strengths and weaknesses in terms of the conduct, quality, and utilization of M&E. Additionally, a diagnosis is invaluable in providing the basis for preparing an action plan. The action plan should be designed according to the desired future uses of monitoring information and evaluation findings.

A diagnosis can be conducted by government or donors, or it may be desirable jointly. The process of conducting a diagnosis provides an opportunity to get important stakeholders within government, particularly senior officials in the key ministries, to focus on the issue of institutionalizing an M&E system. For most if not all developing countries, there will already be a number of M&E activities and systems. But a common challenge is a lack of coordination or

harmonization between them. This can result in significant duplication of effort. A diagnosis that reveals such problems can provide a stimulus to the government to address the problems. By providing a shared understanding of the nature of the problems, it can also help foster a consensus on what is needed to overcome the problem.

In Uganda, for example, the finding that there were 16 M&E sub-systems in existence raised strong concerns among senior officials. Their response led to a decision to create a national, integrated, M&E system to address the problems of harmonization and excessive demands on the suppliers of monitoring information in sector ministries and agencies and at the facility level.

A diagnosis also provides a baseline for measuring a country's progress over time; it is a long-haul effort to build and sustain both demand and supply for M&E. In this environment, it is important to regularly monitor and evaluate the M&E system itself, just as any area of public sector reform should be regularly assessed. Some aspects of an M&E system are amenable to regular monitoring, such as the number of evaluations completed or the extent to which their recommendations are implemented. Other aspects may require more in-depth evaluation from time to time, such as the extent of utilization of M&E information in budget decision making, or the quality of monitoring data. Thus, a diagnosis is a type of evaluation and can identify the degree of progress achieved and any necessary mid-course corrections.

A diagnosis of M&E would be expected to map out a number of key issues as highlighted in Box 2.

Box 2: Key issues for a diagnosis of a government's M&E system

1. Genesis of the existing M&E system – Role of M&E advocates or champions; key events that created the priority for M&E information (for example, election of reform-oriented government, fiscal crisis).
2. The ministry or agency responsible for managing the M&E system and planning evaluations – Roles and responsibilities of the main parties to the M&E system, for example, finance ministry, planning ministry, president's office, sector ministries, the Parliament or Congress; possible existence of several, uncoordinated M&E systems at the national and sector levels; importance of federal/state/local issues to the M&E system.
3. The public sector environment and whether it makes it easy or difficult for managers to perform to high standards and to be held accountable

for their performance – Incentives for the stakeholders to take M&E seriously, strength of demand for M&E information. Are public sector reforms under way that might benefit from a stronger emphasis on the measurement of government performance, such as a poverty-reduction strategy, performance budgeting, strengthening of policy analysis skills, creation of a performance culture in the civil service, improvements in service delivery such as customer service standards, government decentralization, greater participation by civil society, or an anticorruption strategy?

4. The main aspects of public sector management that the M&E system supports strongly – i) Budget decision making, (ii) national or sector planning, (iii) me management, and (iv) accountability relationships (to the finance ministry, to the president's office, to Parliament, to sector ministries, to civil society).
 - Actual role of M&E information at the various stages of the budget process: such as policy advising and planning, budget decision making, performance review and reporting; possible disconnect between the M&E work of sector ministries and the use of such information in the budget process; any disconnect between the budget process and national planning; opportunities to strengthen the role of M&E in the budget.
 - Extent to which the M&E information commissioned by key stakeholders (for example, the finance ministry) is used by others, such as sector ministries; if not used, barriers to utilization; any solid evidence concerning the extent of utilization by different stakeholders (for example, a diagnostic review or a survey); examples of major evaluations that have been highly influential with the government.
5. Types of M&E tools emphasized in the M&E system: regular performance indicators, rapid reviews or evaluations, performance audits, rigorous, in-depth existence of impact evaluations; scale and cost of each of these types of M&E; manner in which evaluation priorities are set – focused on problem programmes, pilot programmes, high expenditure or -visibility programmes, or on a systematic research agenda to answer questions about programme effectiveness.
6. Who is responsible for collecting performance information and conducting evaluations (for example, ministries themselves or academia or consulting firms); any problems with data quality or reliability or with the quality of evaluations conducted; strengths and weaknesses of local supply of M&E; key capacity constraints and the government's capacity-building priorities.
7. Extent of donor support for M&E in recent years; donor projects that support M&E at whole-of-government, sector, or agency levels – Provision of technical assistance, other capacity building and funding for the conduct of major evaluations, such as rigorous impact evaluations.
8. Conclusions: Overall strengths and weaknesses of the M&E system; its sustainability, in terms of vulnerability to a change in government, for example, how dependent it is on donor funding or other support; current plans for future strengthening of the M&E system.

The purpose of a diagnosis is more than a factual stocktaking. It requires careful judgment concerning the presence or absence of the success factors for building an M&E system. It is therefore important to understand the strength of the government's demand for M&E information and whether there is an influential government champion for M&E.

It is also important to know if there are barriers to building an M&E system, such as lack of genuine demand and ownership; lack of a modern culture of evidence-based decision making and accountability (due, in some countries, to issues of ethics or corruption); lack of evaluation, accounting, or auditing skills; or, poor quality and credibility of financial and other performance information. This understanding naturally leads to the preparation of an action plan to strengthen existing M&E systems or to develop a new system entirely.

Although the preceding issues are largely generic to all countries, it is necessary to adjust the focus according to the nature of the country. Middle-income or upper middle-income countries might well possess a strong evaluation community, centered in universities and research institutes. However the supply of evaluation expertise would be much weaker in many of the poorest countries, for example, those that prepare poverty-reduction strategies. Also, poorer countries are likely to have a strong focus on poverty-monitoring systems, in particular, and are likely to experience much greater difficulties in coping with multiple, unharmonized donor requirements for M&E. Donor pressure is often the primary driver of government efforts to strengthen M&E systems, and the strength of country ownership of these efforts may not be strong.

A question that is often asked is: how long should it take to conduct an M&E diagnosis. There is no simple answer to this question. It all depends on the purposes for which a diagnosis is intended, the range of issues under investigation, and the available time and budget. In some cases a week-long mission to a country has provided a sufficient starting point for a broad understanding of the key issues facing a government interested in strengthening its M&E functions. At the other end of the spectrum is a more formal, detailed, and in-depth evaluation of a government evaluation system, such as the one the Chilean government commissioned the World Bank to undertake. The Chile evaluation involved a team of seven people working for many months.

Other issues may need to be investigated in-depth, such as the quality and credibility of monitoring information and of the sector

information systems which provide this information. Another possible issue is the capacity of universities and other organisations that provide training in M&E. Such training is a common element of action plans to help institutionalize M&E.

Depending on the issues to be addressed in a diagnosis, it might well be necessary to assemble a team of experts with a range of backgrounds. A team might therefore include individuals with expertise in some or all of the following: the management of a government M&E system; performance indicators and systems; statistical systems; evaluation; public sector management reform; and, performance budgeting.

Most diagnoses are neither very rapid nor very time consuming or in-depth; they fall between these two extremes. Nevertheless, a sound diagnosis does require considerable care. The expertise and quality of judgment of those who prepare the diagnosis is crucial.

Conclusions

The focus of this paper is on the key lessons for governments in their efforts to build, strengthen, and fully institutionalize their M&E systems, not as an end in itself but to achieve improved government performance. A consistent message argued here is that the bottom-line measure of “success” of an M&E system is utilization of the information it produces. It is not enough to create a system that produces technically sound performance indicators and evaluations. Utilization depends on the nature and strength of demand for M&E information, and this in turn depends on the incentives to make use of M&E. Some governments in developing countries have a high level of demand for M&E; in others the demand is weak or lukewarm. For these latter countries, there are ways to increase demand by strengthening incentives.

One of the key lessons to incorporate into building an M&E system is the importance of conducting a country diagnosis of M&E. It can provide a sound understanding of M&E activities in the government, the public sector environment and opportunities for using M&E information to support core government functions. Such a diagnosis is an important building block for preparing an action plan. A diagnosis can also be a vehicle for ensuring that key government and donor stakeholders have a shared understanding of the issues and of the importance of strengthening M&E.

References

- Bamberger, Michael. Mackay, Keith and Ooi, Elaine. (2005), *Influential Evaluations: Detailed Case Studies*. Washington, DC: World Bank.
- Bedi, Tara. Coudouel, Aline. Cox, Marcus . Goldstein, Markus and Thornton, Nigel . (2006), *Beyond the Numbers: Understanding the Institutions for Monitoring Poverty Reduction Strategies*. Washington, DC: World Bank.
- Curristine, Teresa. (2005), *Performance Information in the Budget Process: Results of the OECD 2005 Questionnaire*. OECD Journal on Budgeting 5 (2): 87–131.
- Hatry, Harry P. (2006), *Performance Measurement: Getting Results*, 2nd ed. Washington, DC: The Urban Institute Press.
- Hauge, Arild. (2003), *The Development of Monitoring and Evaluation Capacities to Improve Government Performance in Uganda*. No. 10 of Evaluation Capacity Development Working Paper Series. Washington, DC: World Bank.
- IEG (Independent Evaluation Group). (2004), *Monitoring and Evaluation: Some Tools, Methods and Approaches*, 2nd ed. Washington, DC: World Bank.
- Mackay, Keith. (2007), *How to Build M&E Systems to Support Better Government*. Washington, DC: World Bank.
- Mackay, Keith. (1998), *Evaluation Capacity Development: A Diagnostic Guide and Action Framework*. No. 6 of Evaluation Capacity Development Working Paper Series. Washington, DC: World Bank.
- Mackay, Keith, Gladys Lopez-Acevedo, Fernando Rojas, Aline Coudouel, and others. (2007), *A Diagnosis of Colombia's National M&E System, SINERGIA*. No. 17 of Evaluation Capacity Development Working Paper Series. Washington, DC: World Bank.
- May, Ernesto, David Shand, Keith Mackay, Fernando Rojas, and Jaime Saavedra, eds. (2006), *Towards Institutionalizing Monitoring and Evaluation Systems in Latin America and the Caribbean: Proceedings of a World Bank/Inter-American Development Bank Conference*. Washington, DC: World Bank.
- OECD (Organisation for Economic Co-operation and Development). (2007), *Performance Budgeting in OECD Countries*. Paris: OECD.
- OECD (Organisation for Economic Co-operation and Development). (2005), *Modernizing Government: The Way Forward*. Paris: OECD.
- Ravindra, Adikeshavalu. (2004), *An Assessment of the Impact of Bangalore Citizen Report Cards on the Performance of Public Agencies*. No. 12 of Evaluation Capacity Development Working Paper Series. Washington, DC: World Bank.
- Sandoli, Robert L. (2005), Budgeting for Performance in the United States Using the Program Assessment Rating Tool (PART). Presented at a World Bank-Korea Development Institute conference, Improving the Public Expenditure Management System, December 8.
- Toulemonde, Jacques. (2005), *Incentives, Constraints and Culture-Building as Instruments for the Development of Evaluation Demand*. In Building Effective Evaluation Capacity: Lessons from Practice, ed. Richard Boyle and Donald Lemaire, 153-174. New Brunswick, NJ: Transaction Publishers.

GETTING THE LOGIC RIGHT. HOW A STRONG THEORY OF CHANGE SUPPORTS PROGRAMMES WHICH WORK!

Jody Zall Kusek, *Lead Coordinator of Global HIV/AIDS
Monitoring and Evaluation Group, the World Bank*

Ray C. Rist, *Advisor, the World Bank, and President,
International Development Evaluation Association (IDEAS)*

Introduction

A vital restaurant area in an urban community, called Ninaville, has been experiencing a recent rash of burglaries. A young couple was even attacked in an adjacent parking garage. Restaurant-goers are also increasingly being harassed on the street by local gangs. As a result, fewer people are frequenting this once-popular eating area. Revenues have plunged and employees are being let go. Over a relatively short time, the area has been transformed from a popular gathering place to one where few venture after dark. Streets are in disrepair, buildings are left vacant, and other fixtures left abandoned. Fortunately, there are funds set aside by the state government for urban renewal in five communities. The Government intends to develop and issue substantial new policies and guidelines for zoning of businesses and residential areas in the State. However, they believed that they need a stronger evidence basis from which to develop the new policy. Thus, they hoped that the five urban renewal projects would serve as pilots to help them understand how to effectively develop the new policy. Ninaville would like to submit a proposal to use the funds to help restore the once-thriving restaurant area. The funds would be made available for three years, with twice yearly reporting on renewal progress in order to maintain funding eligibility.

To achieve the overall goal of restoring security in the restaurant area many questions need to be answered. Are people not coming because they do not feel safe? If so what would make them feel safer? Would hiring more policemen work? Would routing out areas where the gangs congregated be the appropriate thing to do? What about more arrests? Perhaps people are not coming because the res-

restaurant area is no longer on a route for public transportation? What about building a pedestrian mall that would attract other shops and activities for the public? To be a successful candidate for the urban renewal funds, each community would need to develop a strong proposal that described how the funds would be used to achieve key urban renewal goals. Communities were asked to include a programme design, implementation plan, budget and timeline. The city council of Ninaville plans to hold a meeting with all interested stakeholders to identify key concerns, and objectives which they hoped would form the outline of a programme proposal.

Thinking through the logic of good programme design

The first task faced by the city council of Ninaville was to make sure that there was agreement on the nature of the problem. Some people focused on the gangs, and saw the need as to rid the community of these thugs. Others said that while the gangs were important, the real problem was loss of jobs. Others thought that the solution was to bring about economic well-being so that the entire community could benefit. They felt that while once a thriving community, there were many factors besides the gangs and crime that prevented the community from being all it could be. The City Council felt that it was important to outline a set of assumptions that were the likely cause of the recent problems and to identify key risks that had to be managed to achieve renewal of the community.

Ninaville is on the right track. Often referred to as the Programme Logic Model or the Theory of Change approach, a good programme theory is needed to think through the assumptions which will guide an organization (e.g. a community, government, or business), towards the design of effective programme interventions; a strong implementation plan; and, where to best spend resources. A good programme theory provides a strong rationale to: (i) get buy-in from key stakeholders; (ii) expend funds; (iii) suggest achievable outcomes and outputs; and (iv) support scale up of pilot projects to larger and more costly projects and programmes. Ninaville recognizes that in order to compete for one of the five pilots, they have to demonstrate that they are to design and implement a strong programme that will result in positive change. They recognize they need a strong programme theory to demonstrate how the interventions they plan to fund will result in the achievement of their goals.

This discussion fits into the theme of country-led evaluations since to successfully build a strong evaluation culture in developing countries there needs to be an emphasis on how evaluation can help deliver information and analysis which strengthens programme delivery. In short, how evaluation can provide coherent and useful theories of change which countries can deploy as they seek to address the problems they have.

We have identified five questions which need to be answered when thinking through the logic of a programme, or its theory of change. This “CORAL” questionnaire aims to support programme planners in addressing the following:

- C** what is the **concern** or concerns most affecting citizens and other stakeholders?
- O** what is the **outcome** or solution sought? In other words, what would success look like?
- R** what are known or likely **risks** which will stop the programme being successfully implemented?
- A** can key **assumptions** be tested and measured with information readily available to determine what is, or is not, working?
- L** can new programme **logic** and knowledge, gained from implementing programme interventions, be regularly fed back into the programme to revise the design and implementation plan as necessary?

Can performance frameworks or log frameworks provide the basis for good design and evaluation?

In our 2004 book, “Ten Steps to a Results-Based M&E System”¹, we identify the ten steps that we believe are necessary to building and using an monitoring and evaluation system to manage to results. In our book, we present a logic model in five parts – inputs, activities, outputs, outcomes, and impacts. We explain how most programme theory is designed from inputs to outputs to impacts. This leaves out any thinking on how to design successful behav-

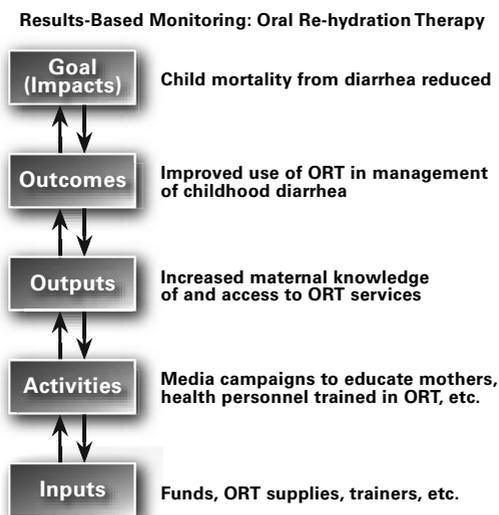
1 The authors summarized the book in an article published in the book “*Bridging the gap: The role of monitoring and evaluation in evidence-based policy making*”. The book is available – free of charge – at: <http://www.unicef.org/ceecis/resources.html>

our change and improvements in utilization rates, such as building schools and then actually measuring if children use them rather than theorizing that building 10 new schools will result in improved literacy rates of children. In short, we argue that one cannot get to impacts without first being very clear about what outcomes are to be achieved.

Over the last few years we have heard from numerous programme planners and programme evaluators on the need to further understand what is behind a good performance or logic framework. Questions such as: “how do I know that the interventions in my programme are being designed and implemented to support the programme change I am seeking” or, “how do I keep myself and my staff looking at the big picture”, are frequent. Short of undertaking expensive and often difficult evaluations, it is not easy to know the answers to these questions. However, paying more attention at the design stage will help ensure that a programme will be able to show the effective use of resources, show the links between inputs, activities, outputs and outcomes, and provide a rationale for setting up an evaluation to later test whether the theory “held” or not during implementation. Attention to the programme theory will also help assess, in the case of a programme failure, whether it was the design that failed or whether the implementation failed, or both. Thus a strong programme theory can support effort to better restructure a project to get it back on track.

Figure 1 presents a typical logic model (or results framework as they are often called), for the design of a project intended to support the achievement of reducing mortality rates for children under five years old. Most development programmes are required to include results frameworks to be eligible for international funding. These frameworks intend to demonstrate cause and effect of planned programme components by linking activities and outputs to higher order outcomes and impacts (goals). The suggestion here is that funding media campaigns to inform mothers about the importance of re-hydrating children sick with diarrhoea will ultimately increase their knowledge of its importance and thus change their behaviour towards its use. These activities are presumed causal to the eventual, or higher order, goal of reducing deaths from diarrhoea.

Figure 1: Example of logic model



Logic models, or results frameworks, make assumptions that a set of activities are causal to achieving the overall goal. Sometimes these assumptions are made based upon what is considered best practices from similar programmes, or from the findings of evaluation research about what works and why. However, in the rush to get development programmes approved by governments as well as institutional boards, projects are not always designed using valid evidence about what works and why. Assumptions are not tested, and there are no plans to manage risks likely to be encountered during implementation. In these cases, it is down to luck whether the programme theory holds or not.

When the assumptions behind a programme or project design are neither tested nor backed by published evidence, regular “testing” of the logic during implementation can help assure that results will be achieved. This requires that each output and outcome be translated into a set of key performance measures that are tracked regularly to see if the assumptions behind the project or programme are valid. A monitoring system that relies on valid and verifiable information to assess the change of each performance indicator will help determine if the project or programme is achieving planned outputs and outcomes and at what speed. Managers need to pay consistent and regular attention to the original design of the programme and, when necessary, make changes in both the design and the original

assumptions. Building the theory “as you go” requires continued feedback on what appears to be working and what is not and a willingness to make necessary changes to both the original design and assumptions.

In evaluation there is a frequently used phrase, “Weak thrust, weak effect.” This essentially points to the fact that a weakly conceived programme theory of change is not likely to produce strong results, but more likely the opposite: you will not get strong effects from weak designs. Essentially we can think of this in terms of a two by two table (figure 2) showing strong and weak designs across the top and strong and weak implementation along the side. In only one of the four boxes is there both strong theory and strong implementation – which is what it takes for a successful policy or programme or project. Any of the other three boxes represent a problem. Box 2 with a weak design and strong implementation does not provide strong results any more than box 3, with strong design and weak implementation. Finally box 4 is obvious – weak designs and weak implementation can only produce failure. The point of this is that treating design considerations carefully is essential to any opportunity for a successful programme. It cannot happen any other way. A well crafted theory of change is essential for success. Stated differently, both a strong design and strong implementation are requirements if programmes, projects, or policies are to be successful. Neither alone (strong design or strong implementation) is sufficient.

Figure 2: Weak thrust, weak effect

		<i>Strength of design</i>	
		Hi	Lo
<i>Strength of implementation</i>	Hi	1	2
	Lo	3	4

The CORAL questions

Certainly there are many questions that need to be answered during both the design phase of a project and when it is implemented. To assist with this, the authors have, as noted above, developed what we call the CORAL Questionnaire as a self-assessment tool that can be used during the initial design of a new programme or project, during implementation and, to support an evaluation of how well the programme or project achieved its intended goals. In the passages below, we further describe this model.

State the problem that is of concern to key stakeholders

This is not necessarily self-evident. Different stakeholders can view a problem quite differently, and still all agree there is a problem. The challenge is one of being clear, and in agreement, on the matter of causality. Agreement on the fact that young people are dropping out of school does not automatically lead to agreement on why they are dropping out, let alone what to do about it. The same holds for our example at the beginning of this paper – why is it that the neighbourhood is in decline? Agreement on decline is not hard, but deciding on why it is so can be most contentious. So, to sort out this issue, we need questions such as:

- do we agree there is a problem?
- do we agree on how to define this problem?
- do we agree on the extent of the problem?
- do we agree on what causes this problem?

Agree on desired outcome or solution. Define what success looks like

If we want to solve our problem, we would have to agree on what a solution would look like. And as our example at the beginning of this paper demonstrates, success can appear very different to different stakeholders. For the owner of the restaurant, it would mean he or she could re-open the restaurant and again make a living; for elderly persons it might mean being able to walk outside without fear of intimidation; for young parents, it might mean being able to again take their children to the playground; and so on. The point is that success is in the eyes of the beholder. But for the evaluator, success is essentially built on the consensus of stakeholders and their view that the theory of change held true; that what was predicted

to take place took place; and, that those who had an input into the discussion on what success would look like, agree that it is what they are seeing. Success is essentially the end point in the theory of change. So, questions that address this issue of success, and what it would look like, might include:

- do we have all the definitions of success on the table?
- do we have ways of measuring success?
- do we agree on when success would be achieved? (Is any neighbourhood ever entirely crime free?)
- can we agree on which definitions of success are most relevant?
- can we articulate the causal model/theory of change on how we will get to that state of success?

Identify and manage risks to success

There are many factors or risks that can cause success not to happen. Some might be anticipated and we can plan for these; others not (the so-called “unanticipated consequences of social change”). But the fundamental point is that change cannot be completely managed and engineered as one might think could be possible with an infrastructure project. Change takes place within parameters of what are and are not acceptable. A programme might have a trajectory towards success, but it is seldom if ever precisely as was planned or initiated. Multiple circumstances such as clashing personalities of the stakeholders; changes in funding levels; loss of key staff; inability to replace those same staff; and, changes in the political climate, are but a few of a much larger number of threats to the successful completion of the project, programme, or policy. Each of these threats is a risk to the initiative. Each could be enough in the right circumstances to ensure the initiative fails.

The point about identifying and trying to manage risks, is that to ignore them pretty much means one is programming failure. Anticipating how to deal with some of the risks helps boost the prospects of success, but it is not guaranteed that being prepared to mitigate some of the risks will ensure success. The challenge is to think through and acknowledge the key risks, attempt to figure out how to address these risks, and be constantly on the look-out for emergent situations which can sabotage the whole effort. The theories of change for a programme should address the presence of these risks; note how they are going to be addressed; and, establish a monitoring and evaluation system that is flexible, nimble, and

sensitive to information on when things are starting to go wrong. Rigidities in the theory of change are harmful as are rigidities in a monitoring and evaluation system.

Questions to pose here can include:

- have we identified key risks at each stage of the theory of change that threaten the success of the initiative?
- have we decided on how to address these key risks should they emerge?
- do we believe that our monitoring and evaluation system is sufficiently nimble and sensitive to picking up data that show the effort is going off track? (Unanticipated risks are emerging.)

Test key assumptions with valid information

Assumptions are all those components of a project or programme which are presumed to hold true, to hold constant, or to hold together for the change to eventually occur. Each assumption should be stated explicitly and then examined as to whether it is likely or highly problematic, whether there is research to support it or not, and whether all the key factors, which will facilitate or hinder progress towards the desired change, have been identified within the cumulative total of all assumptions.

A theory of change needs to be continually tested to see if the logic behind it continues to hold during programme or project implementation. To do this, one must ask key questions during design and implementation and when the programme or project is being evaluated.

A theory of change should be able to answer the following:

- is the model an accurate depiction of the programme?
- are all elements and components well defined?
- are there any gaps in the logical chain of events?
- are all elements necessary?
- is the sum total of all elements sufficient to capture the proposed change model?
- are relationships plausible?
- are relationships consistent?
- is the model realistic in terms of the change being achieved?

As described above, we need to regularly test our assumptions by measuring a set of key performance measures designed to track whether desired outputs and outcomes are being achieved. By measuring performance measures on a regular determined basis, managers and decision makers can find out whether projects, programmes and even policies are on track, off track, or even doing better than expected against the targets for performance of those indicators. This provides an opportunity to make adjustments, correct course, and gain valuable institutional and programme experience and knowledge. Ultimately, of course it increases the likelihood of achieving the desired results. In order to test the logic of a programme or project, there must be a valid source of information that can be used to measure each indicator. In accomplishing this, there are nine questions which need to be answered:

- what are the sources of the data that will be used to measure each indicator?
- what are the data collection methods?
- who will collect the data?
- how often will the data be collected?
- what are the costs to collect the data?
- what is the difficulty to collect the data?
- who will analyze the data?
- who will report the data?
- who will use the data?

It should be noted here that no theory of change can be explicit on all possible assumptions. Not all assumptions should be listed and not all assumptions can be tested. The list would be very long -perhaps stretching out with an infinite number of "if-then" statements. As the philosopher E. B. White once noted, "There is no limit to how complicated things can get, on account of one thing always leading to another." What is important is to be relatively sure of getting down with explicit statements all the key assumptions – those presenting the most risk to the programme, whether by happening or by not happening.

Feedback knowledge during implementation to redesign or improve implementation

Testing key assumptions of the theory of change will produce a continuous flow of information which will support better management of the programme or project, and provide a basis for revising (if necessary) the original design. Thus by allowing flexibility in the programme design logic, decision makers can continuously revise the theory of change if it appears that the original assumptions do not hold. This is not to suggest that poor programme or project performance, due to ineffectual implementation, is a reason for revising the programme logic. If the logic is strong, then the challenge is rightly to improve the implementation – essentially moving from box three to box one in Figure 2.

Key questions which need to be answered, to ensure that knowledge acquired during implementation is used to improve the chances that the programme or project will be successful, include:

- is there a monitoring system in place that allows continual feedback to decision makers?
- is there a champion or individual whose job it is to assess progress towards programme/project implementation?
- are implementation results discussed with key stakeholders?
- what are the opportunities to adjust the original results or performance framework, hence revising the theory of change?

Conclusion

This paper has addressed the issue of why it is important to focus on building coherent logical models so as to be explicit about: (i) what change is anticipated; (ii) what risks there are to that change ever coming into being; (iii) why a system of monitoring is necessary to capture relevant data on whether the change is emerging as planned; and, (iv) how and when relevant stakeholders will be able to decide if the initiative was a success or not. A successful project, programme or policy needs both a strong design and strong implementation. One or other of these two components, by themselves, is not sufficient to ensure success. A well crafted theory of change can help on both accounts, by clearly articulating where the initiative intends to go and, secondly, by matching monitoring data against the theory so as to tell us if the initiative is going in the right direction or not.

References

- Blamey, A. and Mackenzie, M. (2007). *Theories of Change and Realistic Evaluation*. In: Evaluation. Vol. 13, No. 4.
- Kusek, J. Z. and Rist, R. C. (2008). *Ten Steps to a Results-based Monitoring and Evaluation System*. In: Segone, M., Bridging the gap. The role of monitoring and evaluation in evidence-based policy making. UNICEF
- Kusek, J. Z., Rist, R. C. and White, E. M., (2005). *How will We Know the Millennium Development Goal Results When We See Them?* In: Evaluation. Vol. 11, No. 1.
- Kusek, J.Z. and R. C. Rist, (2004) *Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington, D.C., The World Bank.
- Mason, P. and Barnes, M. (2007). *Constructing Theories of Change*. In: Evaluation. Vol. 13, No. 2.
- Mayne, J. (2007). *Challenges and Lessons in Implementing Results-based Management*. In: Evaluation. Vol 13, no. 1.
- Mayne, J. and Rist, R. C. (2006). *Studies are not Enough: The Necessary Transformation of Evaluation*. In: The Canadian Journal of Evaluation. Vol 21, No. 3.
- Rogers, P. (2008). *Using Program Theory of Evaluate Complicated and Complex Aspects of Interventions*. In: Evaluation. Vol. 14, No. 1.
- Vassen, J. (2006). *Programme Theory Evaluation, Multicriteria Decision Aid and Shareholder Values*. In: Evaluation. Vol. 12, No. 4.
- Zapico-Goni, E. (2007). *Matching Public Management, Accountability, and Evaluation in Uncertain Contexts*. In: Evaluation. Vol. 13, No. 4.

REALWORLD EVALUATION. CONDUCTING EVALUATIONS UNDER BUDGET, TIME, DATA AND POLITICAL CONSTRAINTS¹

*Michael Bamberger, Independent consultant,
Jim Rugh, Independent international program evaluator*

The RealWorld Evaluation context

The RealWorld Evaluation (RWE) approach was developed to assist the many evaluators in both developing, transition, and developed countries, who must conduct evaluations with budget, time, data and political constraints. In one common scenario, the client (project implementing agency, national planning or finance ministry; or, international donor agency), delays contracting an evaluator until late in the project when a decision has to be made on whether to continue support to the project or programme, or possibly to launch a larger second phase. Such tardiness occurs even when evaluation has built into the original project agreement. With the decision point approaching, the funding agency may suddenly realize that it does not have solid information on which to base a decision about future funding of the project; or the project implementing agency may realize it does not have the evidence needed to support its claim that the project is achieving its objectives. An evaluator called in at this point may be told it is essential to conduct the evaluation by a certain date and to produce “rigorous” findings regarding project impact although, unfortunately, very limited funds are available and no systematic baseline data has been collected.

In other scenarios, the evaluator may be called in early in the life of the project but then finds that for budget, political, or methodological reasons, it will not be possible to collect comparison data to determine programme impact by comparing participants with non-

1 This article is adapted from the book by Michael Bamberger, Jim Rugh and Linda Mabry. *RealWorld Evaluation: Working under budget, time, data and political constraints* published by Sage in 2006. It also incorporates additional material developed by Bamberger and Rugh for training workshops that have now been offered in 15 countries. Additional materials including more extensive tables are available at www.realworldevaluation.org. The two present authors are entirely responsible for the content and interpretations presented in this chapter.

participants. In some cases, it may not even be possible to collect baseline data on the project participants themselves for purposes of analyzing progress or impact over time. Data constraints may also result from difficulties in collecting information on sensitive topics such as HIV/AIDS; domestic violence; post-conflict reconstruction; or, illegal economic activities (e.g. commercial sex workers, narcotics, or political corruption).

Determining the most appropriate evaluation design under these kinds of circumstances can be a complicated juggling act involving a trade-off between available resources and acceptable standards of evaluation practice. Often the client's concerns are more about budgets and deadlines, and basic principles of evaluation may receive a lower priority. Failure to reach satisfactory resolution of these trade-offs may also contribute to a much lamented problem: low use of evaluation results (see Chelimsky, 1994; Patton, 1997; Operations Evaluation Department, 2004 and 2005). RWE is a response to the all-too-real difficulties in the practical world of evaluation.

The pressures of conducting evaluations under budget and time constraints have often resulted in inattention to sound research design or to identifying and addressing factors affecting the validity of the findings. RWE is based on a seven-step approach, summarized in Figure 1.

Scoping the evaluation

It is important that those charged with conducting an evaluation gain a clear understanding of what those asking for the evaluation (the clients and stakeholders) are expecting – that is, the political setting within which the project and the evaluation will be implemented. It is also important to understand the policy and operational decisions to which the evaluation will contribute and the level of precision required in providing the information which will inform those decisions.

Understanding client's needs

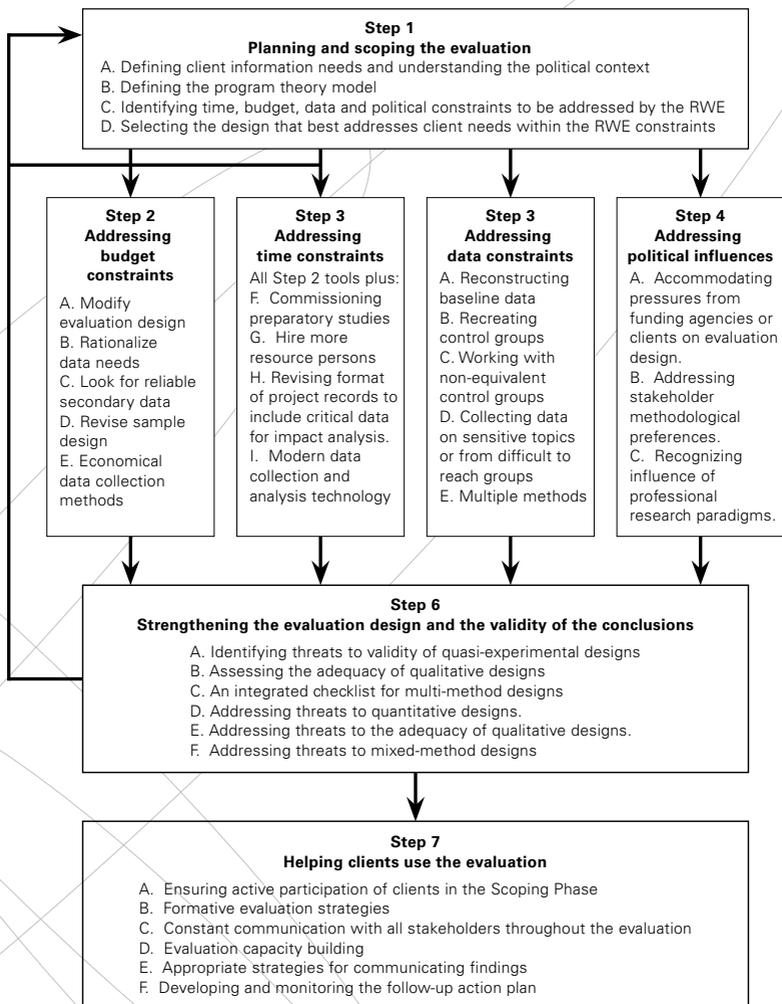
An essential first step in preparing for any evaluation is to obtain a clear understanding of the priorities and information needs of the client (the agency or agencies commissioning the evaluation,) and other key stakeholders (persons interested in or affected by the project). The timing, focus, and level of rigor of the evaluation should be determined by the client information needs and the types of decisions to which the evaluation must contribute.

The process of clarifying what questions need to be answered can help those planning the evaluation to identify ways to eliminate unnecessary data collection and analysis, hence reducing cost and time. The RealWorld evaluator must distinguish between:

- (a) information that is essential to answer the key questions driving the evaluation and,
- (b) additional questions that would be interesting to ask, if there were adequate time and resources, but which may have to be omitted given the limitations faced by the evaluation.

An important function of the scoping phase is to understand whether the lack of consultation with the groups affected by the project (including the poorest and most vulnerable groups), is due to a lack of resources or to the low priority that the client assigns to their involvement. Often, lack of time and money may be used as an excuse, so it is important for the evaluator to fully understand the perspective of the client before deciding what approach to adopt.

Figure 1: The RealWorld Evaluation [RWE] Approach



Understanding the political environment

The political environment includes the priorities and perspectives of the client and other key stakeholders, the dynamics of power and relationships between them and the key players in the project being evaluated, and even the philosophical or methodological biases or preferences of those conducting the evaluation. Table 1 lists some of the ways in which political factors can affect evaluations when they are being designed, while they are being implemented and when the findings are being presented and disseminated.

Table 1: Examples of some of the ways that political influences affect evaluations

During evaluation design	
The criteria for selecting evaluators	<p>Evaluators may be selected:</p> <ul style="list-style-type: none"> • for their impartiality or their professional expertise • for their sympathy towards the program • for their known criticisms of the program (in cases where the client wishes to use the evaluation to curtail the program) • for the ease with which they can be controlled • because of their citizenship in the country of the program's funding agency
The choice of evaluation design and data collection methods	The decision to use either a quantitative or qualitative approach or to collect data that can be put into a certain kind of analytical model (e.g. collecting student achievement or econometric data on an education program) can predetermine what the evaluation will and will not address.
<i>Example of a specific design choice:</i> Whether to use control groups (i.e. experimental or quasi-experimental design)	<p>Control groups may be excluded for political rather than methodological reasons such as:</p> <ul style="list-style-type: none"> • to avoid creating expectations of compensation • to avoid denial of needed benefits to parts of a community • to avoid pressures to expand the project to the control areas • to avoid covering politically sensitive or volatile groups. <p>On the other hand evaluators may insist on including control groups in the evaluation design in order to follow conventional practice in their profession even when they contribute little to addressing evaluation questions.</p>

The choice of indicators and instruments	The decision to only use quantitative indicators can lead (intentionally or otherwise) to certain kinds of findings and exclude the analysis of other, potentially sensitive topics. For example, issues of domestic violence or sexual harassment on public transport will probably not be mentioned if only structured questionnaires are used.
The choice of stakeholders to involve or consult	The design of the evaluation and the issues addressed may be quite different if only government officials are consulted, compared to an evaluation of the same programme in which community organizations, male and female household heads and NGOs are consulted. The evaluator may be formally or informally discouraged from collecting data from certain sensitive groups, for example by limiting the available time or budget, a subtle way to exclude difficult to reach groups.
Professional orientation of the evaluators	The choice of, for example, economists, sociologists, political scientists or anthropologists to conduct an evaluation will have a major influence on how the evaluation is designed and the findings and recommendations that ensue.
The selection of internal or external evaluators	Evaluations conducted internally by project or agency staff have a different kind of political dynamic and are subject to different political pressures compared to evaluations conducted by external consultants, generally believed to be more independent. The use of national versus international evaluators also changes the dynamic of the evaluation. For example, while national evaluators are likely to be more familiar with the history and context of the programme, they may be less willing to be critical of programmes administered by their regular clients.
Allocations of budget and time	While budget and time constraints are beyond the total control of some clients, others may try to limit time and resources to discourage addressing certain issues or to preclude thorough, critical analysis.
During implementation	
The changing role of the evaluator	The evaluator may have to negotiate between the roles of guide, publicist, advocate, confidante, hanging judge, and critical friend.

The selection of audiences for progress reports and initial findings	A subtle way for the client to avoid criticism is to exclude potential critics from the distribution list for progress reports. Distribution to managers only, excluding programme staff, or to engineers and architects, excluding social workers and extension agents, will shape the nature of findings and the kinds of feedback to which the evaluation is exposed.
Evolving social dynamics	Often at the start of the evaluation relations are cordial, but they can quickly sour when negative findings begin to emerge or the evaluator does not follow the client's advice on how to conduct the evaluation (e.g. from whom to collect data).
Dissemination and use	
Selection of reviewers	If only people with a stake in the continuation of the project are asked to review the draft evaluation report, the feedback is likely to be more positive than if known critics are involved. Short deadlines, innocent or not, may leave insufficient time for some groups to make any significant comments or to include their comments, introducing a systematic bias against these groups.
Choice of language	In developing countries, few evaluation reports are translated into local languages, thereby excluding significant stakeholders. Budget is usually given as the reason, suggesting that informing stakeholders is not what the client considers valuable and needed. Language is also an issue in the U.S., Canada and Europe where many evaluations concern immigrant populations.
Report distribution	Often, an effective way to avoid criticism is to not share the report with critics. Public interest may be at stake, as when clients have a clear and narrow view of how the evaluation results should be disseminated or used and will not consider other possible uses.

Source: RealWorld Evaluation Table 6.1

It is important to avoid the assumption that political influence is bad and that evaluators should be allowed to conduct the evaluation in the way that they know is "best" without interference from politicians and other "narrow-minded" stakeholders trying to make sure that their concerns are introduced into the evaluation. The whole purpose of evaluation is to contribute to a better understanding of policies and programmes about which people have strong and, often, opposing views. If an evaluation is not subject to any political pressures or influences, this probably means either that the topic being studied is of no consequence to anyone or that the evaluation

is designed in such a way that the concerned groups are not able to express their views. Evaluators should never assume that they are right and that stakeholders who hold different views on the key issues, appropriate methodology, or interpretation of the findings are biased, misinformed, or just plain wrong.

If key groups do not find the analysis credible, then the evaluator may need to go back and check carefully on the methodology and underlying assumptions. It is never an appropriate response to sigh and think how difficult it is to get the client to “understand” the methodology, findings and recommendations.

One of the dimensions of contextual analysis used in developing the programme theory model (see the following section) is to examine the influence of political factors. Many of the contextual dimensions (economic, institutional, environmental, and socio-cultural), influence the way that politically concerned groups will view the project and its evaluation. A full understanding of these contextual factors is essential to understanding the attitudes of key stakeholders to the programme and to its evaluation. Once these concerns are understood, it may become easier to identify ways to address the pressures placed by these stakeholders on the evaluation.

Not surprisingly, many programme evaluations are commissioned with political motives in mind, whether or not they are openly expressed. A client may plan to use the evaluation to bolster support for the programme and may consequently resist the inclusion of anything but positive findings. On the other hand, the real but undisclosed purpose the client may have had for commissioning the evaluation may be to provide ammunition for firing a manager or closing down a project or a department. Seldom, if ever, are such purposes made explicit. Different stakeholders may also hold strongly divergent opinions about a programme, its execution, its motives, its leaders, and how it is to be evaluated. Persons who are opposed to the evaluation being conducted may be able to pre-empt an evaluation or obstruct access to data, acceptance of evaluation results, or continuation of an evaluation contract.

Before the evaluation begins, the evaluator should anticipate these different kinds of potential political issues and try to explore them, directly or indirectly, with the client and key stakeholders.

Political dimensions include not only clients and other stakeholders. They also include individual evaluators, who have preferred approaches that resonate with their personal and professional back-

ground and views as to what constitutes competent, appropriate practice. Different evaluators, even those who have chosen to work together on a project, may take different stances regarding their public and ethical responsibilities. Evaluators, like everyone else, have their own personal values. However, for many evaluators, it may be more comfortable to think of the work of evaluation not as an imposition of the evaluator's values but, rather, as an impartial or objective evidence-based judgment about programme merit, shortcomings, effectiveness, efficiency, and goal achievement. The evaluators must be aware of their own perspectives (and biases) and seek to ensure that these are acknowledged and taken into consideration.

Clients may base their selection of evaluators on their reputations for uncompromising honesty, counting on those reputations to ensure the credibility and acceptance of findings. Or the choice of evaluator may be based on ideological stances the evaluator has taken that are in agreement with the client's. These decisions may be so understated as to initially go unnoticed in friendly negotiations and enthusiastic statements about the strategic importance of the proposed evaluation.

Evaluators should also be alert to the fact that political orientations of clients and stakeholders can influence how evaluation findings are disseminated and used. Clients can sometimes ignore findings they do not like and can suppress distribution by circulating reports only to carefully selected readers, by sharing only abbreviated and softened summaries, and by taking responsibility for presenting reports to boards or funding agencies and then acting on that responsibility in manipulative ways. Clients have been known to give oral presentations and even testimony that distort evaluation findings, to take follow-up activities not suggested by, and even contraindicated by, evaluation reports and, to discredit evaluations and evaluators who threaten their programmes and prestige.

The wise evaluator should be aware of such realities and be prepared to deal with them in appropriate ways during the evaluation design, the implementation of the evaluation and in the presentation and use of the evaluation findings.

Defining the programme theory²

Before an evaluation can be conducted, it is necessary to identify the explicit or implicit theory or logic model that underlies the design upon which a project was based. An important function of an impact evaluation is to test the hypothesis that the project's interventions and outputs contributed to the desired outcomes, which, along with external factors that the project assumed would prevail, were to have led to sustainable impact.

Defining the programme theory or logic model is good practice for any evaluation. It is especially useful in RWE, where, due to budget, time, and other constraints, it is necessary to prioritize what the evaluation needs to focus on. An initial review of what a project did, in the light of its logic model, could reveal missing data or information that is needed to verify whether the logic was sound, and whether the project was able to do what was needed to achieve the desired impact.

If the logic model was clearly articulated in the project plan, it can be used to guide the evaluation. If not, the evaluator needs to construct it based on reviews of project documents and discussions with the project implementing agency, project participants, and other stakeholders. In many cases, this requires an iterative process in which the design of the logic model evolves as more is learned during the course of the evaluation.

In addition to articulating the internal cause-effect theory on which a project was designed, a logic model should also identify the socio-economic characteristics of the affected population groups, as well as contextual factors such as the economic, political, organizational, psychological and environmental conditions which affect the target community.

Every project is designed and implemented within a unique *setting* or *context* that includes local and regional economic, political, institutional, and environmental factors as well as the socio-cultural characteristics of the communities or groups affected by the project. The programme theory must incorporate all these factors through a *contextual analysis*. Where a project is implemented in a number of different locations, it will often be the case that performance and outcomes will differ significantly from one site to another because of the different configurations of contextual variables.

2 For a more detailed discussion of program theory models see Bamberger, Rugh and Mabry (2006) RealWorld Evaluation, Chapter 9. This includes references to other recent publications.

Customizing plans for evaluation

Those commissioning an evaluation need to consider a number of factors that should be included in the terms of reference (TOR). The client, and an evaluator (or team of evaluators) being contracted to undertake this assignment, might find the following set of questions helpful to be sure these factors are taken into consideration as plans are made for conducting an evaluation. The answers to these questions can help to focus on important issues to be addressed by the evaluation, including ways to deal with RWE constraints.

- Who asked for the evaluation? Who are the key stakeholders? Do they have preconceived ideas regarding the purpose for the evaluation and expected findings?
- Who should be involved in planning/implementing the evaluation?
- What are the key questions to be answered?
- Will this be a formative or summative evaluation? Is its purpose primarily for learning and improving, accountability, or a combination of both?
- Will there be a next phase, or will other projects be designed based on the findings of this evaluation?
- What decisions will be made in response to the findings of this evaluation? By whom?
- What is the appropriate level of rigor needed to inform those decisions?
- What is the scope/scale of the evaluation?
- How much time will be needed/available?
- What financial resources are needed/available?
- What evaluation design would be required/is possible under the circumstances?
- Should the evaluation rely mainly on quantitative (QUANT) methods, qualitative (QUAL) methods, or a combination of the two?
- Should participatory methods be used?
- Can/should there be a survey of individuals, households, or other entities?
- Who should be interviewed?

- What sample design and size are required/feasible?
- What form of analysis will best answer the key questions?
- Who are the audiences for the report(s)? How will the findings be communicated to each audience?

Staffing the evaluation economically

In this section, we address issues concerning external experts (either from another country or from a different part of the country), content area specialists, and locally available data collectors. The ideal is to compose an evaluation team that includes a good combination of persons with different experiences, skill sets, and perspectives. Where RWE constraints are faced, especially funding, compromises may have to be made in the composition of the evaluation team. Although we address each of these categories of persons separately, it is important to consider the overall combination and the effectiveness of the full evaluation team in meeting the requirements of an evaluation.

Use international consultants wisely

International consultants are usually contracted:

- because of lack of local technical expertise (inside the organization or in the local research community);
- to build up local capacity;
- to save time;
- to ensure independence and objectivity;
- to ensure expert credibility, and/or;
- because of a requirement by the funding agency.

While, if well selected and used, international consultants can significantly improve the quality of the present and future evaluations, they are also expensive and sometimes disruptive, so they should be selected and used wisely. Under RWE constraints, the goal should be to limit the use of international consultants to those areas where they are essential. Here are a few general rules for selecting and using consultants:

- Ensure that local agencies and the client are actively involved in defining the requirements for the external consultant and in the selection process.

- Consider carefully the relative merits of international and national consultants. There is often a trade-off between greater technical expertise of the international consultant and the local knowledge (and of course language ability) of the national consultant. Not using any national consultants can also antagonize the local professional community who may be reluctant to cooperate with the international expert. It is often a good idea to have an evaluation team that combines the attributes of one or more international evaluators with the right mix of local expertise.
- If an international consultant is used, give priority to candidates who have experience in the particular country and with local language skills (if required).
- For evaluations with an operational focus, avoid selecting consultants with impressive academic credentials but limited field experience in conducting programme evaluations. The purposes and requirements of programme evaluations are different than for academically oriented research.

International consultants are often not used in the most cost-effective way, either because they are doing many things that could be done as well or better by local staff, or because they are brought in at the wrong time. Here are some suggestions on ways to ensure the effective use of international consultants:

- Define carefully what the consultant is being asked to do and consider whether all these activities are necessary.
- Even when the budget is tight, try to plan sufficient time for the consultant to become familiar with the organization, the project, and settings in which it is being implemented. A consultant who does not understand the project, has not spent some time in the communities, or has not built up rapport with project staff, clients, and other stakeholders will be of very little use.
- Plan carefully at what points the consultant should be involved and coordinate ahead of time to ensure that he or she will be available when required. Get tough with consultants who wish to change the timing, particularly at short notice, to suit their own convenience. Some of the critical times to involve a consultant are these:
 - during the scoping phase when critical decisions are being made on objectives, design, and data collection methods and when agreement is being reached with the client on options for addressing time, budget, and data constraints;

- when decisions are being made on sample size and design;
 - when the results of the initial round of data collection are being reviewed and analyzed;
 - when the draft evaluation report is being prepared;
 - when the findings of the evaluation are being presented to the different stakeholders.
- Arrange for a briefing document (preparatory study) to be prepared, by agency staff or local consultants, before the international consultant starts work. This should summarize important information about the project (including compilation of key documents, including monitoring data and periodic reports), key partner agencies, and the settings where the project is located. The document, which should be prepared in coordination with the consultant (for example through an exchange of e-mail or phone calls), might also include rapid diagnostic studies in a few communities. A well-prepared document of this kind can save a great deal of time for the consultant and can initiate dialogue on key issues and priorities among clients, local researchers and stakeholders before the external consultant even arrives.
 - Consider the use of video or phone conferences so that the consultant can maintain more frequent contact with others involved in planning and implementing the evaluation. This enables the consultant to contribute at critical stages of the evaluation without having to always be physically present. In this way, the consultant can make suggestions about the sample or other stages of the design at a sufficiently early stage for it to be possible to make changes based on these recommendations. Video and phone conferences also have the advantage of flexibility, thus avoiding the extremely costly situation where, for example, a consultant flies from Europe to West Africa to participate in the project design phase, only to discover that everything has been delayed for several weeks.

Consider including content area specialists

In addition to expertise in the relevant evaluation areas (e.g., qualitative interviewing, questionnaire construction, sample design, and data analysis), it is also essential to include at least one team member with the necessary experience in the content area of the evaluation (e.g., agricultural extension, secondary education, micro-credit, health, promoting civil society, etc.). Ideally, if resources permit, the

team should include both a sector expert with experience in many different countries or programmes as well as someone with local knowledge. The school or health system in Chicago or Dushanbe will probably have many unique features (cultural, organizational, and political) which it is important to incorporate into the evaluation.

Collect data efficiently

Simplifying the plans to collect data

Data collection tends to be one of the most expensive and time-consuming items in an evaluation. Consequently, any efforts to reduce costs or time will almost inevitably involve simplifying plans for data collection. This involves three main approaches (see Table 2):

1. Discuss with the client what information is really required for the evaluation and eliminate other information in the TOR, or mentioned in subsequent discussions, which is not essential in answering the key questions driving this evaluation.
2. Review data collection instruments to eliminate unnecessary information. Data collection instruments tend to grow in length as different people suggest additional items that it would be “interesting” to include, even though not directly related to the purpose of the evaluation.
3. Streamline the process of data collection to reduce costs and time. These include the following:
 - simplifying the evaluation design (e.g. eliminating the collection of baseline data or cutting out the comparison group);
 - clarifying client information needs;
 - look for reliable secondary data;
 - reducing sample size;
 - reducing the costs of data collection, input, and analysis (e.g. use of self-administered questionnaires, using direct observation instead of surveys, using focus groups and community fora instead of household surveys, and finding cheaper data collectors).

Commission preparatory studies

It is sometimes possible to achieve considerable cost and time savings by commissioning an agency staff person or local consultant to prepare a preparatory study. This can cover these points:

- a description of the different components of the project being evaluated and how they are organized;
- basic information on the implementing agency;
- rapid diagnostic studies of the project communities and possible comparison communities;
- information on government agencies, NGOs and other organizations involved in or familiar with the project;
- recommendations on community leaders and other key informants with whom the international consultant should meet and preparation of background information on them.

Look for reliable secondary data

A great deal of time and expense can be saved if reliable and relevant secondary data can be obtained. Depending on the country and subjects, it may be possible to find records maintained by government statistical agencies or planning departments; university or other research organizations; schools; commercial banks or credit programmes; mass media; and, many sectors of civil society. Indeed, the evaluator should make use of any relevant records such as monitoring data and annual reports produced by the implementing agency itself.

Caution: never accept secondary data at face value without checking its reliability and relevance to the communities targeted by the programme being evaluated.

Collect only the necessary data

It is important to ensure that only essential information is collected. Long questionnaires and the collection of unnecessary data increases costs and time and also reduces the quality of the information required because respondents become tired if they have to answer large numbers of questions. Therefore, we recommend that all data collection instruments be carefully scrutinized to cut out information that is not relevant and essential to the purpose of the evaluation, and that very likely will never be analyzed or used.

Table 2: Strategies for addressing data constraints

Reconstructing Baseline Data		
<i>Approaches</i>	<i>Sources/Methods</i>	<i>Comments/Issues</i>
Using existing documents (secondary data)	<ul style="list-style-type: none"> • Project records • Data from public service agencies (health, education, etc.) • Government household and related surveys 	Consider when the data was collected, what population was included (or excluded), how reliable and relevant the results are in relation to the indicators and population that is being addressed by the present evaluation.
Assessing the reliability and validity of secondary data	<ul style="list-style-type: none"> • School enrollment and attendance records • Patient records in local health centers • Savings and loans cooperatives' records of loans and repayment • Vehicle registrations (to estimate changes in the volume of traffic) • Records of local farmers markets (prices and volume of sales) 	<p>All data must be assessed to determine their adequacy in terms of</p> <ul style="list-style-type: none"> • Reference period • Population coverage • Inclusion of required indicators • Documentation on methodologies used • Completeness • Accuracy • Freedom from bias

<p>Using recall: asking people to provide numerical (income, crop production, how many hours a day they spent traveling, school fees) or qualitative (the level of violence in the community, the level of consultation of local government officials with the community) at the time the project was beginning</p>	<ul style="list-style-type: none"> • Key informants • PRA (participatory rural appraisal) and other participatory methods 	<p>Recall can be used for</p> <ul style="list-style-type: none"> • School attendance • Sickness/use of health facilities • Income/earnings • Community/individual knowledge and skills • Social cohesion and conflict • Water usage and cost • Major or routine household expenditures • Periods of stress • Travel patterns and transport of produce
<p>Improving the reliability/validity of recall</p>	<ul style="list-style-type: none"> • Refer to previous research or, where possible, conduct small pretest-posttest studies to compare recall with original information • Identify and try to control for potential bias • Clarify the context • Link recall to important reference points in community or personal history • Triangulation (key informants, secondary sources, PRA) 	<ul style="list-style-type: none"> • Where possible refer to previous research that has determined accuracy of recall on certain types of indicators • Be aware of underestimation of small expenditures, truncating large expenditures by including some expenditures made before the recall period, distortion to conform to accepted behavior, intention to mislead. • Context includes time period, specific types of behavior, reasons for collecting the information

<p>Key informants</p>	<ul style="list-style-type: none"> • Community leaders • Religious leaders • Teachers • Doctors and nurses • Store owners • Police • Journalists 	<p>Use to triangulate (test for consistency) data from other sources</p>
<p>Collecting sensitive data (e.g., domestic violence, fertility behavior, household decision making and resource control, information from or about women, and information on the physically or mentally handicapped)</p>	<ul style="list-style-type: none"> • Participant observation • Focus groups • Unstructured interviews • Observation • PRA techniques • Case studies • Key informants 	<p>These issues also exist with project participants, but they tend to be more difficult to address with comparison groups because the researcher does not have the same contacts or access to the community.</p>
<p>Collecting data on difficult-to-reach groups (e.g., sex workers, drug or alcohol users, criminals, informal small businesses, squatters and illegal residents, ethnic or religious minorities, and in some cultures, women.)</p>	<ul style="list-style-type: none"> • Observation (participant and non-participant) • Informants from the groups • Self-reporting • Tracer studies and snowball samples • Key informants • Existing documents (secondary data) • Symbols of group identification (clothing, tattoos, graffiti) 	<p>As for previous point</p>

Similarly, the data analysis plan should be reviewed to determine what kinds of disaggregated data analysis are actually required. If it is found that certain kinds of proposed disaggregation are not needed (e.g. comparing the impacts of the project on participants in different locations), then it will often be possible to reduce the size of the sample.

Find simple ways to collect data on sensitive topics and from difficult-to-reach populations

Another challenge to evaluators, although not unique to RWE, regards the collection of data on sensitive topics such as domestic violence, contraceptive usage, or teenage violence; or from difficult to reach groups such as commercial sex workers, drug users, ethnic minorities, migrants, the homeless, or, in some cultures, women. A number of methods can help to address such topics and reach such groups. However, RWE constraints such as budget, time, or political prejudices could create pressures to ignore these sensitive topics or leave out groups of people who are difficult to reach. There are at least three strategies for addressing sensitive topics:

- Identify a wide range of informants who can provide different perspectives;
- Select a number of culturally appropriate strategies for studying sensitive topics;
- Systematically triangulate;

Difficult-to-reach groups include commercial sex workers, drug or alcohol users, criminals, informal and unregistered small businesses, squatters and illegal residents, ethnic or religious minorities, boy-friends or absent fathers, indentured laborers and slaves, informal water sellers, girls attending boys' schools, migrant workers, and persons with HIV/AIDS, particularly those who have not been tested.

The evaluator may face one of two scenarios. In the first scenario, the groups may be known to exist, but members are difficult to find and reach. In the second scenario, the clients and, at least initially, the evaluator may not even be aware of the existence of such marginalized or "invisible" groups. The techniques for identifying and studying difficult-to-reach groups are similar to those used for addressing sensitive topics and include the following:

- *Participant observation.* This is one of the most common ways to become familiar with and accepted into the milieu where the groups operate or are believed to operate. Often, initial contacts or introductions will be made through friends, family, clients, or in some cases, the official organizations with whom the groups interact.
- *Key informants.* Schedule interviews with persons who are particularly familiar with and well informed about the target groups.

- *Tracer studies.* Neighbors, relatives, friends, work colleagues, and so on are used to help locate people who have moved.
- *Snowball samples.* With this technique, efforts are made to locate a few members of the difficult-to-locate group by whatever means are available. These members are then asked to identify other members of the group so that if the approach is successful, the size of the sample will increase. This technique is often used in the study of sexually transmitted diseases.
- *Socio-metric techniques.* Respondents are asked to identify to whom they go for advice or help on particular topics (e.g., advice on family planning, traditional medicine, or for the purchase of illegal substances). A socio-metric map is then drawn with arrows linking informants to the opinion leaders, informants, or resource persons.

Be creative about data collectors

Creative options are sometimes available for reducing the cost of contracting data collectors. In a health evaluation, it may be possible to contract student nurses; in an agricultural evaluation, to contract agricultural extension workers; and, for many types of evaluation, to contract graduate students as interviewers or enumerators. Arrangements can often be made with the teaching hospital, the Ministry of Agriculture, or a university professor to contract students or staff at a rate of pay that is satisfactory to them but, well below the market rate. Although these options can be attractive in terms of potential cost savings, or for the opportunity to develop local evaluation capacity, there are obvious dangers from the perspective of quality. The interviewers may not take the assignment very seriously; it may be politically difficult to select only the most promising interviewers; or, to take action against people producing poor-quality work. Supervision and training costs may also be high, and the time required to complete data collection may increase. However, experience shows that these kinds of cooperation can work very well if there is a serious commitment on the part of the agency or university faculty.

Another creative option is to employ data collectors from the community. Sometimes a local high school can conduct a community needs assessment study, or a community organization can conduct baseline studies, or monitor project progress. A number of self-reporting techniques can also be used. For example, individuals or families can keep diaries of income and expenditures, daily time

use, or time, mode, and destination of travel. Community groups can be given cameras, tape recorders, or video cameras and asked to make recordings on issues such as problems facing young people, community needs, or the state of community infrastructure. Although all these techniques pose potential validity questions, they are valuable ways to understand the perspective of the community on the issues being studied.

Analyze data efficiently

Look for ways to manage data efficiently

Before data can be analyzed, they must be input into an electronic or manual format. If this is not done properly, the quality and reliability of the data can be compromised or time, money, or both can be wasted. Furthermore, if data are not properly managed, there is the risk that significant amounts of information will be lost. The following are some of the main steps in the development and implementation of an analysis plan:

- *Drafting an analysis plan.* This must specify for each proposed type of analysis, the objectives of the analysis, the hypothesis to be tested, the variables included in the analysis, and the types of analysis to be conducted.
- *Developing and testing the codebook.* If there are open-ended questions, the responses must be reviewed to define the categories that will be used. If any of the numerical data have been classified into categories ("More than once a week," "Once a week," etc.), the responses should be reviewed to identify any problems or inconsistencies.
- *Ensuring reliable coding.* This involves both ensuring that the codebook is comprehensive and logically consistent and also monitoring the data-coding process to ensure accuracy and consistency between coders.
- *Reviewing surveys for missing data and deciding how to treat missing data.* In some cases, it will be possible to return to the field or mail the questionnaires back to respondents, but in most cases, this will not be practical. Missing data are often not random, so the treatment of these cases is important to avoid bias. For example, there may be differences between sexes, age, and economic or education groups in their willingness to respond to certain questions. There may also be differences

between ethnic or religious groups or between landowners and squatters. One of the first steps in the analysis should be to prepare frequency distributions of missing data for key variables and, when necessary, to conduct an exploratory analysis to determine whether there are significant differences in missing data rates for the key population groups mentioned above.

With particular reference to entering the data into the computer or manual data analysis system:

- *Cleaning the data.* This involves the following:
 - Doing exploratory data analysis to identify missing data and to identify potential problems such as outliers. (These are survey variables where a few scores on a particular variable fall far above or below the normal range.) A few outliers can seriously affect the analysis by making it much more difficult to find statistically significant results (because the standard deviation is dramatically increased). Consequently, the data cleaning process must include clear rules on how to treat outliers.
 - Deciding how to treat missing data and the application of the policies
 - Identifying any variables that may require recoding
 - Providing full documentation of how data were cleaned, how missing data were treated and how any indices were created.

While RWE follows most of the standard data analysis procedures, a number of approaches may be required when time or budget are constraints. When *time* is the main constraint and where additional resources may be available to speed up the process, the following approaches can be considered:

- direct inputting of survey data into hand-held computers;
- use of electronic scanning to read questionnaires;
- sub-contracting data analysis to a university or commercial research organization;
- hiring more, or more experienced, data coders and analysts.

When *money* is the main constraint, one or more of the following options can be considered:

- limiting the kinds of statistical analysis to reduce expensive computer time;

- consider acquiring and using popular statistical packages such as SPSS or SAS so that the analysis can be conducted in-house rather than subcontracting. Needless to say this option requires the availability of statistical expertise in-house.

Focus analysis on answering key questions

It is wise advice for any evaluation to focus on the key questions that relate to the main purpose of undertaking an assessment. This is especially important for RWE, because choices need to be made on what can be dropped as a consequence of limitations of time and funding. By being reminded of what the major questions are and what is required to adequately answer them, those planning a RWE can be sure to focus on those issues and not others. Typically, the clients and stakeholders, as well as the evaluators themselves, would like to collect additional information. However, when faced with RWE constraints, what would be “interesting to find out” must be culled from “what is essential” to respond to those key questions that drive the evaluation.

The Real-World evaluator must understand which critical issues must be explored in depth and which are less critical and can be studied less intensively or eliminated completely. It is also essential to understand when rigorous (and expensive) statistical analysis is needed by the client (to legitimize the evaluation findings to members of congress or parliament, or to funding agencies critical of the programme), and when more general analysis and findings would be acceptable. The answer to these questions can have a major impact on the evaluation budget and time required, and particularly on the required sample design and size.

Assessing and addressing threats to the validity of the evaluation findings and conclusions

Validity refers to the extent to which evaluation findings and conclusions are supported by: the conceptual framework and programme theory model on which the evaluation was based; the statistical techniques (including sample design); how the project was designed and implemented; and, the similarities and differences between the project population and the wider population to which findings are generalized. If there are problems with the evaluation design or the way the data is interpreted, there is a danger that programmes not achieving their intended objectives may be continued

or even expanded, that good programmes may be discontinued or, that priority target groups may not have access to project benefits.

The Appendix to this chapter includes an abbreviated portion of a checklist that has been developed by the authors to assess validity³. The checklist⁴ identifies seven dimensions of validity and includes indicators for assessing the adequacy with which the evaluation addresses each threat to validity. These are:

- **Objectivity:** Are the conclusions drawn from the available evidence?
- **Reliability:** Is the process of the study consistent, reasonably stable over time and across researchers and methods?
- **Internal validity:** Are the findings credible to the people studied and to readers, and are the presumed causal linkages between project interventions and outcomes valid?
- **Statistical conclusion validity:** The statistical design and analysis may incorrectly assume that programme interventions have contributed to the observed outputs.
- **Construct validity:** The indicators of outputs, impacts and contextual variables may not adequately describe and measure the constructs (hypotheses, concepts) on which the programme theory is based.
- **External validity:** Do the conclusions fit other contexts and how widely can they be generalized?
- **Utilization:** Were findings useful to clients, researchers and communities studied?

The checklist can be used to assess validity at various points in the evaluation:

- (a) When the evaluation design is submitted by the evaluation consultants;
- (b) during the implementation of the evaluation;
- (c) when the draft final evaluation report is submitted;
- (d) After the evaluation has been completed (this is particularly useful for meta-evaluation).

3 The Appendix includes for illustrative purposes the following sections of the checklist: The cover page, the format for the summary assessment of each validity dimension (only two dimensions are included) and examples of the detailed checklists for two dimensions (Objectivity and External Validity)

4 The complete checklist is available at www.realworldevaluation.org.

Report findings efficiently and effectively

As we mentioned in the section above titled “Customizing Plans for Evaluation”, an evaluation should focus on the key questions which relate to the main reason for the evaluation. This is especially important for RWE, because choices need to be made on what can be dropped because of limitations of time and funding. Those key questions need to be kept in mind not only during the planning for the evaluation, data collection and analysis, but also when the report(s) are being written. There is a temptation to report on all sorts of “interesting findings,” but the evaluator(s) need to keep the report focused on answering the key questions which the client(s) and stakeholders want answered.

One of the most effective ways to increase the likelihood that evaluation findings are used is to ensure that they are of direct practical utility to the different stakeholders.

Some of the factors affecting utilization include:

- timing of the evaluation;
- recognizing that the evaluation is only one of several sources of information and influence on decision makers and ensuring that the evaluation complements these other sources;
- building an ongoing relationship with key stakeholders, listening carefully to their needs, understanding their perception of the political context, and keeping them informed of the progress of the evaluation. There should be “no surprises” when the evaluation report is presented. (Operations Evaluation Department 2005; Patton 1997).

Some steps in the presentation of evaluation findings include the following.

- Understand the evaluation stakeholders and how they like to receive information;
- Use visual presentation to complement written reports or oral presentations. Where appropriate and feasible, make use of presentation tools such as PowerPoint, but do not become a slave to the technology and do be prepared to work without this if the logistics become too complicated. Visual presentations are particularly useful when the presentation is not made in the first language of many people in the audience.

- Share the evaluation results through oral presentations. Many stakeholders are not comfortable with written reports or slide presentations, so talking about the findings can be important.
- Plan the written report to make it simple, attractive, and user-friendly. Consider presenting different versions of the findings in ways that are most understandable and useful to different audiences.
- Involve the mass media. When a goal is to reach and influence a wide audience (e.g. public opinion, all parents of secondary-school-age children, lawmakers), the press can be a valuable ally. However, working with the media requires time and preparation and if their involvement is important, it may be worth hiring a consultant who “knows the ropes.”

Succinct report to primary clients

The impact of many evaluations is reduced because the findings and recommendations do not reach the primary clients in time and in a form they like and understand. There is no one best way to report evaluation findings. It depends on the clients and the nature of the evaluation. A good starting point is to ask clients which previous reports they found most useful and why.

A general rule, particularly for RWE, where time tends to be a constraint, is to keep the presentation short and succinct. It is a good idea to have a physically short document that can be widely distributed; although the executive summary at the start of a large report may be well written, some clients and stakeholders may be intimidated by the size of the document and may not get round to opening the summary.

Vaughan and Buss (1998) present some useful guidelines for figuring out what to say to busy policy-makers and how to say it. They point out that many policy-makers have the intellectual capacity to read and understand complicated analysis, but most do not have the time. Consequently, many will want to be given a flavor of the complexities of the analysis (they do not wish to be talked down to), but without getting lost in details. Other policymakers may not have the technical background and will want a simpler presentation. So, there is a delicate balance between keeping the respect and interest of the more technical while not losing the less technical. However, everyone is short of time. Therefore the presentation must be short, even if not necessarily simple. Vaughan and Buss’s rules for figuring out what to say are as follows:

- Analyze policy but not politics. Evaluators are hired to provide technical expertise, not to advise on political strategies.
- Keep it simple.
- Communicate reasoning as well as bottom lines. Many policymakers will want to know how the evaluator arrived at the conclusions, so that they can assess how much weight to give to the findings.
- Use numbers sparingly.
- Elucidate, don't advocate. If evaluators advocate particular policies they risk losing the trust of the policymaker.
- Identify winners and losers. Decision makers are concerned with how policies affect their constituencies, particularly in the short run. Consequently, if evaluators and analysts want policymakers to listen to them, they must identify winners and losers.
- Don't overlook unintended consequences. People will often respond to new policies and programmes in unexpected ways, particularly to take advantage of new resources or opportunities. Sometimes unexpected reactions can destroy a potentially good programme, and in other cases unanticipated outcomes may add to the programme's success. Policy-makers are sensitive to the unexpected because they understand the potentially high political or economic costs. Consequently, if the evaluation can identify some important consequences of which policy-makers were not aware, this will catch the attention of the audience and raise the credibility of the evaluation.

Practical, understandable, and useful reports to other audiences

A dissemination strategy has to be defined to reach groups with different areas of interest, levels of expertise in reading evaluation reports, and preferences in terms of how they like to receive information. In some cases, different groups may also require the report in different languages. The evaluation team must decide which stakeholders are sufficiently important to merit the preparation of a different version of the report (perhaps even translation into a different language) or the organization of separate presentations and discussions.

These issues are particularly important for RWE because reaching the different audiences, particularly the poorest, least educated, and least accessible has significant cost and time implications. There is a danger that when there are budget or time constraints, the evalu-

ation will reach only the primary clients, and many of the groups whose lives are most affected may never see the evaluation, and may never be consulted on the conclusions and recommendations.

An important purpose of the scoping exercise is to agree with the client who will receive and have the opportunity to express opinions about the evaluation report. If the client shows little interest in wider dissemination, but is not actively opposed, then the evaluator can propose cost-effective strategies for reaching a wider audience. If, on the other hand, the client is actively opposed to wider consultation or dissemination, then the evaluator must consider the options – one of which would be to not accept the evaluation contract.

Assuming the main constraints to wider dissemination are time and budget, the following are some of the options:

- Enlist the support of NGOs and civil society organizations. They will often be willing to help disseminate but may wish to present the findings from their own perspective (which might be quite different from the evaluation team's findings), so it is important to get to know different organizations before inviting them to help with dissemination.
- Meetings can be arranged with organizations in the target communities to present the findings and obtain feedback. It is important that these meetings are organized sufficiently early in the report preparation process so that the opinions and additional information can be incorporated into the final report.
- If the client agrees that the findings of the evaluation would be of interest to a broader public, enlist the support of the mass media. It requires certain talents and the investment of a considerable amount of time to cultivate relationships with television, radio, and print journalists. They might be invited to join in field visits or community meetings and they can be sent interesting news stories from time to time. However, working with the mass media can present potential conflicts of interest for the evaluator, and many would argue that this is not an appropriate role for the evaluator.

Help clients use the findings well

Unfortunately, it is all too common for an evaluation to be completed, a formal report written and handed over to the client, and then nothing more done about it. Following the above advice, including involving the client and other key stakeholders throughout the evaluation process, one would hope that the findings of an evalua-

tion are relevant and taken seriously. However, if there is no follow-up, one can be left with the impression that the evaluation had no value. There are examples where major donor agencies, noting the limited use of evaluation reports, have decided to simply stop commissioning routine evaluations. Wouldn't it be better for more effort to be put into making sure evaluations are focused on answering key questions, well done, and then more fully utilized?

A major purpose of RWE is to help those involved focus on what is most important and to be as efficient as possible in conducting evaluations that add value and are useful. The final step, utilization, must be a part of that efficiency formula. If information is not used to inform decisions that lead to improved programme quality and effectiveness, it is wasted. The point here is that those conducting evaluations need to see that the follow-through is an important part of the evaluation process.

One way to do this is to help the client develop an action plan that outlines steps that will be taken in response to the recommendations of an evaluation and then to monitor implementation of that action plan. Doing this is obvious if this was a formative evaluation, where the findings are used to improve subsequent implementation of an ongoing project. Even in the case of a summative evaluation (where the purpose was to estimate the degree to which project outcomes and impacts had been achieved), or where the project that was evaluated has now ended, follow-up should include helping to utilize the lessons learned to inform future strategy and in the design of future projects. At a minimum, those responsible for an evaluation need to do whatever can be done to be sure that the findings and recommendations are documented and communicated in helpful ways to present and future decision makers.

Conclusion: who uses RWE, for what purposes and when?

There are two main users of RWE. These include *evaluation practitioners* who can use the RWE steps and approaches to:

- identify ways to cope with insufficient time and inadequate budgets for evaluations;
- overcome data constraints, particularly the lack of baseline data;
- and identify and address factors affecting the validity and adequacy of the findings of the evaluation.

The other main users are the *clients*, i.e. *representatives of agencies who commission evaluations and/or use evaluation findings*. Their concerns are similar though from different perspectives, including the need to:

- identify ways to reduce the costs of and time for evaluations, while still meeting the requirement for an adequately credible assessment that meets their needs and will be convincing to those to whom they must report; and
- understand the implications of different RWE strategies on the ability of the evaluation to respond to the purposes for which it was commissioned.

Application of the RWE approach can be helpful at three different points in the life of a project or programme: at the start during the planning stage (M&E plan and baseline), when the project is already being implemented (mid-term evaluation) or at the end (final evaluation). When the evaluation planning process begins at the start of the project, RWE can be used to help identify different options for reducing costs or time of the baseline, minimal but relevant monitoring data to be collected throughout the life of the project, plans for the subsequent evaluation(s), and for deciding how to make the best use of available data, or to understand client information needs and the political context within which the evaluation will be conducted.

When the evaluation does not begin until project implementation is already underway, RWE can be used to identify and assess the different evaluation design options that can be used within the budget and time constraints, and to consider ways to reconstruct baseline data. Attention will be given to assessing the strengths and weaknesses of administrative monitoring data available from the project and the availability and quality of secondary data from other sources. The feasibility of identifying a comparison group may also be considered.

When the evaluation does not begin until towards the end of the project (or after the project has already ended), RWE can be used in a similar way to the previous situation except that the design options are more limited as it is no longer possible to observe the project implementation process.

Under any of these scenarios, one of the innovative RWE approaches is to suggest measures that can be taken to strengthen the validity of the findings from the time of initial negotiations of the ToR, during the process of data collection and analysis, and even up to the point when the draft final evaluation report is being reviewed.

Appendix 1⁵:

CHECKLIST FOR ASSESSING THREATS TO THE VALIDITY OF AN IMPACT EVALUATION⁶

Part I. Cover Sheet

1. Name of project/programme
2. Who conducted this validity assessment? (indicate organizational affiliation)
3. When did the evaluation begin? A. Start of the project ____ B. Mid-term ____ C. Towards the end of the project ____ D. When the project has been operating for several years ____
4. At what stage of the evaluation was this assessment conducted? A. Proposed evaluation design ____ B. Progress report on the evaluation ____ C. Draft final evaluation report ____ D. After the evaluation has been completed ____
5. Reason for conducting the threats to validity assessment
6. Summary of findings of the assessment
7. Recommended follow-up actions (if any)

5 The complete checklist is available at www.realworldevaluation.org

6 Source: Michael Bamberger (2008) adapted from Miles and Huberman (1994) Chapter 10 Section 1; Guba and Lincoln (1989); Shadish, Cook and Campbell (2002) Tables 2.2, 2.4, 3.1 and 3.2; Bamberger, Rugh and Mabry (2006) Chapter 7 and Appendix 1 and Bamberger (2007). The present authors are solely responsible for the adaptation in this abbreviated form.

Part II. SUMMARY ASSESSMENT FOR EACH COMPONENT <i>[see attachments for more detailed assessments]</i>						
	Very strong				Serious problems	Not applicable
	1	2	3	4	5	N/A
Component A. Objectivity <i>(Confirmability): Are the conclusions drawn from the available evidence, and is the research relatively free of researcher bias?</i>						
Summary assessment and recommendations						
Overall rating of this component of the evaluation						
Number of issues/problems identified [indicate no. of 4 and 5 ratings]						
Component B. Reliability: <i>Is the process of the study consistent, coherent and reasonably stable over time and across researchers and methods? If emergent designs are used are the processes through which the design evolves clearly documented?</i>						
Summary assessment and recommendations						
Overall rating of this component of the evaluation						
Number of issues/problems identified [indicate no. of 4 and 5 ratings]						

**** Note: This and the following attachment are examples of the detailed checklists that are included for each of the seven components****

<i>Attachment. OBJECTIVITY</i> (<i>Confirmability</i>)	
Are the conclusions drawn from the available evidence, and is the research relatively free of researcher bias?	Rating
1. Are the conclusions and recommendations presented in the executive summary consistent with, and supported by, the information and findings in the main report.	
2. Are the study's methods and procedures adequately described? Are study data retained and available for re-analysis?	
3. Is data presented to support the conclusions? Is evidence presented to support all findings.	
4. Has the researcher been as explicit and self-aware as possible about personal assumptions, values and biases?	
5. Were the methods used to control for bias adequate?	
6. Were competing hypotheses or rival conclusions considered?	
General comments on this component	
Ratings: 1 = Evaluation design or analysis is very strong; 5 = design or analysis has serious problems	

<i>Attachment. EXTERNAL VALIDITY</i> <i>[Transferability]</i>	
	Rating
Reasons why inferences about how study results would hold over variations in persons, settings, treatments and outcomes may be incorrect.	
1. Sample does not cover the whole population of interest subjects may come from one sex or from certain ethnic or economic groups or they may have certain personality characteristics (e.g. depressed, self-confident). Consequently it may be difficult to generalize from the study findings to the whole population.	
2. Different settings affect programme outcomes. Treatments may be implemented in different settings which may affect outcomes. If pressure to reduce class size forces schools to construct extra temporary and inadequate classrooms the outcomes may be very different than having smaller classes in suitable classroom settings.	
3. Different outcome measures give different assessments of project effectiveness. Different outcome measures can produce different conclusions on project effectiveness. Micro-credit programmes for women may increase household income and expenditure on children's education but may not increase women's political empowerment.	
4. Programme outcomes vary in different settings. Programme success may be different in rural and urban settings or in different kinds of communities. So it may not be appropriate to generalize findings from one setting to different settings	
5. Programmes operate differently in different settings. programmes may operate in different ways and have different intermediate and final outcomes in different settings. The implementation of community-managed schools may operate very differently and have different outcomes when managed by religious organizations, government agencies and non-governmental organizations.	
6. The attitude of policy makers and politicians to the programme identical programmes will operate differently and have different outcomes in situations where they have the active support of policy makers or politicians than in situations where they face opposition or indifference. When the party in power or the agency head changes it is common to find that support for programmes can vanish or be increased.	

7. Seasonal and other cycles. many projects will operate differently in different seasons, at different stages of the business cycle or according to the terms of trade for key exports and imports. Attempts to generalize findings from pilot programmes must take these cycles into account.	
8. Are the characteristics of the sample of persons, settings, processes, etc. described in enough detail to permit comparisons with other samples?	
9. Does the sample design theoretically permit generalization to other populations?	
10. Does the researcher define the scope and boundaries of reasonable generalization from the study?	
11. Do the findings include enough “thick description” for readers to assess the potential transferability?	
12. Does a range of readers report the findings to be consistent with their own experience?	
13. Do the findings confirm or are they congruent with existing theory? Is the transferable theory made explicit?	
14. Are the processes and findings generic enough to be applicable in other settings?	
15. Have narrative sequences been preserved? Has a general cross-case theory using the sequences been developed?	
16. Does the report suggest settings where the findings could fruitfully be tested further?	
17. Have the findings been replicated in other studies to assess their robustness. If not, could replication efforts be mounted easily?	
General comments on this component	
Ratings: 1 = Evaluation design or analysis is very strong; 5 = design or analysis has serious problems	

References

- Bamberger, M. (2008) *Enhancing the utilization of evaluation for evidence-based policymaking*, in Segone, Bridging the Gap: the role of monitoring and evaluation in evidence-based policymaking. UNICEF.
- Bamberger, M. (2007) *A framework for assessing the quality, conclusion validity and utility of evaluations. Experience from international development and lessons for developed countries* Paper presented in the Panel Session *Simply the best? Understanding the market for "Good Practice" advice from government research and evaluations* American Evaluation Association Annual Conference. November 2007. Baltimore
- Bamberger, M. and White, H. (2007) *Using Strong Evaluation Designs in Developing Countries: Experience and Challenges* Journal of Multidisciplinary Evaluation. October 2007 Vol 4 No. 8 pp. 58-73 [http://survey.ate.wmich.edu/jmde/index.php/jmde_1/article/view/31/78]
- Bamberger, M. (editor). (2000) *Integrating Quantitative and Qualitative Research in Development Projects*. Directions in Development Series. World Bank.
- Brewer, J. & Hunter, A. (2006). (eds) *Foundations of Multimethod Research. Synthesizing Styles*. Sage Publications.
- Brown, J. (2000). *Evaluating the impact of water supply projects in Indonesia*. In Bamberger, M (editor). *Integrating Quantitative and Qualitative Research in Development Projects*. Directions in Development Series. World Bank.
- Creswell, J. C., Clark V.L., Guttman, M.L. & Hanson, W. (2003). *Advanced Mixed Method Research Designs*. In Tashakkori, A., & Teddye, C. (Eds) *Handbook of Mixed Methods in Social and Behavioral Science*. (pp. 209-240). Thousand Oaks, California: Sage Publications.
- Glewwe, P., Kremer, M., Sylvie Moulin & Zitzewitz, E. (2004). *Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya*. *Journal of Development Economics* 74: 251-268. <http://www.povertyactionlab.com/projects/project.php?pid=26>
- Guba, E. & Lincoln, Y. (1989) *Fourth Generation Evaluation*. Thousand Oaks, California. Sage Publications.
- Kozel, V. & Parker, B. (2000). *Integrated approaches to poverty assessment in India* in Bamberger, M.(Ed) (2000). *Integrating Quantitative and Qualitative Research in Development Projects. Directions in Development*. (pp. 59-68). Washington D.C: The World Bank.
- Kumar, S. (2002). *Methods for Community Participation. A Complete Guide for Practitioners*. London. ITDG Publishing.
- Miles, M. & Huberman, M. (1994) *Qualitative Data Analysis*. Thousand Oaks, California. Sage Publications.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, California: Sage Publication.

- Patton, M. Q. (1997). (Third Edition). *Utilization-focused evaluation*. Thousand Oaks, California: Sage Publications.
- Ravallion, M. (2006). *Evaluating anti-poverty programmes. Handbook for Agricultural Economics* (edited by Robert Evenson and Paul Schulz) Volume 4. North-Holland
- Rietberger-McCracken, J. & Narayan, D. (1997). *Participatory Rural Appraisal. Module III of Participatory tools and techniques: a resource kit for participation and social assessment*. Environment Department. Washington D.C: The World Bank.
- Roche, C. (1999). *Impact Assessment for Development Agencies. Learning to Value Change*. OXFAM.
- Rugh, J. (1986). *Self-Evaluation: Ideas for Participatory Evaluation of Rural Community Development Projects*. Oklahoma City, Oklahoma: World Neighbors.
- Schwarz, N. & Oyserman, D. (2001). *Asking Questions about Behavior: Cognition, Communication, and Questionnaire Construction*. *American Journal of Evaluation*. Volume 22. No. 2 pp. 127-160.
- Shadish, W. Cook, T. & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, Houghton Mifflin.
- Tashakkori, A & Teddye, C (eds) (2003). *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, California. Sage Publications.
- Valadez, J. & Bamberger, M. (1994). *Monitoring and evaluating social programmes in developing countries: a handbook for policymakers, managers and researchers*. Washington D.C. World Bank.
- Vaughan, R. & Buss, T. (1994). *Communicating Social Science Research to Policymakers*. Applied Social Science Research Methods Series No. 48. Sage Publications.
- White, H. (2006) *Impact Evaluation: The experience of the Independent Evaluation Group of the World Bank*. Washington D.C.: World Bank (www.worldbank.org/ieg/ecd)
- White, H. and Bamberger, M. (2008). *Impact Evaluation in Official Development Agencies*. IDS Bulletin Volume 39 No. 1 March 2008.
- World Bank. Independent Evaluation Group. (2006) *Conducting quality impact evaluations under budget, time and data constraints*. Available free at www.worldbank.org/ieg/ecd
- World Bank. Operations Evaluation Department. (2004). *Influential Evaluations*. World Bank. Operations Evaluation Department. (2005). *Influential Evaluations: Detailed Case Studies*. Available free at www.worldbank.org/ieg/ecd.

STRENGTHENING COUNTRY DATA COLLECTION SYSTEMS. THE ROLE OF THE MULTIPLE INDICATOR CLUSTER SURVEYS

Marco Segone, *Senior Regional Advisor,
Monitoring and Evaluation, UNICEF CEE/CIS*
George Sakvarelidze, *Monitoring and Evaluation Specialist,
UNICEF CEE/CIS*
Daniel Vadnais, *Data Dissemination Specialist,
UNICEF Headquarters*

The role of household surveys in country-led monitoring and evaluation systems

Results-based monitoring and evaluation systems are powerful public management tools to demonstrate accountability, transparency and results, as well as to support evidence-based policy making. Good monitoring and evaluation systems need ownership, efficient management, effective maintenance and credibility. The need to strengthen statistical capacity to support the design, monitoring and evaluation of national development plans has been recognized for at least the last three decades. This has been particularly true in the area of monitoring and evaluating of the situation of children and women.

In 1990, for instance, participants of the World Summit for Children recognized that many countries often lack the institutional capacity, or effective systems, for gathering reliable data in a timely manner. UNICEF answered the call and developed the Multiple Indicator Cluster Survey (MICS) programme, with surveys conducted every five years since 1995. Since the initiation of the programme, around 200 surveys have been implemented in approximately 100 countries.

The UNICEF-supported MICS is one of the few household survey programmes that governments can use for collecting standardized information on the socio-economic condition of households and household members, including women and children. Each round of surveys builds upon the last and offers new indicators to monitor current priorities in addition to the monitoring of trends. MICS also

offers a critical look at sub-national disparities faced by particular communities or groups, for instance, the Roma in FYR Macedonia or Serbia.

MICS, along with USAID-supported Demographic and Health Surveys (DHS), provides countries with the opportunities to strengthen their capacity in collecting data that is relevant to national and international development strategies and priorities. Through capacity building activities and a consultative process of adaptation and customization, MICS promotes national ownership of the household survey tool and of the collected data.

Overview of the third round of the Multiple Indicator Cluster Surveys (MICS3)

The third round of MICS (2005-2007) focused on providing a monitoring tool for the Millennium Development Goals (MDGs) and World Fit for Children Goals, as well as for other major international commitments, such as the United Nations General Assembly Special Session (UNGASS) on HIV/AIDS and the Abuja targets for malaria. Data on nearly half of the MDG indicators were collected in the third round of MICS, offering the largest single source of data for MDG monitoring.

The MICS3 questionnaire collected indicators on a wide range of topics including: child mortality; nutrition; child health; water and sanitation; reproductive health; child development; education; child protection; HIV/AIDS; sexual behaviour; and, children orphaned and made vulnerable by HIV/AIDS.

UNICEF works with a wide range of inter-agency MDG monitoring groups and other inter-agency indicator development groups with the aim of harmonizing, as far as possible, methodologies for measuring priority indicators.¹ UNICEF makes every effort to harmonize MICS – and the indicators measured – with other similar household survey projects, in particular the DHS programme. This level of coordination ensures maximum coverage, analysis of trends over time, and comparability across projects while guaranteeing the acquisition of most of the indicators needed to monitor the situation of children and women locally and globally.

1 These groups include: the Inter-agency Group for Child Mortality Estimation, the Malaria monitoring and evaluation reference group, the Technical advisory group of the WHO/UNICEF Joint monitoring programme on water supply and sanitation, the HIV/AIDS Monitoring and evaluation reference group, the Child health epidemiology reference group, the Global Alliance for Vaccines and Immunization Monitoring and evaluation task force and the Countdown to 2015 technical working group.

More than 50 countries carried out MICS3, including 12 countries in Central and Eastern Europe (CEE) and the Commonwealth of Independent States (CIS) which are at the heart of this paper. MICS3 is generating data representative of close to one in four children living in developing countries; nearly two in five children if India and China are excluded². During that round, some 500,000 households were surveyed and more than 300 experts from developing countries were trained in survey methodology.

Process leading to MICS3 data ownership and use

Strengthening national statistical capacity

Picture 1: First regional MICS3 workshop on Survey planning in Tbilisi, Georgia



The third round of MICS provided a broad avenue for strengthening the national statistical capacity of government institutions and individuals in over 50 countries. A key element of this strategy was UNICEF's implementation of a series

of four regional-level workshops. The purpose of these workshops was to train national officers in charge of implementing MICS3 in their country. Typically, these were government officials representing their national statistical office. For example, in the CEE/CIS region, a total of 12 countries decided to carry out MICS3 and their representatives were invited and trained in the course of the four workshops on household survey planning, data processing, data analysis and report writing and data archiving and dissemination.

The main guidance for MICS3 is available in the Multiple Indicator Cluster Survey Manual 2005, which covers all stages of survey planning and implementation. In addition to the manual, countries that carried out MICS3 were provided with standard software packages, data entry and tabulation programmes, and report templates. Most, but not all countries, followed the guidelines and standard procedures for the implementation of the surveys. UNICEF pro-

2 Source: The State of the World's Children 2008.

vided assistance throughout the survey process, either through the workshops, by distance communications or occasionally by going directly in a country. Throughout this process, all MICS3 participating countries were encouraged to submit to UNICEF key materials such as their national sampling plans, questionnaires, data sets and reports so as to allow the global MICS3 team to review their content and provide feedback.

In 2007-2008, UNICEF commissioned an evaluation of the MICS3 programme. This was carried out by the external consultancy firm John Snow Inc. One component of the evaluation was to assess the guidelines and standard procedures put forward to facilitate the implementation of MICS3. It was found that UNICEF's overall guidance was of high quality and in compliance with current international standards. A vast majority of countries adopted the standard software and data entry and tabulation programmes provided for data processing. This resulted in a significant improvement in standardization of MICS3 data sets. In general, countries that closely followed the MICS3 standards and guidelines and that submitted important materials for review were quite successful in producing data of good quality.

According to the online survey carried out within the framework of the MICS3 Evaluation, 97% of respondents working in implementing agencies felt that the MICS3 helped to build local capacity. The exposure of country level implementation teams to experts; the participation in the regional training workshops; the provision of user-friendly survey guidelines; and, the continuous interaction of the implementation teams with those responsible for the development of tools, have undoubtedly contributed to the development of capacity.

National ownership of MICS3 surveys

Picture 2: Official signature of the memorandum of understanding between the government of FYR Macedonia and UNICEF



MICS3 promoted the use (or establishment, where not existent) of inter-ministerial steering committees and the development of joint memorandums of understanding. Steering committees included not only government institutions but also international

organizations. They promoted joint review and selection of indicators and modules. This process was part of the assessment of data needs in the countries and allowed for the identification of indicators to fill in the deficit of information for monitoring national strategies, local MDGs and other government priorities.

Picture 3: Local interviewers interviewing the mother of a child in Kazakhstan



The emphasis on national ownership has been a major feature of the MICS programme. In the majority of MICS3 countries, national institutions led all stages of survey planning and implementation. The general approach in MICS3

was to empower national counterparts to undertake all survey activities, and to avoid performing any survey activity on behalf of the country implementers (typically the national statistics offices).

Even when a country required significant amounts of support to carry out a specific survey activity, this was implemented with strong involvement of the government counterparts. The aim was always to leave the completion of the activity to the counterparts. In only a few cases, and only after maximum effort, did UNICEF hire external survey experts to complete the survey, where completion would otherwise have been impossible.

One of the lessons learned from MICS3 is that when government ownership is weak and the national counterparts perceive the survey as a "UNICEF" activity, then the resulting commitment of the implementing agency has also been weak, causing delays in the completion of activities and sometimes sub-standard outputs. Another lesson is that a country's perception of the relevance of MICS has implications for national ownership of the survey and of its results.

Use of MICS3 data to inform evidence-based policy advocacy

Making data meaningful: the importance of data dissemination and communication

Picture 4: Two-page information sheet on MICS



The newly created dissemination team at UNICEF Headquarters (HQ) has been coordinating a comprehensive global dissemination and communication strategy for MICS data, in close collaboration with MICS3 colleagues in New York, regional and country counterparts. While dissemination materials and tools are country-designed and country-led, the UNICEF HQ team has liaised with MICS3 countries to encourage and support them in planning and delivering a number of

activities. It has also provided technical assistance to many individual countries. As new activities are implemented at the country and regional level, the HQ team has made efforts to track and collect these activities to make them publicly available at www.childinfo.org. These examples have become dissemination models for other countries and regions to use and adapt to their own needs.

To help raise visibility of the MICS tool and increase knowledge about the information it offers, a two-page information sheet on MICS was produced and made available at: www.childinfo.org.

Starting with the planning phase of MICS3, CEE/CIS made special efforts to ensure that MICS findings would be disseminated to the maximum extent possible. CEE/CIS was the first region to host the 4th Regional MICS3 Workshops on Data archiving and dissemination, and it actively contributed to making sure one full day would be dedicated to Data dissemination, and one to further analysis. As a result, the third round of MICS saw an increased dissemination of key findings, using new and innovative tools as well as the traditional ones. To access dissemination and further analysis materials based on MICS3 findings from the CEE/CIS region, please visit http://www.unicef.org/ceecis/resources_8588.html.

Several countries produced dissemination materials. Serbia and Kyrgyzstan opted for the production of shorter executive versions of MICS3 reports. These are simplified and more user-friendly summaries aiming at conveying the survey messages to the general audience in an efficient manner. Tajikistan designed a calendar highlighting MICS data on a monthly basis; Malawi produced a series of thematic wall charts; Vietnam designed various fact sheets and; Thailand, the first country to have completed MICS3, produced thematic sub-reports and provincial reports, leaflets, fact sheets, and a video.

Almost half of the CEE/CIS countries developed web-pages dedicated to MICS3. Printed materials for dissemination of the survey findings included fact sheets, booklets, leaflets, posters and calendars. Before launching the survey, most countries prepared and distributed media releases which were instrumental to the printing of articles and broadcasting of messages on radio and television.

Picture 5: Press releases were instrumental in producing articles in newspapers highlighting MICS findings



Picture 6: Calendar highlighting MICS findings in Tajikistan



Picture 7: Poster focusing on emerging challenges highlighted by MICS findings in Serbia



In order to make both the process and the content of MICS3 more understandable for the general audience and to promote national ownership of the survey, UNICEF CEE/CIS and HQ supported the development of a comprehensive video on the implementation of MICS3 in Uzbekistan. In addition, Serbia produced 26 episodes of a serial television documentary, called “Serbia fit for children,” based on their MICS findings.

Picture 8: Fact sheet on child nutritional status produced in Tajikistan



Picture 9: Serbia prepared 26 episodes of the TV serial “Serbia fit for children”



To facilitate easy access to MICS3 findings, about 25 countries, including Kyrgyzstan and Tajikistan, created a national version of MICSInfo based on DevInfo - a powerful database system designed to compile and disseminate data. Other countries, including FYR Macedonia and Serbia, included MICS3 data into their existing DevInfo national databases. DevInfo adaptations aim at easier access and dissemination of data on women and children, providing utility for producing charts, tables and maps.

Picture 10: CEE/CIS MICSInfo provides access to key MICS3 findings in 12 countries



The UNICEF CEE/CIS Regional Office produced MICS Info – available at www.micsinfo.org. It includes MICS3 data from 12 countries disaggregated by: family size; children living arrangement; sex; residence (urban/rural), mother’s/caretaker’s; wealth index; ethnicity/language/religion.

UNICEF’s decision to design a standardized MICS3 final report cover template proved to be very useful by ensuring consistency and a common image among all MICS3 participating countries.

Picture 11: Examples of country adaptations of the MICS final report cover.



Picture 12: New Childinfo website home page



Recently, the UNICEF dissemination team has also made a strong effort to improve the look of the www.childinfo.org home page which incorporates a number of original features which make it easier for users to find the statistical information they need on children and women.

The website highlights the leading role UNICEF plays in monitoring the situation of children and women worldwide, particularly in terms of: supporting data collection; maintaining and updating global databases; undertaking data analysis and methodological work; promoting data use; and dissemination, as well as being a leader among UN agencies responsible for the global monitoring of the child-related MDGs. The website also provides the technical resources for conducting MICS.

Access to data facilitates further analysis

MICS3 findings have been instrumental in informing strategic documents produced at global, regional and country level. Further analysis of MICS3 findings has been promoted from the very beginning of the process. One of the major pre-requisites for this was promotion of, and subsequent public access to, the micro datasets through implementing agencies and UNICEF HQ (visit www.childinfo.org). The International Household Survey Network (IHSN) Microdata Management Toolkit was used to document and archive the data sets and other survey information.

At the global level, an increasing number of analyses (such as a Health Equity study), incorporating MICS3 data, are being carried out. MICS3 data are also the basis for policy analyses in the *Global study on child poverty and disparities*, which is in progress across 40 countries³. Country reports, with disaggregated data, are at the heart of the study which will use newly-generated evidence on child poverty from MICS, DHS and other sources, as tools for starting and influencing public policy debates. Study findings will be used to improve access, use, equity and efficacy of social services and benefits, and to strengthen related programmes and partnerships.

3 See the Global Study Guide online at www.unicefglobalstudy.blogspot.com.

MICS3 data are also being used at the global level by interagency monitoring groups. These groups use MICS findings to develop joint estimates on a number of development indicators, in particular on: child labour; malaria coverage and burden; water and sanitation, immunization; AIDS; and, under-five and infant mortality. A good example is the release of CMEInfo, a DevInfo application presenting child mortality estimates using MICS, DHS and other representative data sources. It is available at: <http://www.childmortality.org/>

MICS3 data have informed a number of key publications, including: *Progress for children: a Report card on maternal mortality*; *Progress on drinking water and sanitation*; *Children and AIDS: Second stock-taking report*; *Countdown to 2015: Tracking progress in maternal, newborn & child survival*⁴; *The State of the world's children: Child survival; malaria and children: progress in intervention coverage*; *Progress for children: a World Fit for Children statistical review*.

Picture 13: Key MICS3 findings from 12 countries are presented in the publication “*Emerging challenges for children in Eastern Europe and Central Asia. Focus on disparities*”

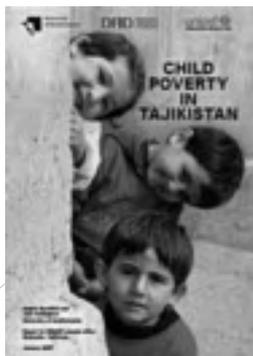


At the regional level, UNICEF CEE/CIS Regional Office used MICS3 data to produce the publication “*Emerging challenges for children in Eastern Europe and Central Asia: Focus on disparities*”. The publication consolidates key findings, focusing on disparities, of 12 MICS3 surveys carried out in CEE/CIS. It comes at a time when there is increasing evidence from a number of sources of growing and disturbing trends towards inequality within countries in the region. The publication presents cross-country tables with data disaggregated by social stratifiers and aims to promote deeper analysis and policy work at country level.

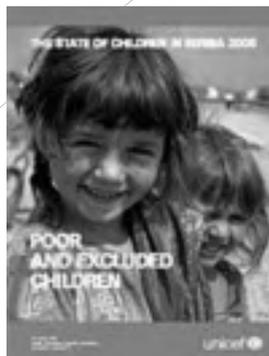
Key regional publications on early childhood development, education and nutrition were also informed by MICS3 data.

4 2015 is the date by which the international community will assess its committed achievement to the MDGs that aim at reducing under-five child deaths by two-thirds, from a baseline set in 1990.

Picture 14: MICS3 data informed the Child poverty study in Tajikistan



Picture 15: MICS3 data informed the study "*The situation of women and children in Serbia. Poor and excluded children*"



Several MICS3 countries, including, in the CEE/CIS region: Albania; Bosnia and Herzegovina; FYR Macedonia; Kyrgyzstan; Serbia; Tajikistan; and, Uzbekistan, used MICS3 data to inform monitoring processes. They use the situation analysis reports related to women and children (including minority groups); child poverty studies; sectoral analysis of early childhood development and child protection; comparative analysis of MICS 2 and MICS3; and monitoring reports for Poverty Reduction Strategies and MDGs.

Use of MICS3 data has enhanced evidence-based policy advocacy and decision making

MICS3 findings provided participating countries with information disaggregated by several background characteristics such as: region; urban/rural residence; gender; age; level of education; wealth index; ethnicity/language/religion, etc. For many indicators valid data has been obtained on the sub-national level. Disaggregated data allowed for the assessment of disparities within the countries. This is an important aspect for country-led monitoring and evaluation systems. This data also facilitated evidence-based policy advocacy and decision making.

Picture 16: Ms. Ann Veneman, UNICEF's executive director, visiting the MICS stand at the OECD World Forum on measuring and fostering the progress of societies.



When Ms. Ann Veneman, UNICEF's Executive Director, officially revealed (based on new data from MICS, DHS and other reliable sources), that the level of annual deaths of children under the age of five fell, for the first time, below the 10 million mark, news of this child survival

milestone spread all over the world on the Internet, as well as in newspapers, radio and television.

Picture 17: MICS3 findings informed the public hearing at the National Parliament on "Child health. Challenges and Solutions" in Serbia.



At country level, MICS3 findings were presented to Government policy makers and major stakeholders, including to Parliament in Kazakhstan. MICS3 findings have been presented in strategic national Conference, such as at the EU Conference on Social Inclusion

in FYR Macedonia and the National Conference on Poverty in Tajikistan. In Serbia, the MICS3 findings informed the public hearing at the National Parliament on "Child health. Challenges and solutions."

Although still at an early stage, some preliminary results achieved through the use of MICS3 findings in policy making are already being reported. In Serbia, for example, MICS3 findings were instrumental in initiating the establishment of the National commission on young children's' nutrition and feeding practices, as well as the initiative to ban corporal punishment, coordinated by the Serbian NGO network in partnership with the Ministry of Labor and Social Policy.

MICS3 was the first round in which there has been a strong emphasis on dissemination. With materials and activities now available online for countries to use as dissemination models, an increasing number of tools will be developed. This should also ensure that MICS4 data will benefit from an even more elaborate and sophisticated dissemination strategy with the goal of increasing the utilization of the data.

REFERENCES

John Snow, Inc. (2008). *Evaluation of UNICEF Multiple Indicator Cluster Surveys Round 3 (MICS3)*. Final Report. (forthcoming)

Segone, M. and Sakvarelidze, G. (2008). *Using MICS3 for Evidence-based policy making. The case of CEE/CIS* (PowerPoint presentation).

Available at: http://www.ceecis.org/mics/MICS3_CEECIS_DissUse.pdf

UNICEF CEE/CIS (2008). *MICS 3 Dissemination: Using MICS3 for Evidence-based policy advocacy. The case of CEE/CIS countries*. Geneva.

Available at: http://www.unicef.org/ceecis/resources_8588.html

UNICEF (2008). *Monitoring the Situation of Children and Women*. New York.

Available at: <http://www.childinfo.org/>

UNICEF (2006). *Multiple Indicator Cluster Survey Manual 2005: Monitoring the Situation of Children and Women*, New York.

Vadnais, Daniel. and Hancioglu, Attila. (2008). *The strategic intent of data collection and analysis. The case of Multiple Indicator Cluster Surveys (MICS)*. In: Segone, M., et al. Bridging the gap. The role of monitoring and evaluation in evidence-based policy making. Switzerland.

STRENGTHENING COUNTRY DATA DISSEMINATION SYSTEMS. GOOD PRACTICES IN USING DEVINFO

Nicolas Pron, DevInfo Global Administrator, UNICEF Headquarters

Kris Oswald, Executive Director, DevInfo Support Group

Marco Segone, Senior Regional Advisor, Monitoring and Evaluation, UNICEF CEE/CIS

George Sakvarelidze, Monitoring and Evaluation Specialist, UNICEF CEE/CIS

Country-led monitoring and evaluation systems are vital to national and decentralized development

Since their adoption by all United Nations Member States in 2000, the Millennium Declaration and the Millennium Development Goals have become a universal framework for development. They are also a means for developing and transition countries, and their development partners, to work together in pursuit of a shared future for all. In 2007, halfway to the MDGs' 2015 target date, there have been gains, but much remains to be done if millions of people are to realize the basic promises of the Millennium Declaration. To achieve sustainable outcomes, country-led development strategies must be backed by adequate financing within the global partnership for development. However, this is only possible if timely evidence is available from policy-relevant and technically-reliable country-led monitoring and evaluation systems. The evidence provided by such systems, owned by developing and transition countries, should inform necessary policies and strategies to ensure progress.

DevInfo is being used to support country-led monitoring and evaluation systems

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 Millennium Development Goals (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 indicators (the 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, as well as the contents of the databases supported by the system, include country-specific branding and packaging options. These options have been designed for broad ownership by national authorities.

The vision that DevInfo supports is a day when Member States use common database standards for tracking national human development indicators, containing high-quality data with adequate coverage and depth, to sustain good governance around the agenda of achieving the MDGs and national development goals.

DevInfo is being used as an advocacy platform to engage a broad spectrum of stakeholders in policy choices for human development. 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 in progress.

DevInfo is being used as a tool to restructure programming processes based on human rights. The system helps planners address disparities and target the most vulnerable sections of society. An important aspect of the DevInfo database structure is that it provides for monitoring multiple levels of sub-national data. The database structure also provides methods for monitoring subgroups (by sex, location (urban/rural), age-groups, ethnicity, education level, wealth index), and other important factors related to groups at risk and in need.

DevInfo can help design cost effective interventions based on facts, not perceptions. The system helps planners evaluate their options to plan for optimum results with limited resources. DevInfo presents the facts from multiple data sources with extensive metadata. This

assists planners to assess all of the available data related to the current situation, weigh alternatives and plan ahead as effectively as possible.

DevInfo. A database system designed to facilitate ownership by national authorities

National ownership and demand-driven monitoring and evaluation systems

Progress in human development is being made even in countries where the challenges are the greatest. This progress testifies to the unprecedented degree of commitment by these countries to achieve results through national ownership of the development process. National ownership of data dissemination processes helps to ensure that all stakeholders can make informed decisions about the future course of development policies that affect them as individuals, communities and the nation as a whole.

A survey conducted by UNICEF CEE/CIS Regional Office in 2008 showed that 68% of countries in the region are in various stages of DevInfo implementation. In most of these countries, the National Statistics Office (NSO) is the owner of the database, while in 32% of them the ownership is shared with other agencies or ministries. For example, in Kosovo, the Ministry of Science and Technology is supporting the DevInfo initiative. In Tajikistan, the Ministry of Economic Development and Trade is a national partner, along with the NSO.

The selection of indicators contained in a DevInfo database is demand-driven. This ensures that a national database will sustain its relevance and importance as a useful tool for monitoring national frameworks. The data's relevance, for tracking these frameworks, is critical to the success of the implementation of the database system. Successful DevInfo implementations have identified stakeholders and ensured their participation in governance of the system. The stakeholders have thoroughly examined the legal framework for gathering and use of statistics in the country, and its ramifications for DevInfo. They have leveraged relevant institutional structures and processes of government and partners to strengthen national data dissemination. Considering these issues helps position DevInfo strategically, creating links to relevant activities, such as in the areas of national strategic planning and support to the statistical system in the country. In this way DevInfo is conceived as a component of a more strategic approach to achieve national development goals.

DevInfo is being used by Member States to monitor comprehensive plans for sustainable development, including poverty reduction strategies, health and nutrition plans, environmental plans and education plans. DevInfo is being implemented by complementing existing databases and bridging data dissemination gaps.

Most of the countries in the CEE/CIS region that are implementing DevInfo have not limited the content of the national databases to the monitoring of the MDGs. Albania, Armenia, Bosnia and Herzegovina, Moldova and Serbia expanded its scope to monitor national development strategies, including poverty reduction strategies (PRSPs). Albania and Turkey are using DevInfo to monitor EU-related strategies, including social exclusion. In some cases DevInfo is being used for monitoring sectoral strategies, such as health care reform in Kyrgyzstan and the education strategy in Kosovo.

Picture 1: *ArmeniaInfo*, national adaptation in Armenia, is used to monitor MDGs as well as national development strategies



There are more than 16 national adaptations of DevInfo database technology in the CEE/CIS region. Some of these adaptations have been deployed online: for example, Tajikistan launched *TajikInfo* at www.tajikinfo.tj and Moldova launched *MoldovaInfo* at www.devinfo.md. Four national databases (Armenia, Azerbaijan, Macedonia and Serbia) are hosted at the global DevInfo website www.devinfo.info. In addition, the websites of the national statistical offices of Serbia (<http://webzrs.statserb.sr.gov.yu/axd/devinfo/indexe.htm>) and Montenegro (www.monstat.cg.yu/EngProjekti.htm) allow users to download their databases to function with the desktop version of DevInfo.

Picture 2: *TojikInfo*, local adaptation in Tajikistan, is available on line.



Picture 3: *Kyrgyzstan HealthInfo*, local adaptation in Kyrgyzstan, is used to monitor health reform.



National ownership processes entail several elements. It starts from the signature of a Memorandum of Understanding among stakeholders, to build a common database to monitor national development priorities. It then moves on to: outline roles and responsibilities of all stakeholders; commit financial and human resources; establish a steering committee to govern the content of the database; assign working groups to update the database; decide on the location of the common database; and finally, to end up with the integration of DevInfo database technology into the internal infrastructure of the government. This results in full institutionalization of the system.

An example of full ownership of the DevInfo system by a government is in the case of the Republic of Serbia. The government declared DevInfo as a database tool of particular interest for the Republic of Serbia in 2006. The technology thereby became part of the regular programme of the Statistical Office of the Republic of Serbia (SORS). This led to the formation of a committee on social indicators and analysis. The unit consists of four people, supported by the government, who have undertaken the task of further development and maintenance of the DevInfo database at the national

level. As a result, the national DevInfo database contains a rich set of 395 indicators at national level, which are classified in 12 sectors with 5 multilateral strategies: Millennium Development Goals (MDGs); Poverty Reduction Strategy (PRS); National Plan of Action for Children (NPA); World Fit for Children; and, World Summit for Children. The database also contains data on 91 indicators at local level (for each of 167 municipalities). A specially designed census database has 62 indicators at the settlement level (for each of 4,715 settlements). These databases are strong tools for monitoring and planning at central and local level.

Important initiatives are also taking place in other regions. For example, the Costa Rica government selected a strategic implementing partner, made them responsible for the system, so they took ownership and so, are developing it further, promoting it, and most importantly, sharing the information it contains.

In Egypt, a Memorandum of Understanding was signed among government agencies in charge of data collection, processing, analysis and dissemination. A major advantage is the linkage of DevInfo adaptations to existing decision-making mechanisms and processes in the country. For that purpose, it is helpful for a government body, directly linked to the decision-making process, to manage the system.

Tanzania's *TSED*, for example, is owned by the National Bureau of Statistics in collaboration with more than 20 ministries, departments and agencies in the country. It is embedded in the monitoring system for the National Strategy for Growth and Reduction of Poverty. In order to ensure the relevance of Tanzania's *TSED*, the database includes data for: the MDGs; the country's National Strategy for Growth and Reduction of Poverty; and, other relevant frameworks, such as Ageing and Aged Population; Labor Market Indicators; Maternal and Child Monitoring Indicators; and, Education for All. In addition, the National Bureau of Statistics implements a process for ensuring the quality, accuracy and reliability of the data. These conditions encourage the use of the database to produce reports to monitor the National Strategy for Growth and Reduction of Poverty, and it enables the government and its partners to gauge the progress being made by various interventions. Civil society organizations are using *TSED* in advocacy work related to policy formulation and budgetary processes. Others have also used the database for reporting, proposal writing and presentations.

Cambodia provides a clear illustration of strategic linkages. The Statistical Literacy Project has partnered with the *CAMInfo* initiative to conduct joint nation-wide trainings on *CAMInfo* and statistical literacy, targeting government officials and users of statistical data, including high-level decision makers. This partnership is expected to promote better coordination between the data manager, the National Institute of Statistics, and the planning and decision-making agency, the Ministry of Planning. As a result, better access to quality data and improved statistical literacy are anticipated to contribute to the improvement of the government's capacity to integrate statistical information into policy making. In St. Lucia, *Helen Info* is designed to be used by the government for Evidence-Based Social Policy. The database has been established in partnership between Government, EU, UNDP and UNICEF. Most important has been government ownership and their commitment to maintain and use the database. Following this successful example, DevInfo is now being rolled out throughout the Eastern Caribbean.

National capacity development

Access to timely and reliable development data plays an important role in helping identify national development issues and, through national capacity development in data dissemination, leads to better information for policy development. Progress is being made in sharpening national monitoring and evaluation systems and this is enhancing the impact of development funding. These efforts are being stepped up to increase awareness of potential problems and to find solutions for extreme disparities and vulnerabilities. Since 2004, more than 20,000 professionals have been trained in the use of DevInfo database technology. These training sessions have focused on best practices in establishing a common database on human development and on how to put the data to use for decision making. The training has targeted a broad audience of planners, politicians, policy analysts, researchers, teachers, youth and statisticians. It has been organized at global, regional, national and local levels. The strategy has been to create teams of master trainers who can assist others to become both trainers and database administrators.

National capacity development is also provided through technical missions and activities to assist national partners and UN agencies in setting up and using DevInfo database technology. In 2007, there were 298 technical support activities carried out. This has resulted in more than 120 countries using DevInfo as the database platform to develop their own national socio-economic databases.

Capacity development activities in Central and Eastern Europe and the Commonwealth of Independent States (CEE/CIS) started with a series of DevInfo roll-out training carried out by the UNICEF CEE/CIS Regional Office. The scope of this training varied from orientation and use of the software to advanced database administration and development of local adaptations of the database technology to meet country-specific requirements. There was also a session devoted to Training of Trainers in the user and data administration modules of DevInfo.

Since 2006 regional training has been implemented in partnership with the United Nations Economic Commission for Europe (UNECE) and UNDP Bratislava Regional Center. The training introduced DevInfo v5.0, a new version with the capability of disseminating data online. The DevInfo regional training brought together national partners and UN staff members already working together on monitoring national development priorities. These regional capacity building activities have been supplemented by the UN Development Group Office (UNDGO, now UNDOCO) which facilitated training in priority countries and included the roll-out of the UN Development Assistance Framework (UNDAF). These training activities were organized through the countries' UN Resident Coordinators.

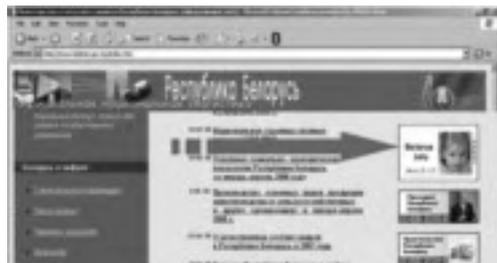
Promoted by these regional activities, much in-country training has been carried out. According to an e-mail survey carried out by the UNICEF CEE/CIS Regional Office in February 2008, more than 1000 people in CEE/CIS have been trained in DevInfo. This provides a critical mass of technical capacity to convey knowledge about the system and to carry out national and sub-national training.

In-country training is vital to the implementation of DevInfo database technology. This training, organized on behalf of national authorities, is integrated into a broad framework for monitoring national development priorities. Training focuses on the demand for data to monitor local circumstances.

An example of national capacity building is the step-by-step introduction of DevInfo in the Republic of Belarus. It started with a needs assessment in 2005, followed by participation in the DevInfo 5.0 regional roll-out training in Geneva (2006). The regional roll-out training was followed by a country request to carry out a session on DevInfo database administration in Belarus. This covered an overview for a wider international and national community and hands-on training for Ministry of Statistics and Analysis staff members. In 2006, database administration training was attended by 22

participants. This was facilitated in Russian by the UNICEF Regional Office, in collaboration with the UNDP and UNICEF country offices, and with the technical and logistical support of the Ministry of Statistics and Analysis. As a result of the training, the Ministry finalized a national adaptation of DevInfo for Belarus in 2007. The current version of *BelarusInfo* contains 126 indicators, focuses on national MDGs and provides access to socio-economic indicators related to human development in the country.

Picture 4: *BelarusInfo* is accessible at the website of the Ministry of Statistics and Analysis of the Republic of Belarus



Information on *BelarusInfo* can be obtained at www.belstat.gov.by. The database is currently available in Russian. The Ministry of Statistics and Analysis, in collaboration with UNDP and UNICEF, is plans to update, translate and further disseminate *BelarusInfo*, to insure wide access and usage of the database for informed decision making on national and the sub-national levels. Sub-national level training is also being planned.

Monitoring UN contribution to national development strategies and priorities

The United Nations Development Assistance Framework (UNDAF) is the strategic programme framework for the national development strategies supported by the UN Country Team. It describes UN contribution to the priorities in the national development framework. The outcomes of the framework show where the UN Country Team can bring its unique comparative advantages to bear in advocacy, capacity development, policy advice and programming for the achievement of related national priorities. A successful UNDAF is dependent on a strong, relevant national data dissemination system.

In India, the features of *DevInfo India* are being implemented to generate information on the overall situation with respect to sustainable development. The monitoring framework is inclusive of indicators to measure UNDAF outcomes/outputs, information on trends/mechanism for coordination, tracking of national development over time, progress of joint-sector programmes and responses to humanitarian emergencies. In Lesotho, *MalutiInfo* helps make information easily accessible to policy-makers, development practitioners and others, thus allowing them to monitor and evaluate the performance of identified indicators related to the UNDAF, PRS and MDGs. To increase the usefulness of the database, the country has created report templates to generate regular progress reports on thematic development agendas such as those related to the UNDAF; UN Common Country Assessment; National Human Development Reports; and, the Situational Analysis of Women and Children. Similarly, Malawi's *MASEDA* contains indicators for monitoring the country's development strategies, MDGs, and the UNDAF monitoring and evaluation (M&E) matrix, supplemented by indicators from other relevant areas such as governance. In Cambodia, *CAMIInfo* was adapted to include not only the indicators specific to monitoring the UNDAF, but additional indicators in the areas of governance and human rights, in order to capture more qualitative information and results at the output/outcome level.

Local monitoring and evaluation systems to strengthen decentralization

Successful national development strategies are built on sound economic and technical information which are used to design programmes to overcome key development challenges. These strategies are aimed to reduce child and maternal mortality, extreme poverty, lack of basic sanitation, unemployment and increasing inequalities. To be effective, national development strategies must be universal while targeting the most vulnerable and marginalized to reduce disparities. Policymakers must know where disparities exist within their own countries in order to develop relevant solutions which benefit the poor. The poor are often those living in rural areas or urban slums, children of mothers with no formal education, and living in the poorest households. National monitoring and evaluation systems focusing on disaggregated data, as well as decentralized systems, are fundamental to provide the information needed for policy makers to design and implement such developing strategies.

In Albania, UNDP (in partnership with UNICEF and UNFPA), supported local authorities, in all 12 regions of Albania, in developing Regional Development Plans. The decentralized monitoring and evaluation system is being supported by DevInfo. In Serbia, in compliance with the National Plan of Action for Children, 16 municipalities initiated Local Plans of Action for Children (LPA). These are strategic documents to define and guide optimal child development in local settings. The municipalities have been introduced to DevInfo to monitor progress, assess the local situation and inform decision making. Similarly, municipal databases are being developed in Montenegro. In Bosnia and Herzegovina, ten municipalities are working on the adaptation of DevInfo to strengthen child rights monitoring. In some municipalities, DevInfo is also used for monitoring the child protection systems reform. Data from municipalities is being sent to the Department of the Economic Development at central level where a consolidated dataset is used for national level planning and fund allocation. In the Russian Federation, the municipality of Moscow is exploring the opportunity of using DevInfo to monitor the Child Friendly Cities Initiative.

DevInfo is being used to monitor regional development challenges

DevInfo is being used at transnational level to highlight and monitor specific development challenges common to a group of countries or regions. For example, the UNICEF CEE/CIS Regional Office developed three adaptations: *MONEEInfo*, *MICS Info* and *Regional MGDInfo*.

MONEEInfo – available in online at www.moneeinfo.org – consists of 128 indicators related to the MDGs and beyond. *MONEEInfo*, based on the UNICEF IRC TranMonee database, allows monitoring of the situation of women and children in 27 countries of the region using time series from 1989 to the most recent year for which data are available. It is available in Russian and English. *MONEEInfo* provides a rich resource to access and analyze child protection indicators related to the institutionalization of children, living arrangements and juvenile justice, among other related issues.

Picture 5: *MONEE Info*, a regional adaptation developed by UNICEF CEE/CIS Regional Office, is based on TransMONEE data



MICSInfo (accessible at www.micsinfo.org), presents the findings for the third round of Multiple Indicator Cluster Surveys carried out in 12 countries of the CEE/CIS region. This DevInfo adaptation consists of a DevInfo gallery provides access to the charts with the key findings; the downloadable tables; the report “*Emerging challenges for children in Eastern Europe and Central Asia – Focus on disparities*”; and, provides full access to data on 59 indicators, including new indicators on child protection and early childhood development. Data are disaggregated by age, gender, family size, children living arrangement, residence, mother’s education, wealth index and ethnicity/language/religion.

Picture 6: *MICSInfo*, a regional adaptation developed by UNICEF CEE/CIS Regional Office, presents MICS3 data



The *Regional MDGInfo* database – accessible at www.regional-mdginfo.org – has been developed through partnership of UNICEF, UNDP and UNECE in an effort to strengthen national capacities in MDG literacy and monitoring. The database is used in advocacy for improvements in data quality and comparability. There are 78 indicators stratified by different background variables in the database. The gallery provides easy access to presentations of the key find-

ings related to progress towards the MDGs. *Regional MDGInfo* contains indicators from both national and international sources, as well as regionally-specific indicators, to maximize the relevance of MDG monitoring to the national context and to promote evidence-based advocacy for policy making.

Picture 7: Regional MDGInfo was developed by UNICEF, UNECE and UNDP



Data disseminated through DevInfo contributed to achieving results for children

Most of the countries in the CEE/CIS region that are using DevInfo report that the system is being used for preparing progress reports on MDGs and national development strategies. Serbia and Moldova reported that DevInfo was able to trigger important policy changes, including in public budgets, both at national and decentralized level.

According to Salah (2008), in Moldova, the DevInfo database of the Ministry of Economy and Trade provides central public authorities with relevant and internationally comparable statistical data on a regular basis. By using the same technology and the same lists of indicators in building two integrated national databases – Economic Growth and Poverty Reduction Strategy database (EGPRSP), and MDG database – the team avoided duplication in collecting statistics and increased the reliability of reporting. They also avoided the complexity which traditionally occurs in maintaining statistical data systems. With the objective of improving national capacity in decision-making, the Ministry of Economy and Trade developed two different types of comprehensive, analytical reports which are also DevInfo based. One, the Annual Evaluation Report on the Implementation of the Economic Growth and Poverty Reduction Strategy Paper, helped social sector ministries to discuss budgetary questions with the Ministry

of Finance. As a result, investments in social sectors were raised by 21 per cent in 2006. The other, the 2005 Poverty and Policy Impact Report, provided an overview of national development and included detailed analyses on child poverty and on poverty in rural areas.

These reports did not replace economic evaluations and public expenditures reviews. They did however provide useful information for decision-making since they contained analyses which indicated those elements which influenced programme results, and how the programme elements interacted among themselves. The reports were produced through an inclusive and nationally owned process where staff from MoET interacted with key decision-makers in line ministries. Because they provided objective analyses of local realities, they were also used by external donors. MoET organized an annual event which was a major opportunity for an evidence-based and participatory reflection on Moldova's performance in the economic and social sectors, and for a comparison with other countries. The reports were used for strategic planning including by teams developing the National Development Plan (NDP) 2008–2011. DevInfo played a role in facilitating a common understanding among the government, civil society organizations (CSOs) and development partners. Data analyses and maps were used as platforms for the national dialogue on poverty reduction. As information was easily accessible, DevInfo was used to produce a bulletin on EGPRSP implementation which was published in Moldovan newspapers and posted on government websites. This bulletin led to increased CSO participation and involvement in EGPRSP implementation. The materials developed by MoET for monitoring the *Poverty Reduction Strategy* helped a coalition of 14 non-government organizations (NGOs) develop the *State of the Nation Report* which presented civil society's view of development in Moldova. The main purpose of the Report was to play a role in decision-making and, in particular, to influence the content of the NDP for 2008–2011.

At the decentralized level, the municipality of Pirot in Serbia (Vasic, Petrovic and Jancovic, 2008) used DevInfo for reviewing the municipal budget allocation in favor of children. As a result, investment for children was increased seven-fold in just two years starting in 2005. In addition, an increasing demand from the local population for better quality of child social services prompted local authorities to provide additional funds. Firstly, additional funds were invested to equip the antenatal service. Secondly, there was increased funding of the Social Welfare Centre, schools and NGOs. Additionally, a new pre-school was built which tripled access to early childhood

education, raising it to 90% in the municipality. In the same municipality, DevInfo enabled local government to identify that none of the Roma children were attending pre-school facilities and that most of the children in the specialized institutions for children with disabilities were Roma. As a result, 50 children from Roma settlements were enrolled into pre-school (rather than in specialist institutions), and in one school year the proportion of Roma children in specialized institutions was reduced by 50%.

In Bosnia and Herzegovina, data disseminated through DevInfo are producing policy changes in education. Previously municipal authorities thought enrollment to primary school was 100 per cent. Now, thanks to data disseminated through DevInfo, local authorities realized that the situation is different for marginalized children. DevInfo also helped local municipalities to have a better insight in the area of social protection services, including for vulnerable and excluded groups, as well as on municipal budget allocation for children.

Conclusions

The DevInfo database initiative is proving that progress in human development can be accelerated through nationally-owned systems to strengthen data dissemination. The progress being made in use of data for decision-making bears witness to the unparalleled degree of advancement that can be achieved through ready access to relevant development data.

DevInfo is being used by the United Nations to strengthen its strategic national programme frameworks to deliver as One UN based on new approaches to create a common database on human development indicators supported by a strong data dissemination system.

National ownership of such data dissemination system is vital to the future course of human development where all stakeholders are able to be actively involved in evidence based policy decision making processes.

References

Djokovic-Papic, Dragana, Oliver Petrovic and Vladica Jankovic. (2008). *Using DevInfo to support Governments in monitoring National Development Strategies. The case of the Republic of Serbia*. In: Bridging the gap. The role of monitoring and evaluation in evidence-based policy making. UNICEF CEE/CIS Regional Office, Switzerland.

Pron, Nicolas. (2008). *The strategic intent of data dissemination. The case of DevInfo*. In: Bridging the gap. The role of monitoring and evaluation in evidence-based policy making. UNICEF CEE/CIS Regional Office, Switzerland.

Salah, Mohamed Azzedine. (2008). *Using DevInfo as a strategic tool for decision making. Achievements and lessons learned in Moldova*. In: Bridging the gap. The role of monitoring and evaluation in evidence-based policy making. UNICEF CEE/CIS Regional Office, Switzerland.

Segone, Marco. (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. UNICEF CEE/CIS Regional Office, Switzerland.

Vasic, Vladan, Oliver Petrovic and Vladica Jankovic. (2008). *Using DevInfo as a strategic tool to facilitate local communities' empowerment. The case of the Municipality of Piro*. In: Bridging the gap. The role of monitoring and evaluation in evidence-based policy making. UNICEF CEE/CIS Regional Office, Switzerland.

United Nations. (2007). *The Millennium Development Goals Report. 2007*.

MAKING DATA MEANINGFUL. WRITING STORIES ABOUT NUMBERS.¹

Colleen Blessing, *United States Department of Energy*
Vicki Crompton, *Statistics Canada*
Dag Ellingsen, *Statistics Norway*
Patricia Fearnley, *Office for National Statistics, United Kingdom*
John Flanders, *Statistics Canada*
John Kavalionas, *United States Census Bureau*
David Marder, *Office for National Statistics, United Kingdom*
Steve Matheson, *Australian Bureau of Statistics*
Kenneth Meyer, *United States Census Bureau*
Hege Pedersen, *Statistics Norway*
Sebastian van den Elshout, *Statistics Netherlands*
Don Weijers, *Statistics Netherlands*
Marianne Zawitz, *United States Bureau of Justice Statistics*

Making Data Meaningful. A guide to writing stories about numbers was prepared within the framework of the United Nations Economic Commission for Europe (UNECE) Work Session on Statistical Dissemination and Communication, under the programme of work of the Conference of European Statisticians.

The guide is intended as a practical tool to help managers, statisticians and media relations officers use text, tables, graphics and other information to bring statistics to life using effective writing techniques. It contains suggestions, guidelines and examples – but not golden rules. This publication recognizes that there are many practical and cultural differences among statistical offices, and that approaches vary from country to country.

What is a statistical story?

On their own, statistics are just numbers. They are everywhere in our life. Numbers appear in sports stories, reports on the economy, stock market updates, to name only a handful. To mean anything, their value to the person in the street must be brought to life.

¹ *Making Data Meaningful: A guide to writing stories about numbers* was originally published by the United Nations Economic Commission for Europe (UNECE). Reprinted with the permission of UNECE.

A statistical story is one that doesn't just recite data in words. It tells a story about the data. Readers tend to recall ideas more easily than they do data. A statistical story conveys a message that tells readers what happened, who did it, when and where it happened, and hopefully, why and how it happened. A statistical story can:

- provide general awareness/perspective/context; and
- inform debate on specific issues.

In journalistic terms, the number alone is not the story. A statistical story shows readers the significance, importance and relevance of the most current information. In other words, it answers the question: *Why should my audience want to read about this?*

Finally, a statistical story should contain material that is newsworthy. Ask yourself: Is the information sufficiently important and novel to attract coverage in the news media? The media may choose a different focus. But they have many other factors to consider when choosing a story line.

Statistical story-telling is about:

- catching the reader's attention with a headline or image;
- providing the story behind the numbers in an easily understood, interesting and entertaining fashion, and;
- encouraging journalists and others to consider how statistics might add impact to just about every story they have to tell.

Why tell a story?

A statistical agency should want to tell a story about its data for at least two reasons. First, the mandate of most agencies is to inform the general public about the population, society, economy and culture of the nation. This information will guide citizens in doing their jobs, raising their families, making purchases and in making many other decisions. Secondly, an agency should want to demonstrate the relevance of its data to government and the public. In such a way, it can anticipate greater public support for its programmes, as well as improved respondent relations and greater visibility of its products.

Most agencies rely mainly on two means of communicating information on the economic and social conditions of a country and its citizens: the Internet and the media. The Internet has become an

important tool for making access easier to the agency's information. More and more members of the public access an agency's data directly on its website. Still, most citizens get their statistical information from the media, and, in fact, the media remain the primary channel of communication between statistical offices and the general public. An effective way for a statistical office to communicate through both means is to tell a statistical story that is written as clearly, concisely and simply as possible. The goal for the Internet is to better inform the public through direct access. When writing for the media, the aim is to obtain positive, accurate and informative coverage. Statistics can tell people something about the world they live in. But not everyone is adept at understanding statistics by themselves. Consequently, statistical stories can, and must, provide a helping hand.

Last, but certainly not least, the availability of statistics in the first place depends on the willing cooperation of survey respondents. Statistical agencies cannot just rely on their legal authority to ensure a suitable response rate. The availability of statistics also depends on the extent to which survey respondents understand that data serve an important purpose by providing a mirror on the world in which we live. The more a statistical agency can show the relevance of its data, the more respondents will be encouraged to provide the data.

Considerations

Statistical agencies must take into account a number of key elements in publishing statistical stories.

First, the public must feel that it can rely on its national statistical office, and the information it publishes. Statistical stories and the data they contain must be informative and initiate discussion, but never themselves be open to discussion. In other words, the information must be accurate and the agency's integrity should never come into question. Statistical agencies should always be independent and unbiased in everything they publish. Stories must be based on high-quality data which are suitable to describe the issues they address. Changes in statistical values over time, for example, should be discussed only if they are determined by statisticians to be statistically significant.

Agencies should always guarantee the confidentiality of data on individual persons or businesses. Indeed, statistical stories may not iden-

tify, or in any way reveal, data on individuals or businesses. In their statistical storytelling, agencies must take into account the position and feelings of certain vulnerable groups in society. Information on these groups should be made available, but the goal should always be to inform the public. Agencies should never seek publicity for themselves at the expense of these particular target groups.

The authors of this guide suggest that statistical agencies should, for the benefit of the citizens they serve, formulate a policy that explains how their practices protect the privacy and confidentiality of personal information. This policy should be given a prominent position on the agency's website.

How to write a statistical story

Do you have a story?

First and foremost, you need a story to tell. You should think in terms of issues or themes, rather than a description of data. Specifically, you need to find meaning in the statistics. A technical report is not a story, nor is there a story in conducting a survey. A story tells the reader briefly what you found and why it is important to the reader. Focus on how the findings affect people. If readers are able to relate the information to important events in their life, your article becomes a lot more interesting.

Statistical offices have an obligation to make the data they collect useful to the public. Stories get people interested in statistical information and help them to understand what the information means in their lives. After they read good statistical stories, people should feel wiser and informed, not confused.

Possible topics/themes for stories:

- current interest (policy agenda, media coverage, etc.);
- reference to everyday life (food prices, health, etc.);
- reference to a particular group (teens, women, the elderly, etc.);
- personal experiences (transportation, education, etc.);
- holidays (Independence Day, etc.);
- current events (statistics on a topic frequently in the news);
- calendar themes (spring, summer, etc.);
- new findings;

- a regular series (“This is the way we live now”, “Spotlight on xxxx”, etc.).

Write like a journalist. The “inverted pyramid”

How can statisticians communicate like journalists? By writing their stories the way journalists do. The bonus is that the media are more likely to use the information.

Journalists use the “inverted pyramid” style. Simply, you write about your conclusions at the top of the news story, and follow with secondary points in order of decreasing importance throughout the text. Think of a typical analytical article as a right-side-up pyramid. In the opening section, you introduce the thesis you want to prove. In following sections, you introduce the dataset, you do your analysis and you wrap things up with a set of conclusions. Journalists invert this style. They want the main findings from those conclusions right up top in your news story. They don’t want to have to dig for the story.

You build on your story line throughout the rest of the text. If the text is long, use subheadings to strengthen the organization and break it into manageable, meaningful sections. Use a verb in subheadings, such as: “Gender gap narrows slightly.”

The lead. Your first paragraph

The first paragraph, or lead, is the most important element of the story. The lead not only has to grab the reader’s attention and draw him or her into the story, but it also has to capture the general message of the data. The lead is not an introduction to the story. On the contrary, it should tell a story about the data. It summarizes the story line concisely, clearly and simply. It should contain few numbers. In fact, try writing the first sentence of the lead using no figures at all.

Don’t try to summarize your whole report. Rather, provide the most important and interesting facts. And don’t pack it with assumptions, explanations of methodology or information on how you collected the data.

The lead paragraph should also place your findings in context, which makes them more interesting. Research shows that it is easier to remember a news report if it establishes relevance, or attempts to explain a particular finding. Exercise caution, though. It is not a good

idea to speculate, especially if your statistical office cannot empirically establish causality, or does not produce projections.

Give enough information so the reader can decide whether to continue reading. But keep it tight. Some authors suggest five lines or fewer – not five sentences – for the opening paragraph.

Poor: A new study probes the relationship between parental education and income and participation in post-secondary education from 1993 to 2001.

Good: Despite mounting financial challenges during the 1990s, young people from moderate and low-income families were no less likely to attend university in 2001 than they were in 1993, according to a new study.

Finally: there is no contradiction between getting attention and being accurate.

Remember:

- focus on one or two findings;
- write in everyday language (the “popular science” level);
- create images for your readers;
- focus on the things you want readers to remember;
- choose the points you think are newsworthy and timely.

Good writing techniques

Write clearly and simply, using language and a style that the layperson can understand. Pretend you are explaining your findings to a friend or relative who is unfamiliar with the subject or statistics in general. Your readers may not be expert users who often go straight to the data tables. Terms meaningful to an economist may be foreign to a layperson, so avoid jargon. Use everyday language as much as possible. If you have to use difficult terms or acronyms, you should explain them the first time they are used.

Remember: on the Internet, people want the story quickly. Write for the busy, time-sensitive reader. Avoid long, complex sentences. Keep them short and to the point. Paragraphs should contain no more than three sentences.

Paragraphs should start with a theme sentence that contains no numbers.

Example: Norway’s population had a higher growth last year than the year before. The increase amounted to 33,000 people, or a growth rate of 0.7%.

Large numbers are difficult to grasp. Use the words millions, billions or trillions. Instead of 3,657,218, write “about 3.7 million.” You can also make data simpler and more comprehensible by using rates, such as per capita or per square mile. Some suggestions follow.

<i>Use</i>	<i>Avoid</i>
<ul style="list-style-type: none"> • Language that people understand; • Short sentences, short paragraphs; • One main idea per paragraph; • Subheadings to guide the reader’s eye; • Simple language: “Get,” not “acquire.” “About,” not “approximately.” “Same,” not “identical”; • Bulleted lists for easy scanning; • A good editor. Go beyond Spell-Check; ask a colleague to read your article; • Active voice. “We found that...” Not: “It was found that....”; • Numbers in a consistent fashion: For example, choose 20 or twenty, and stick with your choice; • Rounded numbers (both long decimals and big numbers); • Embedded quotes (these are sentences that generally explain “how” or “why”, and which journalists like to use verbatim in their news stories in quotes); • URLs, or electronic links, to provide your reader with a full report containing further information. 	<ul style="list-style-type: none"> ✗ “Elevator statistics”: This went up, this went down, this went up; ✗ Jargon and technical terms; ✗ Acronyms; ✗ All capital letters and all italics: Mixed upper and lower case is easier to read; ✗ “Table reading”, that is, describing every cell of a complex table in your text.

Not Good: From January to August, the total square metres of utility floor space building starts rose by 20.5% from the January to August period last year.

Better: In the first eight months of 2004, the amount of utility floor space started was about 20% higher than in the same period of 2003.

Headlines. Make them compelling

If your agency's particular style calls for a headline on top of a statistical story, here are some suggestions to keep in mind.

Readers are most likely to read the headline before deciding to read the full story. Therefore, it should capture their attention. The headline should be short and make people want to read on. It should say something about the findings presented in the article, not just the theme.

Write the headline after you have written your story. Headlines are so important that most newspapers employ copy editors who craft the headlines for every story. Because the information is likely to be new to them, these editors can focus more readily on the most interesting aspects of the story.

In the same vein, statistical agencies might consider a similar arrangement. The individual who writes the headline could be different than the story's author.

Headlines should:

- be informative, appealing, magnetic, interesting and newsy, and incorporate:
 - the highest since, the lowest since...;
 - something new;
 - the first time, a record, a continuing trend;
- make you want to read the story, not scare you off;
- summarize the most important finding;
- be no longer than one line of type;
- not try to tell everything;
- contain few numbers, if any at all;
- have a verb or implied verb.

- Not Good: New report released today (the report is not the news)
Energy conservation measures widespread (too vague)
Prices up in domestic and import markets (what prices?)
- Good: Gasoline prices hit 10-year low
Crime down for third year in a row
July oil prices levelled off in August

Tips for writing for the Internet

The principles of good writing also apply to writing for the Internet, but keep in mind some additional suggestions.

People scan material on the Internet. They are usually in a hurry. Grabbing their attention and making the story easy to read are very important. You also have different space limitations on the Internet than on paper. Stories that make the reader scroll through too many pages are not effective. Avoid making the reader scroll horizontally.

Format the page so the story can be printed properly, without text being cut off by margin settings. A common solution is to include a link to a 'print friendly version', usually another page with navigation menus and banners removed.

Write your text so the reader can get your point without having to force themselves to concentrate. Use structural features such as bulleted lists, introductory summaries and clear titles that can stand alone.

Don't use ALL CAPITAL LETTERS on the Internet. It looks like you're shouting. Underline only words that are electronic links. Use boldface rather than underlining for emphasis. Avoid italic typefaces because they are much harder to read.

Make sure your story is printed on a contrasting background colour: either light lettering on a dark background or the reverse. High contrast improves readability on the Internet. Make sure items are clearly dated so readers can determine if the story is current.

Graphs

A picture is indeed worth a thousand words, or a thousand data points. Graphs (or charts) can be extremely effective in expressing key results, or illustrating a presentation.

An effective graph has a clear, visual message, with an analytical heading. If a graph tries to do too much, it becomes a puzzle that requires too much work to decipher. In the worst case, it becomes just plain misleading. Go the extra mile for your audience so that they can easily understand your point.

Good statistical graphics:

- show the big picture by presenting many data points;
- are “paragraphs” of data that convey one finding or a single concept;
- highlight the data by avoiding extra information and distractions, sometimes called “non-data ink” and “chart-junk”;
- present logical visual patterns.

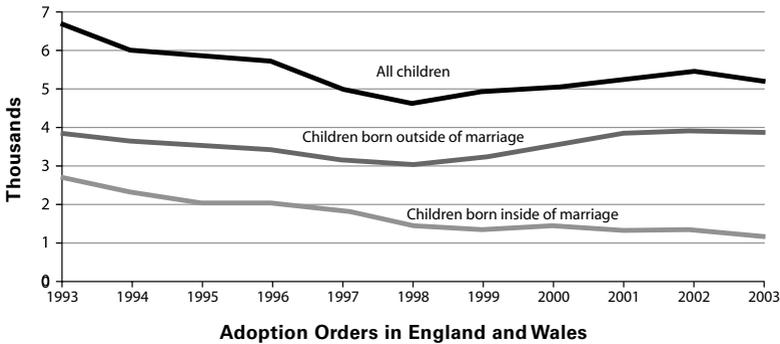
When creating graphics, let the data determine the type of graph. For example, use a line graph for data over time, or a bar graph for categorical data. To ensure you are not loading too many things into a graph, write a topic sentence for the graph.

Achieve clarity in your graphics by:

- using solids rather than patterns for line styles and fills;
- avoiding data point markers on line graphs;
- using data values on a graph only if they don’t interfere with the reader’s ability to see the big picture;
- starting the Y axis scale at zero;
- using only one unit of measurement per graphic;
- using two-dimensional designs for two-dimensional data;
- making all text on the graph easy to understand;
 - not using abbreviations;
 - avoiding acronyms;
 - writing labels from left to right;
 - using proper grammar;
 - avoiding legends except on maps.

For example:

Adoptions fall by 2.4% in 2003²



Tables

Good tables complement text. They should present numbers in a concise, well-organized fashion to support the analysis. Tables help minimize numbers in the statistical story. They also eliminate the need to discuss insignificant variables that are not essential to the story line.

Make it easy for readers to find and understand numbers in your table. Standard presentation tables are generally small. One decimal place will be adequate for most data. In specific cases, however, two or more decimal places may be required to illustrate subtle differences in a distribution.

Presentation tables rank data by order or other hierarchies to make the numbers easily digestible. They also show the figures that are highest and the lowest, as well as other outliers. Save large complex tables for supporting material. Always right-justify the numbers to emphasize their architecture. The guidelines listed for graphics above, such as highlighting data by avoiding “non-data ink”, also apply to the presentation of tables. While graphics should be accompanied by an analytical heading, titles are preferred for tables. They should be short and describe the table’s precise topic or message.

2 Graph from United Kingdom Office of National Statistics. Available online at <http://www.statistics.gov.uk/cci/nugget.asp?ID=592> [accessed 28 September 2005].

For example:

Race of Juvenile Offenders³

<i>Race of juvenile offender(s)</i>	<i>Average annual percent of violent crimes committed by juvenile(s)</i>
Total	100.0%
White	59.1
Black	25.2
Other	11.4
More than 1 racial group	2.6
Unknown	1.7

Maps

Maps can be used to illustrate differences or similarities across geographical areas. Local or regional patterns, which may be hidden within tables or charts, are often made clear by using a well designed map.

Maps are a rapidly expanding area of data presentation, with methods of geographic analysis and presentation becoming more accessible and easier to use. The cost of Geographic Information Systems (GIS), or software capable of mapping statistics, has decreased rapidly in the last ten years. Mapping that was once expensive, or required specialist hardware, is now within reach of most organizations. GIS analysis and presentation are now taught in schools and universities.

Producing statistical maps can be a simple process. The most common type of statistical map is the choropleth map, where different shades of a colour are used to show contrast between regions (usually a darker colour means a larger statistical value). This type of map is best used for ratio data (e.g. population density), where the denominator is usually area (e.g. square kilometers) or population. 'Count' data which has no denominator (e.g. total number of sheep

3 Table from Juvenile Victimization and Offending, 1993-2003, Bureau of Justice Statistics, Special Report, August 2005, NCJ 209468 (page 8). Available online at <http://www.ojp.usdoj.gov/bjs/pub/pdf/jvo03.pdf> [accessed 28 September]

in each region), are best illustrated using proportional or graduated symbol maps. With proportional symbol maps, the size of a symbol, such as a circle, increases in proportion to the value of the statistic. All mapping software should be capable of producing these two map types. Other types of map are possible but are best retained for specialist audiences.

When designing a map, always think about the audience and try to make it quick and easy for them to understand. If there is a natural association between a colour and a topic (e.g. blue for cold temperatures) then it would be sensible to use that colour for the legend. When choosing your legend classes, do not use complex methods unless your audience will understand them. Choosing classes of equal size, or classes containing similar numbers of events, are the most common methods. When choosing how many coloured classes to use, less is often more. Fewer classes emphasize similarity between areas and more classes emphasize the differences.

It should be possible for any statistical map to be read by a user without reference to other information and knowledge. Maps should always have a title and a legend that adequately explain the statistical units, the date that the statistical information was collected or produced and the geographic area type used. The source of statistical data should also be clearly stated. Footnotes may be used to clarify this information where needed and help to simplify titles.

Average Annual Rainfall 1961 – 1990, Europe⁴



4 Graph from United Nations Economic Commission for Europe. Available online at <http://www.unece.org/stats/trends2005/environment.htm> [accessed 30 September 2005].

How to encourage good writing

Each statistical agency may have its own ideas on ways to reward quality writing. But here are some general suggestions.

- set goals, such as a number of stories to be written each year.
- reward good writers for the best headline, most contributions, etc.
- make writing an expected part of the job rather than a sideline.
- explore techniques for building enthusiasm for writing.
- show staff the results of their writing: Post newspaper or magazine coverage initiated by their stories on an office bulletin board.
- provide training.

Writing about data. Make the numbers “stick”

Numbers don’t “talk”. But they should communicate a message, effectively and clearly. How well they do this depends a lot on how well authors use numbers in their text.

In a sense, journalists and statisticians are from two cultures. They tend not to talk the same language. Journalists communicate with words; statisticians communicate with numbers. Journalists are often uncomfortable when it comes to numbers. Many are unable even to calculate a percentage increase. So here are some suggestions for making the data “stick:”

Don’t peel the onion. Get to the point:

Poor: The largest contributor to the monthly increase in the CPI was a 0.5% rise in the transportation index.

Better: Higher auto insurance premiums and air fares helped push up consumer prices this month.

Avoid proportions in brackets:

Poor: Working seniors were also somewhat more likely than younger people to report unpaid family work in 2004 (12% versus 4%).

Better: About 12% of working seniors reported unpaid family work in 2004 compared with 4% for younger people.

Watch percentage changes vs. proportions: A percentage change and a percentage point change are two different things. When you subtract numbers expressed as proportions, the result is a percentage point difference, not a percentage change.

Wrong: The proportion of seniors who were in the labour force rose 5% from 15% in 2003 to 20% in 2004.

Right: The proportion of seniors who were in the labour force rose five percentage points from 15% in 2003 to 20% in 2004.

Avoid changing denominators:

Confusing: Two out of every five Canadians reported that they provided care for a senior in 2001, compared with one in seven in 1996, according to the census.

Clearer: About 40% of Canadians reported that they provided care for a senior in 2001, up from 14% in 1996, according to the census.

Reduce big numbers to understandable levels:

Cumbersome: Of the \$246.8 billion in retail spending last year consumers spent \$86.4 billion on cars and parts, and \$59.3 billion on food and beverages.

Easy to grasp: Of every \$100 spent in retail stores last year, consumers spent \$31 on cars and parts, compared with only \$23 on food and beverages.

What's wrong with this article?

A NEW REPORT RELEASED TODAY SAYS THAT THE PRICES OF MANY PETROLEUM PRODUCTS WILL BE HIGHER IN THE FUTURE

The tight global markets and elevated crude oil prices are expected to result in higher **prices for petroleum products**. The cost of imported crude oil to refineries this winter is projected to average 98.3 c/g (about \$40 per bbl) compared to 70.1 c/g last year. During the winter, WTI prices are expected to

decline from their current record levels but remain in the \$40 per bbl range, but despite above-average natural gas stocks, average winter natural gas prices, both at the wellhead and retail levels, are expected to be above those of last winter, particularly during the fourth quarter of 2004, in response to the hurricane-induced production losses in the Gulf of Mexico during September.

Increases in heating fuel prices are likely to generate higher expenditures even in

regions where demand for fuel is expected to fall. Average residential natural gas prices this winter are expected to be 10 percent higher year-over-year and household expenditures are expected to be 15 percent higher.

Therefore, residential space-heating expenditures are projected to increase for all fuel types compared to year-ago levels.

Demand is expected to be up by 1.637 percent. This increase reflects greater heating degree days in key regions with larger concentrations of gas-heated homes and continued demand increases in the commercial and electric power sectors. Due to the availability of primary inventories, many petroleum products are expected to be reasonably well

protected against the impact of demand surges under most circumstances. As of October 1, working natural gas inventories were estimated to be 3.6tcf, up 2 percent from three years ago, 3 percent from two years ago and 1 percent from last year.

Other interesting findings from this report are that the spot price for crude oil continues to fluctuate. Prices continue to remain high even though OPEC crude oil production reached its highest levels in September since OPEC quotas were established in 1982. Overall inventories are expected to be in the normal range, petroleum demand growth is projected to slow, and natural gas prices will increase.

- Headline is too long and doesn't make a clear point.
- All-cap headline looks like the author is shouting.
- Don't underline words unless they are an electronic link.
- Lead paragraph is background.
- Report title and release date aren't stated.
- Jargon: Readers might not know that gasoline and heating oil are petroleum products.
- Spell out units: c/g is cents per gallon; bbl is barrel.
- Acronyms: OPEC is the Organization of Petroleum Exporting Countries.
- First paragraph is too long: Too much detail, too many numbers.
- Sentences are too long.
- The main story line is in the third paragraph.
- Unexplained references: demand for what is expected to be up?

- Round numbers: not 1.637 percent.
- Elevator economics: this is up, this is down.
- Bullets preferable in the last paragraph.
- No URL link cited at the end.
- No contact or phone number provided.
- Proof read! In the last paragraph, “thought” should be “though”; “it’s” should be its” and “will be will increase” should read “to increase”.

A Revised Version

Released: September 16, 2004

Consumers will spend more to heat their homes this winter

Homeowners will pay much more this winter to heat their homes, according to the latest Heating Usage report released today by the Energy Minister. It predicts an 8% increase in spending over last winter.

Increases in prices for heating fuel are likely to generate higher spending, even in regions where demand for fuel is expected to fall. Average residential prices for natural gas are expected to be 10% higher than last winter, while household spending is expected to rise by 15%.

Tight global markets and elevated crude oil prices are expected to result in higher prices for petroleum products. The cost of imported crude oil to refineries this winter is projected to average 98 cents per gallon (about \$40 dollars per barrel), compared with 70 cents per gallon last year.

Despite above-average stocks of natural gas, average winter natural gas prices, both at the wellhead and retail levels, are expected to be above those of last winter.

Other interesting findings from this report:

- The spot price for crude oil continues to fluctuate. Prices continue to remain high even though the Organization of Petroleum Exporting Countries (OPEC) production of crude oil reached its highest levels in September since OPEC was established in 1982.
- Overall petroleum inventories are expected to be in the normal range.

See the entire report at www.HeatingUsage.gov. Contact John Smith in the Press Office at 123.4567 for more information.

Evaluating the impact

Media analysis

It is a good idea for statistical agencies to monitor the impact of their statistical stories in the print and electronic media from the point of view of both the number of “hits” and the quality of coverage. Useful resources for gauging the breadth, balance and effectiveness of media coverage include Google News, LexisNexis, blogs, and electronic and paper subscriptions.

Monitoring coverage can help managers determine if more work is needed to educate journalists, statisticians or key stakeholders about better ways of conveying the meaning of numbers in language that laypeople can understand. Monitoring would include:

- keyword searches to measure extent of media coverage;
- total coverage for a pre-determined period of time;
- daily coverage to identify spikes;
- comparing coverage to established baselines;
- prior releases of the same data product;
- qualitative methods to analyse media coverage;
- correct interpretation of the numbers;
- coverage of target audiences;
- inclusion of key story-line messages;
- inclusion of core corporate messages;
- effective use of illustrative embedded graphics;
- tone of story (positive/negative);
- tone of quotes from external spokespersons (positive/negative).

Website analysis

Monitoring Internet traffic with website usage software can help determine types of stories most in demand. You should look for:

- the number of page views, visits, etc., to specific pages;
- where visitors are coming from;
- where visitors are going when they leave your pages.

In addition, surveys of users of your site – both media and general users – can help target and improve the information available. You should:

- ask the customer if they found what they were looking for when they came to the site;
- target specific questions to known users of the site;
- ask how the site is used and how often;
- assess general satisfaction with the site;
- solicit recommendations for change or additional topics;
- use focus groups with media representatives to explore needs, approaches and reactions.

Before and after: Applying good writing techniques

To illustrate how to turn a routine statistical story into one with a much stronger story-line and more effective use of data, here is a ‘before’ and ‘after’ example. Note the differences.

BEFORE

Divorces – 2003

In 2003, 70,828 couples divorced, up a slight 1.0% from the recent low of 70,155 in 2002.

The number of divorces has remained relatively stable over the last few years. The year-to-year change has been below two percent for every year since 1999.

The increase in the number of divorces between 2002 and 2003 kept pace with the increase in the Canadian population over this period. As a result, the crude divorce rate for 2003 remained the same as in 2002, at 223.7 divorces for every 100,000 people in the population.

The 1.0% increase in the number of divorces across Canada is primarily due to

a 5.1% increase in the number of divorces in Ontario and a 1.4% increase in Quebec between 2002 and 2003. Prince Edward Island and Saskatchewan were the only other provinces to experience an increase in the number of divorces between these years. Newfoundland and Labrador showed the largest percentage decrease by far in the number of divorces, down 21.4%.

Repeat divorces, involving people who had been divorced at least once before, are accounting for an increasing proportion of divorces.

In 1973, only 5.4% of divorces involved husbands who had previously been divorced. Thirty years later this proportion has tripled to 16.2% of all divorces.

The proportion of divorces involving wives who had previously been divorced is similar, rising from 5.4% to 15.7% over this thirty year period.

Marriage stability can be assessed using divorce rates based on years of marriage. The proportion of marriages expected to end in divorce by the 30th wedding anniversary inched up to 38.3% in 2003, from 37.6% in 2002.

The divorce rate varies greatly depending on how long couples have been married, rising rapidly in the first few years of marriage. The peak divorce rate in 2003 occurred after three years of marriage, when 26.2 out of 1,000 marriages ended in divorce. The risk of divorce decreased slowly for each additional year of marriage.

The custody of dependents, the vast majority of whom are children aged 18 and under, was granted through divorce court proceedings in 27% of 2003 divorces

In the remaining divorces, couples arrived at custody arrangements outside the divorce proceedings, or they did not have dependents. The number of dependents in these divorces is not available.

There has been a 17-year trend of steady increases in joint custody arrangements. Of the 33,000 dependents for which custody was determined through divorce proceedings in 2003, 43.8% were awarded to the husband and wife jointly, up 2.0% from 2002. Under a joint custody arrangement, dependents do not necessarily spend equal amounts of their time with each parent.

The custody of 47.7% of dependents was awarded to the wife and 8.3% to the husband in 2003. In 2002, these percentages were 49.5% and 8.5%, respectively.

The shelf tables Divorces, 2003 (84F0213XPB, \$22) are now available.

For general information or to order custom tabulations, contact Client Custom Services (613-951-1746; hd-ds@statcan.ca). To enquire about the concepts, methods or data quality of this release, contact Brent Day (613-951-4280; brent.day@statcan.ca) or Patricia Tully (613-951-1759; patricia.tully@statcan.ca), Health Statistics Division.

AFTER

Divorces – 2003

Repeat divorces, those involving people who had been divorced at least once before, are accounting for an increasing proportion of divorces in Canada, according to new data.

In 1973, only 5.4% of divorces involved husbands who had previously been divorced. Some 30 years later, this proportion has tripled to 16.2% of all divorces. Similarly, the proportion of divorces involving wives who had previously been divorced rose from 5.4% to 15.7% during this three-decade period.

The number of couples getting a divorce in 2003 edged up 1.0% from a year earlier to 70,828. This slight increase was due primarily to a 5.1% jump in divorces in Ontario, and a 1.4% increase in Quebec. Prince Edward Island and Saskatchewan were the only other provinces to experience an advance.

The number of divorces fell 21.4% in Newfoundland and Labrador, by far the largest decline. No information on the reason for this decrease is available.

The number of divorces has remained relatively stable over the last few years.

The year-to-year change has been below 2% since 1999. The slight rise in divorces in 2003 kept pace with the increase in the Canadian population.

Divorces

	2002	2003	2002 to 2003
	number		% change
Canada	70,155	70,828	1.0
Newfoundland and Labrador	842	662	-21.4
Prince Edward Island	258	281	8.9
Nova Scotia	1,990	1,907	-4.2
New Brunswick	1,461	1,450	-0.8
Quebec	16,499	16,738	1.4
Ontario	26,170	27,513	5.1
Manitoba	2,396	2,352	-1.8
Saskatchewan	1,959	1,992	1.7
Alberta	8,291	7,960	-4.0
British Columbia	10,125	9,820	-3.0
Yukon	90	87	-3.3
Northwest Territories	68	62	-8.8
Nunavut	6	4	-33.3

Total divorce rate, by the 30th wedding anniversary

	2002	2003	2002 to 2003
	per 100 marriages		increase/decrease
Canada	37.6	38.3	0.7
Newfoundland and Labrador	21.8	17.1	-4.7
Prince Edward Island	25.2	27.3	2.1
Nova Scotia	30.4	28.9	-1.5
New Brunswick	27.2	27.6	0.4
Quebec	47.6	49.7	2.1
Ontario	34.9	37.0	2.1
Manitoba	30.3	30.2	-0.1
Saskatchewan	28.7	29.0	0.3
Alberta	41.9	40.0	-1.9
British Columbia	41.0	39.8	-1.2
Yukon	43.4	40.0	-3.4
Northwest Territories and Nunavut ¹	31.2	27.6	-3.6

1. Northwest Territories and Nunavut are combined to calculate the rates in this table because marriage and divorce data are not available for these territories separately for the 30-year period required for the calculation of the total divorce rate.

As a result, the crude divorce rate for 2003 remained stable at 223.7 divorces for every 100,000 people in the population.

Marriage stability can be assessed using divorce rates based on years of marriage. The proportion of marriages expected to end in divorce by the 30th wedding anniversary inched up to 38.3% in 2003, from 37.6% in 2002.

The divorce rate varies greatly depending on how long couples have been married. It rises rapidly in the first few years of marriage. The peak divorce rate in 2003 occurred after three years of marriage, when 26.2 out of 1,000 marriages ended in divorce.

The risk of divorce decreased slowly for each additional year of marriage.

The custody of dependents, the vast majority of whom are children aged 18 and under, was granted through divorce court proceedings in 27% of 2003 divorces.

Available on CANSIM: table 053-0002. Definitions, data sources and methods: survey number 3235.

The shelf tables Divorces, 2003 (84F0213XPB, \$22) are now available. For general information or to order custom tabulations, contact Client Custom Services (613-951-1746; hd-ds@statcan.ca). To enquire about the concepts, methods or data quality of this release, contact Brent Day (613-951-4280; brent.day@statcan.ca) or Patricia Tully (613-951-1759; patricia.tully@statcan.ca), Health Statistics Division.

Examples of well-written statistical stories

There are many sources of well-written stories and this guide can only touch on some. You can find more examples on the Internet, in newspapers and in statistical publications. Here are a few areas to start looking:

- Statistics Norway publishes their Statistical Magazine online. It features a wide range of topics and shows examples of clear tables and graphics.<http://www.ssb.no/english/magazine/>
- The United States Bureau of Justice Statistics website links to their online publications and press releases.<http://www.ojp.usdoj.gov/bjs/>
- The United Kingdom's Office of National Statistics has a 'Virtual Bookshelf' that provides quick access to their online press releases, papers and publications, sorted by theme.<http://www.statistics.gov.uk/onlineproducts/>
- Statistics Netherlands regularly publishes short articles on the Internet as part of their 'Webmagazine' series. The articles show how to incorporate graphics to make the message clear.
<http://www.cbs.nl/en-GB/menu/publicaties/webpublicaties/webmagazine/>
- Statistics Canada has a section on their website called 'The Daily'. Here you will find many examples of brief articles and press releases.<http://www.statcan.ca/english/dai-quo/>
- Look at websites of other statistical agencies. A good starting point is the UNECE's list of links to national and international agencies.<http://www.unece.org/stats/links.htm>

References

Few, Stephen. (2004), *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Oakland, CA: Analytics Press.s

Kosslyn, Stephen M.(1994), *Elements of Graph Design*, New York: W. H. Freeman and Company.

Miller, Jane E., (2004), *The Chicago Guide to Writing About Numbers*, The University of Chicago Press,

Tufte, Edward R., (1997), *Visual Explanations Cheshire*, CN: Graphics Press.

Tufte, Edward R., (1990), *Envisioning information*, Cheshire, CN: Graphics Press.

Tufte, Edward R., (1983), *The Visual Display of Quantitative Information*, Cheshire, CN: Graphics Press.

Truss, Lynne, *Eats, Shoots, and Leaves: The Zero Tolerance Approach to Punctuation*, (London: Profile Books Limited, 2003)

UNECE, *Communicating with the Media: A guide for statistical organizations*, (United Nations, Geneva, 2004) <http://www.unece.org/stats/documents/media/guide/>

UNECE (2005), *Making Data Meaningful: A guide to writing stories about numbers*, Geneva

Wallgren, Anders; Wallgren, Britt; Persson, Rolf; Jorner, Ulf; and Haaland, Jan-Aage, *Graphing Statistics & Data: Creating Better Charts*, (Thousand Oaks: SAGE Publications, 1996).



Annexes

Authors Vitæ	294
Abbreviations	303
What is DevInfo?	311

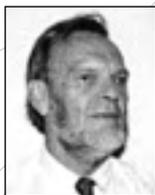
AUTHORS VITÆ



ADRIEN, Marie-Hélène is President of Universal Management Group, a Canadian consulting firm specializing in evaluation and project management. She was the President of the International Development Evaluation Association (IDEAS) from 2005 to 2008. Dr. Adrien has 20 years of consulting experience in evaluation, organizational capacity development and training, representing work in 36 countries around the world. She has published a number of articles and books on evaluation, including *Organizational assessment: 25 years of lessons learned* (2005), a *Framework for improving performance* (2002), and *Enhancing organizational performance. A toolbox for self-assessment* (1999 and 2000).



BAER, Petteri has been working since March 2006 as Regional Adviser at the Statistical Division of the United Nations Economic Commission for Europe (UNECE) in Geneva, Switzerland. Before that he worked for 13 years in different positions at the National Statistical Office of Finland. In October 1992 he started as a marketing planner, and was soon promoted to act as Head of the Regional Services of that institution. In 2001 Mr. Baer was appointed to take the post of Marketing Manager at Statistics Finland. Mr. Baer's working career also includes elements other than statistics. He has worked in the publishing sector, as marketing manager in two publishing houses and as director of the culturally oriented bookshop "Gogol's Nose" in the city of Helsinki. In all parts of his working career he has creatively implemented new ideas and approaches to marketing.



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 Programme (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 Programme; and for several private consulting firms. Professor 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.



FEINSTEIN, Osvaldo is advisor to the Spanish Evaluation Agency and Professor at the Masters Degree Programme on Evaluation at the Universidad Complutense de Madrid. Professor Feinstein is a member of the monitoring and evaluation panel of the Scientific Council of CGIAR (the Consultative Group on International Agricultural Research) and evaluation consultant with the World Bank; the International Fund for Agricultural Development (IFAD); the Global Environment Fund (GEF); UNDP; CEPAL; ILPES and ILO. He was a former manager at the World Bank's Evaluation Department and former IFAD's senior evaluator responsible for Latin America, where he created PREVAL (Latin American programme for the development of evaluation capacities). Feinstein is also a former professor at the FLACSO Ecuador Master in Development Studies. He has worked in monitoring and evaluation and development 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.



GIOVANNINI, Enrico graduated in Economics at "La Sapienza" University of Rome. He continued his studies at the Institute of Economic Policy of 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 back to the Italian National Institute of Statistics. 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 the Statistics on Institutions and Enterprises. Since January 2001 Professor Giovannini has been the Director of Statistics and Chief Statistician of OECD. He is a full professor of economic statistics at the Rome University “Tor Vergata”.



JOBIN, Denis is a Canadian expert in the fields of programme evaluation, performance measurement and performance audit. He is the Vice President of the International Development Evaluation Association (IDEAS) and he currently manages the evaluation unit of the National Crime Prevention Center – Department of Public Safety Canada, delivering impact evaluation studies. Mr. Jobin has been involved in evaluation-related activities for more than 13 years, having worked for the Quebec Provincial Government (Department of Industry and Trade), and the Canadian government (in the Departments of Health and Environment Canada). In addition, he has a solid experience in performance auditing, having worked for the Office of the Auditor General of Canada (OAGC). He also worked and resided in West Africa. Mr. Jobin also sponsors the Theory-based evaluation discussion group (<http://groups.yahoo.com/group/TheoryBasedEvaluation/>), and has contributed to and authored several publications related to evaluation.



KENNEDY, Megan Grace is a consultant with the OECD DAC Network on Development Evaluation in Paris, France, focusing on evaluation capacity development and formulating guidance on evaluating peace-building activities. Ms. Kennedy is completing a Masters of Public Administration in International Management at the Monterey Institute of International Studies in Monterey, California, USA. She holds a Bachelors degree in Economics and in Peace & Global studies from Earlham College and received a Thomas J. Watson Fellowship for independent development research in 2004. She has held various programme management positions, notably in the US, Mexico, and Tanzania.



KHAYRI BA TALL, Oumoul is currently the president of the International Organization for Cooperation in Evaluation (IOCE) (2008-10), 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 delivered speeches on topics such as aid and development, and evaluation capacity. She has 21 years of professional experience in various but related field from auditing, accounting, evaluation, organisational development, micro-entreprise, micro-finance, 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. Khayri Ba Tall is an MBA (1995) and 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 currently heads up the Bank's Global HIV/AIDS Monitoring and Evaluation Group (GAMET) which aims to strengthen the use of HIV/AIDS data to support national and sub-national policy and programme decision-making in over 50 countries, world-wide. Previously, 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 which is now in use, Bank-wide. Earlier, Ms. Kusek worked for the Clinton-Gore Administration in the United States, designing and implementing the Government Performance and Results Act. She is 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.



LUNDGREN, Hans manages the OECD/ DAC Network on Development Evaluation which brings together evaluation managers and experts from 30 bilateral and multilateral development agencies. He joined the OECD in 1987 and has since worked on development policy and aid effectiveness issues, with an increasing focus over time on development evaluation. He has published on evaluation systems in aid agencies and written a number of articles on DAC's evaluation work. He led the drafting of the DAC Principles for evaluation of development assistance and co-ordinated the work on the DAC Glossary of terms in evaluation and results-based management, and the DAC Evaluation quality standards. He also has experience from international expert reviews of monitoring and evaluation systems and is a member of UNESCO's Oversight Advisory Committee. Prior to joining the OECD in 1987, he worked in field operations with the UNDP in West Africa, and at UNESCO headquarters managing trust fund operations.



MACKAY, Keith is a senior evaluation officer in the Independent Evaluation Group of the World Bank, where he is also the coordinator for evaluation capacity development. His current work is focused on helping countries strengthen their national monitoring and evaluation systems to support a performance orientation within their public sectors. Countries with which he is currently working include Brazil, Chile and Colombia. Before joining the Bank in 1997, Mr. Mackay worked for 22 years in the Australian government, including 11 years in the Department of Finance. From 1991 to 1997 he was the senior adviser to the government on its national evaluation strategy. He has written 75 articles, papers and books, principally on monitoring and evaluation.



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.



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. Mr Oswald 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. Mr Oswald 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; and, the Management Sciences for Health and U.S. Library of Congress.



PICCIOTTO, Robert is a graduate of the Woodrow Wilson School of Public and International Affairs (Princeton University). He is Visiting Professor at Kings College, London. He sits on the council of the United Kingdom Evaluation Society and on the board of the European Evaluation Society. At the World Bank, he served as Vice President for Corporate Planning and Budgeting and, for ten years, as Director-General, Evaluation, reporting directly to the executive directors. Prior to this, he held senior operational management assignments in three of the World Bank's regions. Since 2002, Professor Picciotto has been a senior evaluation adviser to governments and international institutions. He currently serves as a member of the International Advisory Committee on Development Impact set up by the UK Secretary of State for International Development and acts as a trustee of the Oxford Policy Institute.



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. Mr. 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. Mr Pron 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 Professor 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 was also Director of Policy Coordination, Management Improvement and International Development Programmes in Asia, Africa, Latin America and the Caribbean.



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, and 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, and he has been an advisor to senior government officials in more than 50 countries. Professor Rist is currently an advisor to the World Bank, co-director of the International programme 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.



RUGH, Jim has been professionally involved for 44 years in rural community development in Africa, Asia, Appalachia and other parts of the world, specializing in international programme evaluation for 28 years. In 2007 he retired after serving for 12 years as head of Design, Monitoring and Evaluation for Accountability and Learning for CARE International. Rugh is recognized as a leader in evaluation among colleagues in the international NGO community, including InterAction's Evaluation Interest Group, as well as active involvement for many years in the International and Cross-Cultural Evaluation Topical Interest Group of the American Evaluation Association (AEA). He currently serves as the AEA

representative to the International Organization for Cooperation in Evaluation (IOCE). He co-authored the popular and practical *Real-World Evaluation* book and has led numerous workshops on that topic for many organizations and networks in many countries.



SAKVARELIDZE, George is a monitoring and evaluation specialist at the UNICEF Regional Office for CEE/CIS. He studied Pediatrics in Georgia, Tbilisi and earned a Master degree in Public Health in USA, New York at the School of Public Health in Albany. He worked with UNICEF in health and monitoring and evaluation fields. Since 2005 he has been the Regional Coordinator for Multiple Indicator Cluster Survey in CEE/CIS, coordinating 13 surveys. He also delivers technical assistance for DevInfo implementation in the Region.



SEGONE, Marco has been serving as the Senior regional advisor, Monitoring and Evaluation in the UNICEF Regional Office for Central and Eastern Europe and the Commonwealth of Independent States (CEE/CIS) since 2005. He represents UNICEF on the Board of Trustees of the International Programme Evaluation Network (IPEN). During his 17 years in international development, Segone has worked in Bangladesh, Pakistan, Thailand, Uganda and Albania in integrated development projects. In 1996 he joined UNICEF to work for the Regional UNICEF Office for Latin America and the Caribbean. From 1999 to 2001 he worked as Monitoring and Evaluation Officer for UNICEF Niger, where he founded, and for two years coordinated, the Niger Monitoring and Evaluation Network (ReNSE). From 2001 to 2004 he was the UNICEF Monitoring and Evaluation Officer for 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). Mr Segone has authored / edited about 30 books and articles, including *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*.



VADNAIS, Daniel joined UNICEF Headquarters at the end of 2006 as Data Dissemination Specialist. Prior to that, Mr. Vadnais worked for 12 years with the Demographic and Health Surveys (DHS) project as Deputy Advisor for Communication, with a focus on the dissemination of findings. He also worked closely with media representatives. Mr. Vadnais provided technical assistance in numerous countries throughout Asia and Africa. In 2006, he contributed to the publication of *Women's lives and experiences: Changes in the past 10 years*. Before that, he co-wrote *Connecting people to useful information: guidelines for effective data presentations* with members of the Dissemination working group of the MEASURE Programme. Mr. Vadnais also worked as Information officer for the Global Committee of Parliamentarians on Population and Development. In 1989 -1990, after coordinating the local arrangements of the Moscow Global Forum on Environment and Development, he served as Public Affairs Officer for Religious and Parliamentary Affairs at UNICEF/New York, at the time of the World Summit for Children. With UNICEF, he helped organize the first global inter-faith conference to focus solely on children's issues which took place at Princeton University. Mr. Vadnais, a native from Québec, holds a Masters Degree in Demography from the University of Montreal.

ABBREVIATIONS

ADB	Asian Development Bank
AEA	American Evaluation Association
AfrEA	African Evaluation Association
CEE	Central and Eastern Europe
CIS	Commonwealth of Independent States
CES	Canadian Evaluation Society
CGD	Center for Global Development
CLE	Country-Led Evaluation
CLEF	Country-Led Evaluation Fund
CLES	Country-Led Evaluations Systems
CLIE	Country-Led Impact Evaluations
CoP	Communities of Practice
CPI	Consumer Price Index
CsOs	Civil Society Organisations
DAC-OECD	Development Assistance Committee of the Organization for Economic Cooperation and Development
DEReC	An on-line evaluation resource centre
DHS	Demographic and Health Surveys
ECD	Evaluation Capacity Development
ECG	Evaluation Cooperation Group
EGPRSP	Economic Growth and Poverty Reduction Strategy Paper
EO/UNDP	Evaluation Office of the United Nations Development Programme
EU	European Union
GBS	General Budget Support
GDP	Gross Domestic Product
IDEAS	International Development Evaluation Association
IHSN	International Household Survey Network
IOB	Dutch Ministry of Foreign Affairs
IOCE	International Organization for Cooperation in Evaluation
IPDET	International Programme for Development Evaluation Training
LFA	Logical Framework Analysis

LPA	Local Plans of Action for Children
MBO	Management by Objectives
MDGs	Millennium Development Goals
MICS	Multiple Indicator Cluster Surveys
MICS3	Multiple Indicator Cluster Surveys – third round
MfDR	Management for Development Results
MoET	Ministry of Economy and Trade
M&E	Monitoring and evaluation
MES	Malaysian Evaluation Society
NGO	Non-Government Organization
NONIE	Network of Networks for Impact Evaluation
NPA	National Plan of Action for Children
NSOs	National Statistical Offices
OECD	Organization for Economic Cooperation and Development
OECD-DAC	Development Assistance Committee of the Organization for Economic Cooperation and Development
ORET/ MILIEV	Development and Environment Related export Transactions
PRSP	Poverty Reduction Strategy Papers
QED	Quasi-experimental design
RéNSE	Réseau Nigérien de Suivi et Evaluation (Niger monitoring and evaluation network)
RWE	Real World Evaluation
SEDESOL	Mexican Secretariat for Social Development
SFE	Société Française d’Evaluation (French Evaluation Society)
SORS	Statistical Office of the Republic of Serbia
TOR	Terms of Reference
TRIPS	Trade Related Intellectual Property Rights
UNDAF	United Nations Development Assistance Framework
UNDGO	United Nations Development Group Office – now UNDOCO
UNDOCO	United Nations Development Operations Coordination Office – formerly UNDGO
UNDP	United Nations Development Programme
UNECE	United Nations Economic Commission for Europe

UNEG	United Nations Evaluation Group
UNFPA	United Nations Population Fund
UNGASS	United Nations General Assembly Special Session
UNICEF	United Nations Children's Fund
UNICEF IRC	UNICEF Innocenti Research Center
3ie	International Initiative for Impact Evaluation

WHAT IS DEVINFO?

DevInfo is a powerful database system which monitors progress towards the Millennium Development Goals and Human Development. It generates tables, graphs and maps for reports and presentations. DevInfo has been developed by United Nations organizations. It was adapted from UNICEF ChildInfo technology. The database maintains indicators, by time periods and geographical areas, to monitor commitments to sustained human development.

UNICEF Regional Office for Eastern and Central Europe and the Commonwealth of Independent States developed three regional databases. The Regional MDGInfo database, developed in cooperation with UNECE and UNDP, makes MDGs as well as regionally specific indicators easily available. It is accessible at www.regionalmdg.org. The MICSInfo database presents the key findings of the third round of Multiple Indicators Clusters Surveys carried out in 12 countries in the region, with data disaggregated by regions, urban and rural, ethnicities, wealth quintiles, mother's education and age of children. It is accessible at www.micsinfo.org. Last but not least, the MoneeInfo database makes data on the situation of children and women, with a specific focus on child protection, easily accessible at www.moneeinfo.org.

All three databases are now available in the CD ROM attached to this report. In the CD ROM, you can also download ready-made graphs and maps on key indicators, the full database in Excel format and produce your own maps, graphs and table using the DevInfo technology.

For additional information on DevInfo, and a quick guide on how to produce maps, graphs and tables using the DevInfo technology, please visit www.devinfo.org.

Instructions on installation and use of DevInfo

Ready-made graphs and maps on the key indicators, as well as the full database in Excel format, are accessible immediately. To produce your own maps, graphs and table using the DevInfo technology, you need to install DevInfo in your computer. Below the instructions.

System requirements for DevInfo

The recommend hardware requirements to install this software application are:

- Pentium IV
- 512 MB of RAM
- 1 GB of free hard disk
- Display resolution 1024 x 768
- Microsoft Windows XP
- Microsoft Office XP
- Microsoft Internet Explorer 6.0

Installing DevInfo

To install this software application on your computer, follow the steps given below:

- Insert the CD ROM into the drive of your computer
- Wait for auto-run to open the setup screen
- Click on the icon “Database”
- Follow the instructions on screen to complete the setup
- Double-click on the DevInfo icon on desktop to start the application

If the setup program does not load automatically:

- Choose Start | Run
- Type d:\setup where d is the letter of your CD-ROM drive and press Enter key
- Follow the instructions on screen to complete setup
- Double-click on the DevInfo icon on desktop to start the application

Note: Computers with Windows 98 Operating System need to be restarted after installing DevInfo.

UNICEF Regional Office for CEE/CIS
Palais des Nations
CH 1211 Geneva 10
Switzerland
www.unicef.org/ceecis

2009