



Bridging the gap

The role of monitoring and evaluation
in Evidence-based policy making

In partnership with:



The Evaluation Working Papers (EWP) are documents that present strategic evaluation findings, lessons learned and innovative approaches and methodologies. We would like to encourage you to propose relevant papers that 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 Monitoring and Evaluation Advisor, msegone@unicef.org.

ISSUE #1: Regional strategy to strengthen the monitoring & evaluation function in CEE/CIS, 2005.

ISSUE #2: Comparative analysis of major trends in evaluations and studies supported by UNICEF 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, 2006 (joint paper with IPEN, with prefaces by Presidents of IDEAS and IOCE)

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.

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.

*Photo Credit:
UNICEF/SWZK00095/Pirozzi
ROMANIA 2004.*

Bridging the gap

The role of monitoring and evaluation
in evidence-based policy making

EDITOR

Marco Segone

AUTHORS

Marie-Helene Adrien

Michael Bamberger

Ross F. Conner

Dragana Djokovic-Papic

Attila Hancioglu

Vladica Jankovic

Dennis Jobin

Ove Karlsson Vestman

Jody Zall Kusek

Keith Mackay

Debora McWhinney

David Parker

Oliver Petrovic

Nicolas Pron

Marco Segone

Ray Rist

Mohamed Azzedine Salah

Daniel Vadnais

Vladan Vasic

Azzedina Vukovic

Contents

Editorial

Part 1: Evidence-based policy making

Evidence-based policy making and the role of monitoring and evaluation within the new aid environment, by Marco Segone, Senior Regional Monitoring and Evaluation Advisor, UNICEF CEE/CIS and past Vice President, IOCE	16
The relationship between evaluation and politics, by Ove Karlsson Vestman, Director, Mälardalen Evaluation Academy, and Vice President of the Swedish Evaluation Society; and Ross F. Conner, University of California, former President of the American Evaluation Association and current President of IOCE	46
Monitoring and evaluation, and the knowledge function, by David Parker, Deputy Director, UNICEF Innocenti Research Center	73
Helping countries build government monitoring and evaluation systems. World Bank contribution to evidence-based policy making, by Keith Mackay, Coordinator, Evaluation Capacity Development, Independent Evaluation Group, World Bank	88
Ten step to a results based monitoring and evaluation systems, by Jody Zall Kusek, Chief, Global HIV/AIDS Monitoring and Evaluation Group, World Bank, and Ray Rist, Advisor, Public sector management	98

Part 2: The strategic intent of evaluations, studies and research

Enhancing the utilization of evaluations for evidence-based policy making, by Michael Bamberger, former Senior Evaluator, World Bank	120
Country-Led Evaluation, by Marie-Helene Adrien, President, IDEAS and Dennis Jobin, Vice President, IDEAS	143
Joint Country-Led Evaluation of the policies related to child well-being within the social protection sector in Bosnia and Herzegovina, by Azzedina Vukovic, Directorate for Economic Planning, Council of Ministers of Bosnia & Herzegovina, and Debora McWhinney, Deputy Representative, UNICEF Bosnia & Herzegovina	154

Part 3: The strategic intent of data collection and dissemination

The strategic intent of data collection and analysis. The case of Multiple Indicator Cluster Surveys (MICS), by Daniel Vadnais and Attila Hancioglu.....	168
The strategic intent of data dissemination. The case of DevInfo, by Nicolas Pron, DevInfo Global Administrator, Division of Policy and Planning, UNICEF Headquarters.....	185
Using DevInfo as a strategic tool for decision making. Achievements and lessons learned in Moldova, by Mohamed Azzedine Salah, Deputy Representative, UNICEF Moldova.....	195
Using DevInfo to support Governments in monitoring National Development Strategies. The case of the Republic of Serbia, by Dragana Djokovic-Papic, Statistical Office of the Republic of Serbia, and Oliver Petrovic and Vladica Jankovic, UNICEF Serbia.....	200
Using DevInfo as a strategic tool to facilitate local communities' empowerment. The case of Serbia, by Vladan Vasic, Mayor of Pirot, and Oliver Petrovic and Vladica Jankovic, UNICEF Serbia.....	203
 <i>Annexes</i>	
Annex I: Authors Vitae.....	210
Annex II: Abbreviations.....	217

Bridging the gap

The role of monitoring and evaluation in evidence-based policy making



Editorial

Evidence-based policy making

This publication offers a number of strong contributions from senior officers in institutions dealing with Evidence-based policy making. These institutions are national and local governments, UNICEF, the World Bank and the International Development Evaluation Association. It tries to bring together the vision and lessons learned from different stakeholders on the strategic role of monitoring and evaluation in evidence-based policy making. These stakeholders are policy-makers, in their role of users of evidence, and researchers and evaluators, in their role of suppliers of evidence.

The concept of 'evidence-based policy making' has been gaining currency over recent years. The use of strong evidence can make a difference to policy making in at least five ways:

- Achieve recognition of a policy issue.

The first stage in the process of policy formation occurs when the appearance of evidence reveals some aspect of social or economic life which had, until then, remained hidden from the general public and from policy-makers. Once this information is revealed, a variety of groups, such as civil servants, non-government organizations, development agencies or the media, lobby for a new policy issue to be recognised and addressed.

- Inform the design and choice of policy.

Once a policy issue has been identified, the next step is to analyze it, so that the extent and nature of the problem can be understood. This understanding provides the basis for any subsequent policy recommendations.

- Forecast the future.

Attempting to read the future is also required in order to know whether a policy measure taken to alleviate a problem in the short-run will be successful in the long-run as well. When a government is committed to attaining targets in the future, forecasting models allow an assessment of whether these targets are likely to be met.

- Monitor policy implementation.

Once policies are being executed, information is required by policy-makers to monitor the expected results associated with the poli-

cies. Careful monitoring can reveal when key indicators are going off-track, which prompts further analysis leading to a change of policy.

- Evaluate policy impact.

Measuring the impact of a policy intervention is more demanding of methodology and of information than is monitoring policy implementation. Incorporating an explicit mechanism for evaluating policy impact into the design of a policy is a key step to ensure its evaluability.

The role of monitoring and evaluation in evidence-based policy making

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? These and other issues are analyzed in this publication.

Segone introduces the concept of evidence-based policy making, exploring the apparent tension between authority and power on the one side, and knowledge and evidence on the other. He suggests that monitoring and evaluation should inform evidence-based policy options, to facilitate public argumentation among policy makers and societal stakeholders and facilitate the selection of policies. To do so, monitoring and evaluation should be both technically sound and politically relevant. Therefore, the dialogue between the suppliers and users of evidence should be strengthened to bridge the gap between the information needs of policy-makers and the information offered by researchers and evaluators.

Karlsson Vestman and Conner explore the relationship between evaluation and politics. They explain that the evaluation family traditionally has included good researchers with their ideal of neutral, objective research as the prototype for evaluation. Evaluation work, however, is always couched within a political context, and this reality brings different kinds of partners into the relationship. These partners, including politicians and policymakers, often make

the evaluation family uneasy. If politics and evaluation are destined to be “life partners”, then what forms could the relationship take – marriage, cohabitation or living apart? This chapter considers some of these possibilities.

Parker attempts to place monitoring and evaluation in the wider context of knowledge management, as an element of organizational learning and performance strengthening. It focuses on the case of UNICEF, within the UN system, which may be seen as representative of many other agencies working in the field of human and social development. It begins with a brief overview of the knowledge function, and examines the experience of monitoring and evaluation, pointing to strengths as well as gaps in the context of UNICEF. An example is then presented of a monitoring system linked to research and policy development in the region of Central and Eastern Europe and the Commonwealth of Independent States.

Mackay suggests that monitoring and evaluation is necessary to achieve evidence-based policy making, evidence-based management, and evidence-based accountability. Policy making, especially budget decision-making and national planning, focuses on government priorities among competing demands from citizens and groups in society. The information provided by monitoring and evaluation systems can support government’s deliberations by providing evidence about the most cost-effective types of policy options. Mackay presents how the World Bank, in cooperation with other partners, is supporting countries in building national monitoring and evaluation systems. He ends by presenting the major lessons learned in developing such systems over the past decade.

Zall Kuseck and Rist focus their contribution on how to design and implement a results-based monitoring and evaluation system. Over recent years, there has been an evolution in the field of monitoring and evaluation involving a movement away from traditional implementation-based approaches towards new results-based approaches. The latter help to answer the “so what” question. In other words, governments and organizations may successfully implement policies but the key question is, have these policies produced the actual, intended results? Have government and organizations delivered on promises made to their stakeholders? The introduction of a results-based monitoring and evaluation system takes decision-makers one step further in assessing whether and how policy goals are being achieved over time. These systems help to answer the all important “so what” questions, and respond to

stakeholders growing demands for results. The presented “Ten Steps Model” addresses the challenge of how governments in general, but those in developing and transition countries in particular, can begin to build results-based monitoring and evaluation systems so as to provide credible and trustworthy information for their own use and to share with their citizens.

The strategic intent of evaluations, studies and researches

Bamberger analyzes why the utilization of evaluation findings is disappointingly low, despite the significant resources devoted to evaluation and its growing importance in industrialized, transition and developing countries. The World Bank Independent Evaluation Group recently concluded that for all development agencies, monitoring and evaluation remains the weakest link in the risk management chain. The Swedish International Development Agency, in a recent assessment of evaluation practices, found that most stakeholders never even saw the findings and that few who did found nothing very new or useful. Bamberger therefore focuses his contribution on why evaluation findings are not used to the maximum extent possible and he explains how to ensure that evaluation does influence policy decisions.

Adrien and Jobin acknowledge that contemporary discussions in the international development arena have broadened the scope and design of evaluation, from an earlier, narrower focus on projects or programmes to broader assessments that encompass policy, policy coherence, and development outcomes. At the same time, there has been increasing pressure to make evaluation central to a country’s own development process and more relevant and meaningful to the people whose lives are affected by development interventions. The field of evaluation is being reshaped by the evolving context of international aid, and particularly, by the emerging recognition that effective development assistance requires that donor agencies respect partner country leadership and help strengthen their capacity to exercise it. Country-led evaluation is a relatively new concept, and one that reflects the world’s growing recognition of the importance of a nation’s self-determination in its own development. Conventional forms of evaluation, typically mandated and funded by development agencies, are now being challenged by emerging independent forms of assessment which put the recipient country in the driver’s seat. The rationale for Country-led evalu-

ation is clear, but the question now is how to do it. What are the obstacles to Country-led evaluation? What needs to be done to support it? Adrien and Jobin present an analysis of enabling factors and barriers to Country-led evaluation, based on the outcome of recent regional consultations in Eastern Europe and Africa.

Vukovic and McWhinney presents the experience of one of the first Country-led evaluations implemented worldwide: the Joint Country-led evaluation of the policies related to child well-being within the social protection sector in Bosnia and Herzegovina. They explain how it was possible to accommodate the information gap of both the Government and UNICEF, while ensuring an independent and objective evaluation process. The joint Country-led evaluation turned out also to be an opportunity to strengthen Government's evaluation capacity on the one side, while better positioning UNICEF within the social protection sector on the other side.

The strategic intent of data collection and dissemination

Data collection and dissemination are essential elements to both policy making and the evaluation function. Vadnais and Hancioglu explain the impact of a UNICEF supported international household survey initiative, designed to assist countries in filling data gaps for monitoring human development in general and the situation of children and women in particular. This survey, named Multiple Indicators Cluster Survey (MICS), has been instrumental in strengthening national statistics capacities, highlighting and filling gaps in quality data, monitoring and tracking progress toward national and international development goals and, in identifying emerging development issues and disparities among groups in societies.

However, collecting quality data is necessary but not sufficient. Data must also be disseminated in a user-friendly way to ensure that they are understood and used, and therefore inform policy decisions. Pron explains how a database system, DevInfo, has been a success story. This database system harnesses the power of advanced information technology to compile and disseminate data on human development. The system, which is endorsed by the UN Development Group to assist countries in monitoring achievement of the Millennium Development Goals, has been adapted and used by more than 80 governments and donor agencies and more than 10,000 professionals (approximately 60% government and 40%

UN), have been trained. Pron gives concrete examples of how DevInfo has been used to strengthen integrated national monitoring and evaluation systems and to enhance advocacy and public awareness on priority development issues. He describes DevInfo's use as a common system for evidence-based planning and results-based monitoring.

Salah, Djokovic-Papic, Petrovic, Jankovic and Vasic explain (in separate articles), how data disseminated through DevInfo have been instrumental in policy and budget allocation decisions in Moldova and Serbia, both at national and at decentralized level.

In Moldova, data presented through DevInfo are gradually playing a strategic role in facilitating a common understanding among the government, civil society organizations and development partners. Data analyses and maps are used as platforms for the national dialogue on poverty reduction. DevInfo is used to produce a bulletin on the national Poverty Reduction Strategy implementation, which is published regularly in Moldovan newspapers and posted on government websites. This national process was instrumental in the government's decision to invest up to 21% more in the social sectors in 2006.

In Serbia, the National DevInfo database contains a rich set of indicators which are used to monitor the Millennium Development Goals, the Poverty Reduction Strategy, and the National Plan of Action for Children. In addition, DevInfo is used to inform policy decisions at municipal level. In the municipality of Pirot, DevInfo highlights important social trends which may otherwise be overlooked. The first DevInfo report was quite shocking in terms of the number of children left out of the education system. Together with local situation analysis, DevInfo helped to reveal that social services had overlooked many children, in particularly Roma children and those with disabilities. Now a local team of Roma representatives and educational experts are working to prepare the ground for continuing education of Roma children. A further use of DevInfo is for review of the municipality budget allocation. Informed by DevInfo data, an increasing demand from the local population for a better quality of child social services prompted local authorities to provide additional funds. As a result, investment for children was increased 7 fold in just two years.

Marco Segone
Editor



Bridging the gap

The role of monitoring and evaluation in evidence-based policy making



Part 1

Evidence-based policy making

Evidence-based policy making and the role of monitoring and evaluation within the new aid environment, by Marco Segone, Senior Regional monitoring and evaluation Advisor, UNICEF CEE/CIS and past Vice President, IOCE	16
The relationship between evaluation and politics, by Ove Karlsson Vestman, Director, Mälardalen Evaluation Academy, and Vice President of the Swedish Evaluation Society; and Ross F. Conner, University of California, former President of the American Evaluation Association and current President of IOCE.....	46
Monitoring and evaluation, and the knowledge function, by David Parker, Deputy Director, UNICEF Innocenti Research Center	73
Helping countries build government monitoring and evaluation systems. World Bank contribution to evidence-based policy making, by Keith Mackay, Coordinator, Evaluation Capacity Development, Independent Evaluation Group, World Bank.....	88
Ten step to a results based monitoring and evaluation systems, by Jody Zall Kusek, Chief, Global HIV/AIDS Monitoring and Evaluation Group, World Bank, and Ray Rist, Advisor, Public sector management	98

EVIDENCE-BASED POLICY MAKING AND THE ROLE OF MONITORING AND EVALUATION WITHIN THE NEW AID ENVIRONMENT

*by Marco Segone, Senior Regional monitoring and
evaluation Advisor, UNICEF CEE/CIS and past Vice President, IOCE*

The Paris Declaration and the new aid environment¹

Aid is in the spotlight as never before. Following the recent commitments made at the G8 Summit at Gleneagles (2005), the UN Millennium Summits (2005 and 2006) and the EU, the amount of aid provided to Least Developed Countries is expected to increase by nearly 60% (about an additional USD 50 billion), by 2010. Yet, in the absence of major improvements in the quality of aid, budget increases will not help to reduce poverty. What is required is ambitious reforms in the aid system. Action is required not only from donors, who could do a much better job at delivering aid more effectively, but also from developing countries who could improve the way they manage aid. For many years, reforms in these areas have been slow to materialise and, all too often, it has been business as usual within the development community.

Today, however, there are good reasons for believing the situation will change. For example, on 2 March 2005, over one hundred donors and developing countries agreed, in Paris, to undertake landmark reforms in the way they do business together. The Paris Declaration marks an unprecedented level of consensus and resolve, to reform aid to increase its effectiveness at combating global poverty.

1 Based on OECD, 2005

Box 1: Representation at the Paris High-Level Forum

The Paris Forum on Aid Effectiveness – in which the Paris Declaration was agreed – was hosted by the French government and was co-sponsored by eight organisations who were represented at the highest level:

- OECD – Secretary-General Donald Johnston and Chair of the Development Assistance Committee, Mr. Richard Manning.
- World Bank – President James Wolfensohn.
- United Nations Development Programme – Administrator Mark Malloch Brown.
- Asian Development Bank – President Haruhiko Kuroda.
- African Development Bank – President Omar Kabbaj.
- European Bank for Reconstruction and Development – President Jean Lemierre.
- Inter-American Development Bank – Chief Development Effectiveness Officer, Mr. Manuel Rapoport.

The meeting was also attended by President Enrique Bolaños (Nicaragua), Commissioner Louis Michel (EC), more than 60 ministers and many other heads of agencies and high level officials.

One reason why reform to the aid system has been slow to materialise is the weakness of accountability mechanisms within this system. All too often, neither donors nor developing country governments are truly accountable to their citizens on the use of development resources. Significant progress towards making aid more effective requires stronger mechanisms for accountability for both donors and partner countries. The Paris Declaration seeks to address this “accountability gap” by promoting a model of partnership to improve transparency and accountability on the use of development resources.

From donorship to ownership. Aid has more impact when there is strong and effective leadership by partner countries on their development policies and strategies. Ownership is therefore the fundamental tenet underpinning the Paris Declaration. It means that governments of developing countries are accountable for their development policies to their own parliaments and citizens, not to donor organizations. In many countries, this requires a strengthening of parliamentary oversight of development policies and budgets

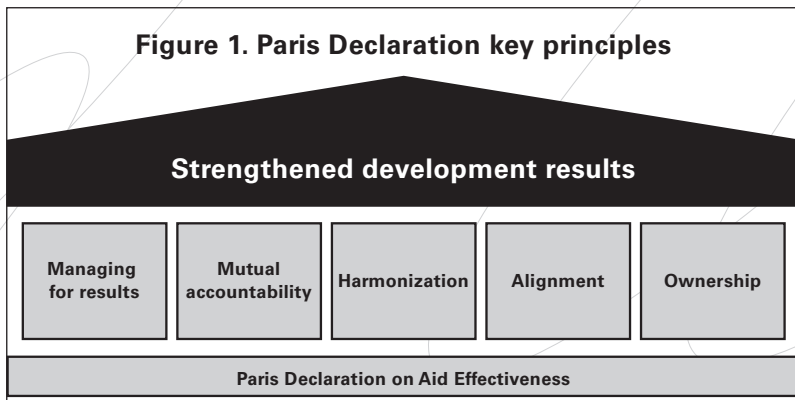
and reinforcing the role of civil society. It also requires donors to scale down their sometimes excessive demands for accountability from developing countries by:

- Relying on country systems and procedures, as much as possible.
- Avoiding intrusive conditionality.
- Decreasing the number of project implementation units (PIUs), since they can undermine national administrations.
- Providing timely and transparent information on aid flows, so as to enable partner authorities to present comprehensive budget reports to their legislature and citizens.

Stronger and more balanced mechanisms for mutual accountability. At present accountability requirements are often stricter for developing countries than for donors. The Paris Declaration recognises that for aid to become truly effective, stronger, more balanced accountability mechanisms are required at different levels. At the international level, the Paris Declaration provides a mechanism in which aid donors and recipients are held mutually accountable to each other and, in which compliance in meeting the commitments will be publicly monitored.

Compared with previous agreements, the Paris Declaration goes beyond a statement of general principles and lays down a practical, action-oriented roadmap to improve the quality of aid and its impact on development.

The Paris Declaration includes 56 partnership commitments which are organised around five key principles:



- **Managing for results** – Both donor and partner countries will manage resources and improve decision-making to focus on required results. Donors should fully support developing countries efforts in implementing performance assessment frameworks to measure progress against key elements of national development strategies.
- **Mutual accountability** – Donors and developing countries pledge that they will hold each other mutually accountable for development results as outlined in the aid effectiveness pyramid below.
- **Harmonisation** – Donors aim to be more harmonised, collectively effective and less burdensome especially on countries, such as fragile states, which have weak administrative capacities. This means, for instance, establishing common arrangements, at country level, for planning, funding and implementing development programmes.
- **Alignment** – Donors will base their overall support on partner countries' national development strategies, institutions and procedures. For example, donors will, wherever possible, use conditions from the development strategy of the developing country government, instead of imposing multiple conditions based on other agendas.
- **Ownership** – Developing countries will exercise effective leadership over their development policies and strategies and, co-ordinate development efforts themselves. Donors are responsible for supporting and enabling developing countries' ownership by respecting their policies and helping strengthen their capacity to implement them.

Unless donors change the way they provide aid and, unless developing countries enhance the way they currently manage it, increased aid flows are unlikely to make a serious dent into global poverty. Business as usual will erode the credibility of development assistance in the North and South, and, more importantly, undermine the international community's ability to reach the Millennium Development Goals (MDGs), by 2015. Disappointing results could make aid, not poverty, history. (Institute of Development Studies, 2005)

This is why the Paris Declaration challenge is to reform the way donor and partner countries work together on common objectives to make best use of limited development resources. Put simply, the Paris Declaration is about changing behaviour. Taken together,

the agenda set out by the Paris Declaration and the strengthened mechanisms for mutual accountability create powerful incentives to change behaviour patterns.

The evolving role of the evaluation function within the new aid environment

Within the changing development framework, the evaluation function is expected to play a strategic role. It should better serve the increased demand for mutual accountability, evidence for decision-making and learning. Although the process of reshaping the evaluation function is just beginning, in order to stimulate debate, it is desirable to attempt to formulate key trends, using as framework the Paris Declaration's five principles.

Managing for results



Due to the greater attention to quality, in particular, in MDG-based national development plans such as the Poverty Reduction Strategies (PRS), the focus of evaluation is shifting from small projects to national programmes and policies. This shift requires a systemic approach to evaluation so that policy decisions can be informed not only by individual evaluation reports, but also by knowledge streams resulting from continuous analysis. Knowledge streams are produced by relevant, integrated monitoring and evaluation systems whose data inform major evaluations designed strategically to inform key decision-making milestones. To ensure such use of evaluations, monitoring and evaluation is being institutionalized within government institutions.

Mutual accountability

Paris Declaration Commitment

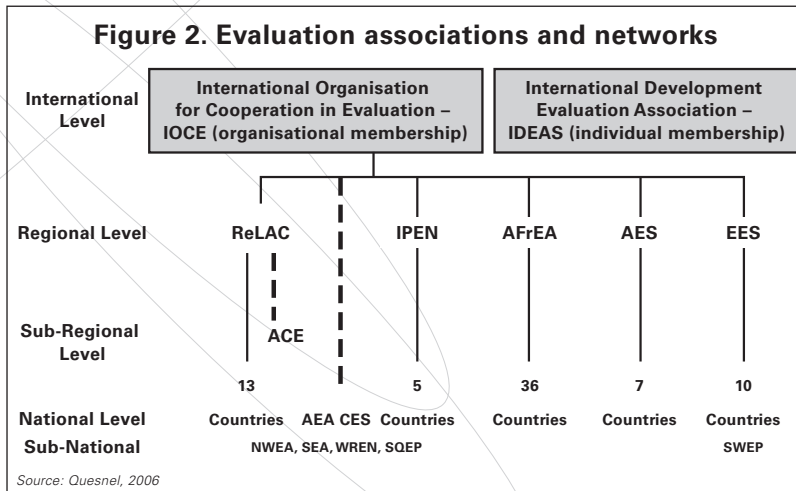
Partner countries reinforce participatory approaches by systematically involving a broad range of development partners when formulating and assessing progress in implementing national development strategies

Implications to the Evaluation Function

foster democratic approaches to evaluation, providing a forum for greater dialogue among civil society, academia, governments and donors; and reporting to Parliaments

In line with the commitment to reinforce participatory approaches in decision-making processes, the evaluation function should also embrace such principle. In this context, organizations for evaluation professional have a potentially significant role to play, more so given the dramatic growth in their number at national and regional level. In the last 10 years their number has increased from half a dozen in 1997 to more than 50 in 2007. Most of these new organizations are located outside Western Europe and North America. Two global organizations have also been created. These are the International Organization for Cooperation in Evaluation (IOCE) and the International Development Evaluation Association (IDEAS), (Segone, 2006).

Figure 2. Evaluation associations and networks



IOCE², launched in Peru in 2003, is the coalition of regional and national evaluation organizations from around the world. Membership is made up of organizations and not individuals. IOCE seeks to legitimate and strengthen evaluation societies, associations or networks so that they can better contribute to good governance and strengthen civil society. It aims to build evaluation capacity; develop evaluation principles and procedures; encourage the development of new evaluation societies and associations or networks; undertake educational activities to increase public awareness of evaluation; and, seek to secure resources for co-operative activity.

IDEAS³ held its first conference in India in 2005. IDEAS was created to attract individual members from all over the world (particularly from developing countries and transition economies), who will:

- a) promote development evaluation for results, transparency and accountability in public policy and expenditure;
- b) give priority to evaluation capacity development;
- c) foster the highest intellectual and professional standards in development evaluation; and
- d) encourage national and regional development evaluation groups.

The national, regional and global professional organizations for evaluation can foster democratic approaches to evaluation. They do this not only by helping to share experience and expertise, but also by providing a forum for greater dialogue among civil society, academia, governments and donors, in line with the increasingly important role of civil society, academia and the private sector in national development. Quesnel (2006) suggests the importance of strategies in promoting partnerships with the mass media and parliaments, to further the use of evaluation as instrument for transparency and accountability.

2 See <http://internationalevaluation.com>

3 See <http://www.ideas-int.org/>

Harmonization

Paris Declaration Commitment

Donors implement, where feasible, common arrangements for monitoring and evaluation

Implications to the Evaluation Function

- **OECD/DAC Working Group on Evaluation**
- **Evaluation Cooperation Group**
- **UN Evaluation Group**
- **UN Development Assistance Framework and Integrated monitoring and evaluation Plan**
- **Joint Evaluations**

Developing countries demand more and more coordination and harmonization among donor countries and international agencies. This trend is visible in the evaluation arena too.

Quesnel (2006) mentions that the Expert Group on Evaluation of the Development Assistance Committee (DAC) of the Organization for Economic Cooperation and Development (OECD) was the main body which introduced greater harmonization in the evaluation of official development assistance. It provides a forum for evaluation specialists, from 30 governmental and multilateral agencies, to work together to improve the relevance and effectiveness of development cooperation.

The system of international financial institutions is another potent lever used by governments for greater systematization of the use of evaluation. These institutions, such as the International Monetary Fund; the World Bank Group⁴; regional and sub-regional multilateral development banks, or international funds; are governed by assemblies of government representatives. Each organization has an evaluation unit. The Evaluation Cooperation Group (ECG)⁵ brings together the heads of evaluation of the global and regional organizations. They have done much to harmonize and develop new evaluation approaches in response to evolving development policy challenges.

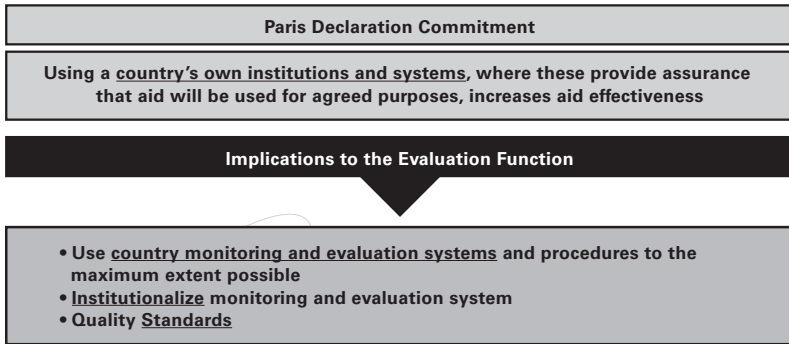
4 See <http://www.worldbank.org/oed/>

5 See <http://www.ecgnet.org/>

The United Nations Evaluation Group (UNEG)⁶ brings together some 38 heads of evaluation. UNEG aims to contribute to harmonization across the UN System by improving the UN use of evaluation through simplification and the undertaking of joint evaluation work, especially at the country level.

At country level, more and more evaluations are carried out jointly to ensure harmonization and synergy among donors. In the case of the United Nations, several UN specialized agencies developed the UN Development Assistance Framework (UNDAF), to ensure that results are delivered in a coordinated manner and are aligned to national development priorities. The UNDAF document also includes a common monitoring and evaluation plan.

Alignment



International development agencies should focus efforts on supporting existing national monitoring and evaluation systems, aligning their monitoring and evaluation assistance with national monitoring and evaluation plans and priorities.

Strategic contributions by international development agencies include supporting sustainable national monitoring and evaluation capacity development, taking into consideration the value of diversity in evaluation approaches and always focusing on the quality of the knowledge produced by evaluative processes.

Respect for evaluation standards should be a priority in order to ensure the quality of knowledge produced through evaluative processes.

6 See <http://www.uneval.org/uneval>

Ownership

Paris Declaration Commitment

- **Partner countries exercise leadership in developing and implementing their national development strategies**
- **Donors respect partner country leadership and help strengthen their capacity to exercise it**

Implications to the Evaluation Function

- **Country-led Evaluations and Systems**
- **Evaluation capacity development**

Integrated monitoring and evaluation systems should be nationally owned. Country-Led Evaluations in South Africa and in Bosnia & Herzegovina were the first attempts to translate this principle into reality. In 2007, UNICEF in Bosnia & Herzegovina supported the Government in carrying out a Country-Led Evaluation of the Social Protection Chapter of the local PRS, which focuses on good governance for children. The Government led the evaluation process, notably by identifying the scope of the evaluation; the evaluation questions which responded to the national information needs (to inform the design of the next cycle of the PRS); and, by involving all major stakeholders (including civil society organizations and international development organizations, such as the World Bank and, the European Commission). UNICEF thus aligned itself to the Government's monitoring and evaluation needs, by providing the required technical assistance and strengthening national evaluation capacities.

IDEAS is also working to strengthen country-led evaluations and systems (CLES). With this aim, IDEAS has organized regional workshops on CLES in Central and Eastern Europe, and Africa.

The strategic intent of monitoring and evaluation⁷

Decision-makers are looking to monitoring and evaluation as the strategic function to turn the Paris Declaration's key principles into reality. Monitoring and evaluation can provide unique information about the performance of government policies, programmes and

7 Based on WB, 2007

projects. It can identify what works, what does not work, and the reasons why. Monitoring and evaluation also provides information about the performance of a government, of individual ministries and of agencies, managers and their staff. Information on the performance of donors supporting the work of governments is also provided.

It is tempting, but dangerous, to view monitoring and evaluation as having inherent value. The value of monitoring and evaluation comes not from conducting monitoring and evaluation or from having such information available; rather, the value comes from using it to help improve government performance. There are several ways in which monitoring and evaluation information can be highly useful to governments and to others:

- **To enhance results-based management**, by supporting governments in managing public policies and programmes, including government service delivery and the management of staff.
- **To enhance transparency and support accountability relationships**. These include the accountability of government to the parliament or congress, to civil society, and to the donors. Monitoring and evaluation also supports the accountability relationships within government, such as between sector ministries and central ministries, and between ministers, managers and staff. Monitoring and evaluation provide a vehicle to magnify the voice of civil society and to put additional pressure on government to achieve higher levels of performance.

Civil society can play an important role in monitoring and evaluation in at least four ways. Firstly, it can present the views of beneficiaries on government service delivery. Secondly, it can produce analysis and reviews of government performance, from activities such as budget analyses and citizen report-cards. Thirdly, it provides independent scrutiny of monitoring and evaluation findings provided by governments. Finally, civil society is a user of monitoring and evaluation information through media reporting and also the activities of universities, think-tanks and NGOs.

- **To support evidence-based policy making**, particularly in the context of the national budget cycle and of national planning. These processes focus on identifying government priorities from the competing demands of citizens and groups in society.

Monitoring and evaluation information can support government's deliberations by providing evidence about the most cost-effective types of government policies.

Since evidence-based policy making is at the heart of the new aid environment, the next chapter focuses on analyzing the role and strategic intent of monitoring and evaluation in evidence-based policy making.

Evidence-based policy making

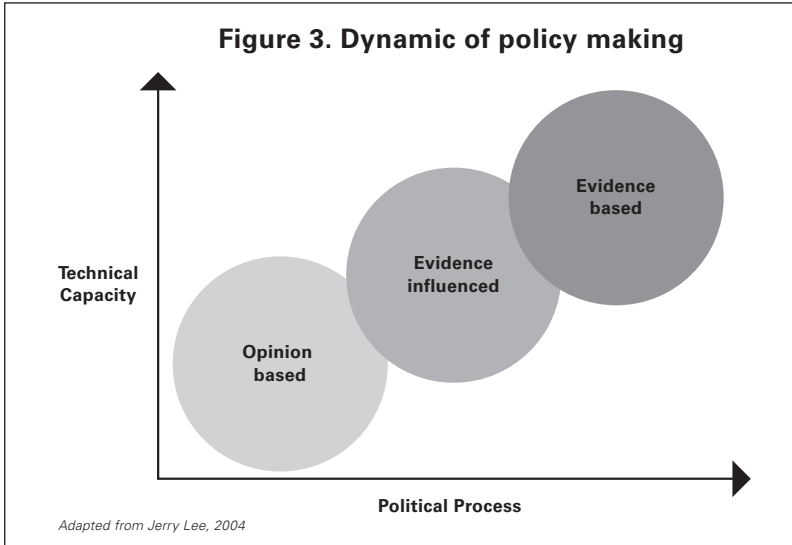
What is evidence-based policy?

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 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". Evidence may include information produced by integrated monitoring and evaluation systems, academic research, historical experience and "good practice" information.

This approach stands in contrast to **opinion-based policy**, which relies heavily on either the selective use of evidence (e.g. on single studies 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**". As we will see later, this is mainly due to the fact that the policy making process is inherently political and, that the processes through which evidence translates into policy options often fails to meet required quality standards.

Proponents of evidence-based policy and practice acknowledge that not all sources of evidence are sufficiently sound to form the basis of policy making (Davies, Nutley and Smith, 2000). Much research and evaluation is flawed by unclear objectives; poor design; methodological weaknesses; inadequate statistical reporting and analysis; selective use of data; and, conclusions which are not supported by the data provided (Davies, 2003).



The concept of 'evidence-based policy' has been gaining currency over the last two decades. Literature suggests that this new interest, in bringing impartial evidence to the policy making process, comes in response to a perception that government needs to improve the quality of decision-making. Poor quality decision-making has been related to the loss of public confidence suffered in recent years. Traditionally, politicians and policy makers operated based on the belief that their voters were unquestioning. However, citizens are less and less inclined to take policy views on trust. Policy-makers are increasingly asked to explain not just what policy options they propose, and why they consider them appropriate, but also their understanding of their likely effectiveness.

Box 2: Modernising policy making

The UK government's vision of modernised policy making was set out in ***Professional Policy making*** (SPMT, 1999). Nine core features were identified:

- Forward looking: takes a long term view of the likely impact of policy
- Outward looking: takes account of influencing factors and learns from elsewhere
- Innovative and creative: questions the status quo and is open to new ideas
- Evidence based: uses the best available evidence from a wide range of sources
- Inclusive: is fair and takes account of the interests of all

- Joined up: works across institutional boundaries and considers implementation
- Reviews: keeps policy under review
- Evaluates: builds evaluation into the policy process
- Learns lessons: learns from experience of what works and what does not

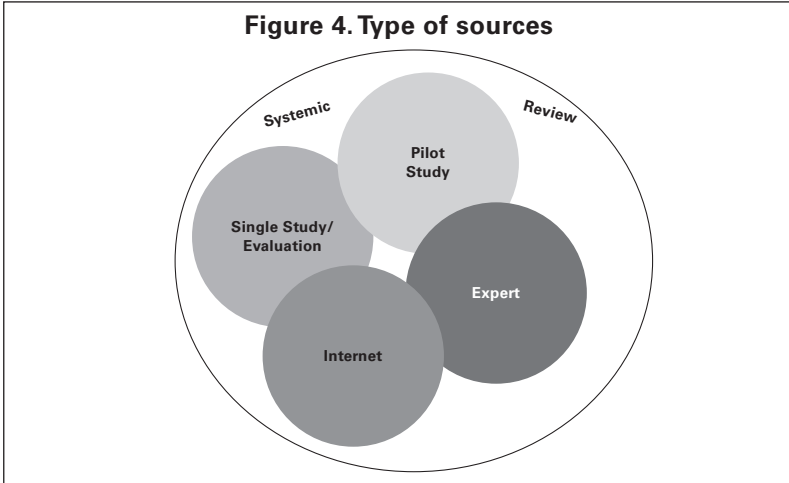
Source: NUTLEY, S., DAVIES, H. and WALTER I., 2002.

The nature of evidence

“The driving force for evidence in government should be the type of question being asked, rather than any particular research method or design.” Jerry Kee

If we are indeed interested in developing an agenda in which evidence is to be more influential, then first we need to develop some agreement as to what constitutes evidence and, in what context, to address different types of policy/practice questions. This means being more explicit about the role of research and evaluation vis-à-vis other sources of information. In addition, greater attention is needed on the relative strengths and weaknesses of different methodological stances. Such methodological development needs to emphasise a ‘horses for courses’ approach. That is, identifying which policy and practice questions are amenable to analysis and through what kind of specific research and evaluation technique. It also needs to emphasise methodological pluralism, rather than continuing paradigmatic antagonisms, seeking complementary contributions from different research and evaluation designs and sources rather than epistemological competition.

The disputes between researchers and evaluators about the superiority or inferiority of quantitative versus qualitative studies, or experimental versus experiential research designs, are not productive. They can lead to poor evidence, or to evidence that is technically very good but of little use to policy makers or anyone else. In the meantime, there are many other policy forces, from lobbyists to pressure groups, which are less thorough but more readily available to policy makers. It is not surprising that such forces are often more successful in finding their way into policy making.



The figure above presents the different types of sources of Evidence:

Systematic reviews

The evidence-based policy movement has built its claim, to influencing policy and practice, on the basis of using research and evaluation evidence which has been *systematically* searched, critically appraised, and rigorously analysed, according to explicit and transparent criteria. Systematic reviews and meta-analyses of existing evidence are accorded such a high premium amongst proponents of evidence-based policy and practice because they overcome the shortcomings of single studies (Cooper and Hedges, 1994; Davies, 2003). Single studies and evaluations can provide an unbalanced and unrepresentative view of the total available evidence on a topic or policy issue. This is because they are almost always sample-specific, time-specific, and context-specific. Also, some single studies and evaluations lack methodological rigour or are not undertaken to high quality standards. Such studies should not be included in the evidence base for policy making or practice. Systematic reviews, by contrast, use explicit and transparent quality criteria, and rigorous standards for searching and critical appraisal, in order to establish 'the consistencies and account for the variability of similar appearing studies' (Cooper and Hedges, 1994:4). Systematic reviews are able to provide generalisations, and specify the limits of these generalisations, amongst existing research evidence, by accumulating only the sound evidence, and identifying studies which are sample, time, or context specific.

Single studies and evaluations

Single studies and evaluations are more commonly used to support government policy and practice than are systematic reviews. Indeed, the vast majority of the research and evaluation undertaken by, or on behalf of, Governments and the Development community consists of single studies or evaluations. These are often undertaken without any accumulation of existing evidence through systematic review methodology. If undertaken to the highest possible standards, single studies and evaluations can provide valuable and focussed evidence for particular policies, programmes and projects in specific contexts. Unlike systematic reviews, however, single studies are less able to say much about the variability of populations, contexts and conditions under which policies may or may not work.

Pilot studies and case studies

Pilot studies and case studies provide the other sources of evidence for policy making and policy implementation. The UK Cabinet Office recommended that “the full-scale introduction of new policies and delivery mechanisms should, wherever possible, be preceded by closely monitored pilots” (Cabinet Office, UK, 2003).

It is sometimes argued that the tight timetables and schedules of the policy making process make it impossible for systematic reviews, single empirical studies, pilots or case studies to be undertaken before rolling out a policy, programme or project. This reasoning is often deployed to justify the use of whatever evidence is readily available, regardless of its scientific quality or source. Such urgency and rapidity of action may be understandable, especially in the absence of a well established evidence base for many areas of public policy. However, it is short sighted and possibly counter productive. Evidence that is selective, and not subjected to careful, critical appraisal and risk assessment, can often lead to inappropriate courses of action which cause more harm than that which they are intended to prevent.

Experts’ evidence

Expert opinion is also commonly used to support government policy and practice, either in the form of expert advisory groups or special advisers. However, using experts as a basis for policy making and practice again raises the problems of selectivity of knowledge and expertise. There is also the need to ensure that the expertise being provided is up to date and well grounded in the most recent research evidence.

Internet evidence

The internet age has brought a revolution in the availability of information and knowledge. Most, though not all, government departments have desktop access to the internet. It is anticipated that more and more government departments will have internet access within the near future. This means that there is uneven access across government departments to these important sources of potential evidence

Not all of the information available *via* the internet, however, is of equal value or quality. Many sites provide 'evidence' that is either scientifically or politically biased, or both. The uncertain scientific and political basis of much of the information and knowledge on the internet makes it difficult to be sure it meets the required quality and to determine if it constitutes sound, valid and reliable evidence.

An optimistic scenario for the future is that initiatives which encourage consultation, through devices such as policy action teams and service planning fora, will widen the membership of policy and practice communities. The involvement of wider interests in these teams is likely to set a different agenda and lead to a more practice-based view of policy and delivery options. The use of research and other evidence under such a scenario is likely to be wide ranging.

Box 3: Types of research/evaluation utilisation

1. Instrumental use

Research feeds directly into decision-making for policy and practice.

2. Conceptual use

Even if policy makers or practitioners are blocked from using findings, research and evaluation can change their understanding of a situation, provide new ways of thinking and offer insights into the strengths and weaknesses of particular courses of action. New conceptual understandings can then sometimes be used in instrumental ways.

3. Mobilisation of support

Here, research and evaluation becomes an instrument of persuasion. Findings – or simply the act of research – can be used as a political tool and can legitimate particular courses of action or inaction.

4. *Wider influence*

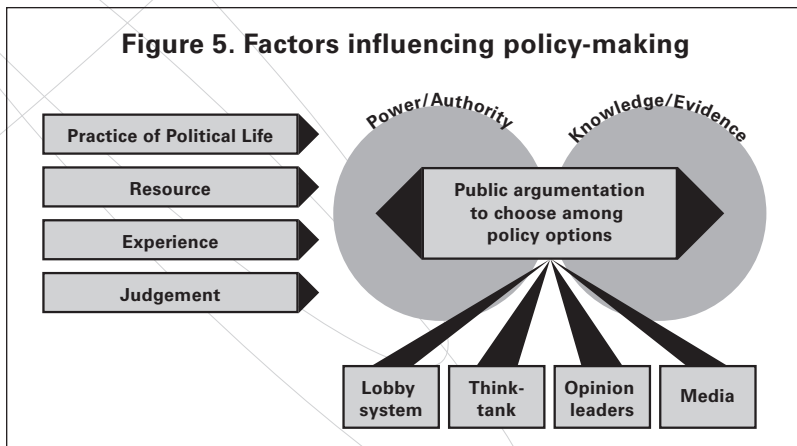
Research and evaluation can have an influence beyond the institutions and events being studied. Evidence may be synthesised. It might come into currency through networks of practitioners and researchers, and alter policy paradigms or belief communities. This kind of influence is both rare and hard to achieve, but research adds to the accumulation of knowledge that ultimately contributes to large-scale shifts in thinking, and sometimes action.

Source: adapted by NUTLEY, S., DAVIES, H. and WALTER I., 2002.

Knowledge as power? The need for evidence-based policy options

As mentioned earlier, the policy making process is political. 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 then to be a tension between power and knowledge in the shaping of policy. A similar tension exists between authority and expertise in the world of practice.

Emphasising the role of power and authority at the expense of knowledge and expertise in public affairs seems cynical; emphasising 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 part of the armoury of those engaged in the discourse. If this seems too crude a metaphor, bear in mind that, to be effective, weapons must be handled with care, their use confined to skilled personnel aware of their limits and capabilities. They must not be deployed casually or wastefully, and must always be used with as full a regard to the risks for those who use them as to those against whom they are used. Knowledge is open to misuse just as much as other sources of power.

The lobby system and pressure groups are other factors competing with evidence to influence policy making and policy implementation. Think-tanks, opinion leaders and the media are also major influences. The ways in which these groups work to influence policy can be under-estimated and misunderstood by proponents of evidence-based policy and practice. It is not that these groups fail to use evidence to promote particular policies, programmes or projects. Rather, it is that such evidence is often less systematic, and more selective, than that used by supporters of evidence-based policy and practice.

Once we acknowledge that evidence is used in various ways by different people in the policy process, governments do not appear to be the 'evidence-free zone' that is sometimes depicted. The evidence that is used is wide-ranging. Policy makers need information, not only about the effectiveness of a procedure and the relationship between the risks and the benefits, but also about its acceptability to key constituencies. They use information in the way they do because the central challenge is not just to exercise sound technical judgement, but to develop consensus between all the interests and institutions of society, and between the interests and institutions represented in the policy making process (Perri, 2002). The quest for evidence-based policy should not, it is argued, be driven by a desire to reduce policy making to technical analysis. Accommodating divergence rather than imposing convergence appears to be the key to a well functioning democratic polity.

Thus, evidence-base 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.

Other factors which influence policy making and policy implementation are the sheer pragmatics of political life such as parliamentary terms and timetables, the procedures of the policy making process, the capacities of institutions, and unanticipated contingencies which may arise. These factors need not be the enemy of evidence-based policy and practice. First, evidence-based policy is a strategic as well as an operational activity. Part of its role is to build an evidence-base for future generations of policy-makers and practitioners. Second, evidence-based policy and practice should be the first line of response to unanticipated events in the sense of identifying what is already known about the problem and what is not.

Policy making and policy implementation take place within the context of finite (and sometimes declining) resources. This means that policy making is not just a matter of 'what works', but what works at what cost and with what outcomes (both positive and negative). This requires sound evidence not only of the cost of policies, programmes or projects, but also the cost-effectiveness, cost-benefit, and cost-utility of different courses of action.

Last but not least, another important factor that clearly influence policy and practice is the experience, expertise and judgement of decision makers. These factors often constitute valuable human and intellectual capital and include the tacit knowledge that has been identified as an important element of policy making (Nutley, Walter and Davies, 2003). Such influences may, or may not be informed by sound evidence. Indeed, judgement based on experience and expertise may be of critical significance in those situations where the existing evidence is equivocal, imperfect, or non-existent (Grimshaw, *et al*, 2003). Consequently, a major goal of evidence-based policy is to ensure that policy making *integrates* the experience, expertise and judgement of decision makers with the best available external evidence from systematic research.

In conclusion, evidence for policy has three components:

- first is the hard data (research, evaluations, etc);
- second is the analytical argumentation that puts the hard data into a wider context;
- third is an evidence base comprising stakeholder opinion.

Box 4: Evidence into policy

Attention is more likely to be paid to research findings when:

- The research and evaluation is timely, the evidence is clear and relevant, and the methodology is relatively uncontested.
- The results support existing ideologies, are convenient and uncontentious to the powerful.
- Policy makers believe in evidence as an important counterbalance to expert opinion, and act accordingly.
- The research and evaluation findings have strong advocates.
- Research and evaluation users are partners in the generation of evidence.
- The results are robust in implementation.
- Implementation is reversible if need be.

Source: adapted by NUTLEY, S., DAVIES, H. and WALTER I., 2002.

Evidence into practice: increasing the uptake of evidence in both policy and practice

A stronger commitment to make research and evaluation 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. The primary concern for those wishing to improve the utilisation of research and evaluation is how to tackle the problem of under-use, where findings about effectiveness are either not applied, or are not applied successfully. However, concerns have also been raised about overuse, such as the rapid spread of tentative findings, and about misuse, especially where evidence of effectiveness is ambiguous (Walshe and Rundall, 2001).

A strategic approach to knowledge creation: the Integrated monitoring and evaluation Strategy (IMES)

Whichever part of the public sector one is concerned with, one observation is clear: the current state of research and evaluation based knowledge is insufficient to inform many areas of policy and practice. There remain large gaps and ambiguities in the knowledge base, and the research literature is dominated by small, ad hoc stud-

ies, often diverse in approach and of dubious methodological quality. In consequence, there is little accumulation of a robust knowledge base on which policy-makers and practitioners can draw.

Furthermore, additions to the research literature are more usually research producer-driven than led by research users' needs. Recognition of these problems has led to attempts to develop integrated monitoring and evaluation strategy (IMES) and plans to address these problems. Developing such strategies and plans necessarily requires a number of key issues to be addressed. These are:

- What research and evaluation designs are appropriate for specific research questions, and what are the methodological characteristics of robust research?
- What is an appropriate balance between new primary research and the exploitation of existing research through secondary analysis?
- How can the need for rigour be balanced with the need for timely findings of practical relevance?
- What approaches can be used to identify gaps in current knowledge provision, and how should such gaps be prioritised?
- How should research and evaluation be commissioned (and subsequently managed) to fill identified gaps in knowledge?
- How can research and evaluation capacity be developed to allow a rapid increase in the availability of research based information?
- How are the tensions to be managed between the desirability of 'independent' researchers and evaluators free from the more overt forms of political contamination, and the need for close co-operation (bordering on dependence) between research users and research providers?
- How should research and evaluation findings be communicated and, more importantly, how can research and evaluation users be engaged with the research and evaluation production process to ensure more ready application of its findings?

Stakeholder involvement in the creation of wide-ranging integrated monitoring and evaluation strategies is crucial, and such strategies need to address capacity building as well as priority areas for future research and evaluation.

Getting appropriate 'buy-in'

The literature suggests that getting policy-makers and practitioners to own and use evidence also involves getting commitment and buy-in at the most appropriate levels. In central government this usually means getting ministers and senior policy officials to sign up to the ownership of a project *and* the evidence that goes to support it. This in turn means a commitment to use findings which are contrary to expectations, and not to continue with a policy, programme or project if the available research evidence indicates that this is ineffective. At the level of 'front line' service delivery it means getting key decision-makers to 'own' and champion the evidence that supports good practice (Davies, 1999b, 2004). This is most likely to take place, and most likely to be effective, in organizational structures which are non-hierarchic, open and democratic (Dowd, 1994; Martin, 1997).

The need to improve the dialogue between policy-makers and the research/evaluation community

A closely related issue is getting policy-makers and practitioners to own the evidence needed for effective support and implementation of policy. This is in contrast to the position where evidence is solely the property and domain of researchers and evaluation 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, improvement of the dialogue between policy-makers and the research and evaluation community is paramount. It is sensible that such dialogues should not be constrained by one single policy issue or a single research project. This raises questions about what the nature of ongoing relationship between policy-makers and external researchers and evaluators should be. Using the analogy of personal relationships, it has been suggested that promiscuity, monogamy and bigamy should all be avoided. Instead, polygamy is recommended, where policy makers consciously and openly build stable relationships with a number of partners, each of whom offers something different, know of each other and can understand and respect the need to spread oneself around (Solesbury, (2001).

Overall, a striking feature of the existing literature on ways of improving uptake of evidence, in both policy and practice, is the common conclusion that developing better, ongoing interaction between evidence providers and evidence users is the way forward (Nutley et al, 2002). The traditional separation between the policy arena, practitioner communities and the research and evaluation community has largely proven unhelpful. Much of the more recent thinking in this area now emphasises the need for partnerships if common ground is to be found (Laycock, 2000; Nutley et al, 2000).

This echoes Huberman's (1987) call for 'sustained interactivity' between researchers and practitioners throughout the process of research, from the definition of the problem to the application of findings. Closer and more integrated working over prolonged periods would seem to be capable of fostering cross-boundary understanding. Doing so, however, is not cheap or organizationally straightforward, and it raises some serious concerns about independence and impartiality.

The vision should be of a society where analysts and experts are 'on tap but not on top' – a society, which is active in its self-critical use of knowledge and social science (Etzioni, 1968, 1993). In such a vision research evidence may well be used as a political weapon but 'when research is available to all participants in the policy process, research as political ammunition can be a worthy model of utilisation' (Weiss, 1998). Of course, a problem arises when certain groups in society do not have access to research and other evidence, or if their ability to use this evidence is restricted because of their exclusion from the networks that shape policy decisions

Matching strong demand with a good supply of appropriate evidence

A distinction can be made between people who are users of research and evaluation and those who are doers of research and evaluation. Whilst it may be unrealistic for professional decision-makers and practitioners to be competent doers of research and evaluation, it is both reasonable and necessary for such people to be able to understand and use research and evaluation in their professional practice. Integrating research and evaluation 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 social, economic and policy research and evaluation which are available; how to gain access to them; and, how to critically appraise them. Without such knowledge and understanding it is

difficult to see how a strong *demand* for research and evaluation can be established and, hence, how to enhance the practical application of research and evaluation. 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.

Improving “understandability” of evidence

A further challenge for researchers and evaluators is making the findings of research and evaluation accessible to the policy-making community. Too often research and evaluation is presented in an unclear way with as much, if not more, emphasis given to the caveats and qualifications of research findings (the ‘noise’ of social research) than to the message and implications of these findings for policy and practice (the ‘signal’). Researchers and evaluators often need to ‘translate’ social science evidence into a language that is useful to the users of evidence, without distorting or misrepresenting the research evidence.

Effective dissemination and wide access

Whether the focus is on primary research or on the systematic review of existing studies, a key issue is how to communicate findings to those who need to know. The strategies used to get research and evaluation findings 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 research and evaluation users can tap into). For example, some UN and development agencies make their own evaluation database available on the Internet.

Much effort has gone into improving the dissemination process, and good practice guidance abounds (see Box 5). This has developed our appreciation of the fact that dissemination is not a single or simple process. Different messages may be required for different audiences at different times. It appears that the promulgation of individual research findings may be less appropriate than distilling and sharing pre-digested research summaries. Evidence to date also suggests that multiple channels of communication (horizontal and vertical networks and hierarchies), may need to be developed in parallel (Nutley and Davies, 2000).

Box 5: Improving dissemination

<i>Recommendations for research/evaluation commissioners</i>	<i>Recommendations for researchers/evaluators</i>
<ul style="list-style-type: none"> • Time research/evaluation to deliver solutions at the right time to specific questions facing practitioners and policy-makers • Ensure relevance to current policy agenda • Allocate dedicated dissemination and development resources within research/evaluation funding • Include a clear dissemination strategy at the outset • Involve professional researchers/evaluators in the commissioning process • Involve service users in the research/evaluation process • Commission research reviews to synthesise and evaluate research 	<ul style="list-style-type: none"> • Provide accessible summaries of research • Keep the research/evaluation report brief and concise • Publish in journals or publications that are user friendly • Use language and styles of presentation that engage interest • Target material to the needs of the audience • Extract the policy and practice implications of research/evaluation • Tailor dissemination events to the target audience and evaluate them • Use a combination of dissemination methods • Use the media • Be proactive and contact relevant policy and delivery agencies • Understand the external factors likely to affect the uptake of research

Source: adapted by NUTLEY, S., DAVIES, H. and WALTER I., 2002.

Incentives to use evidence

Recent government reports aimed at improving the process by which policy is made set out a number of recommendations for increasing the use of evidence (see Box 6).

Practitioners need incentives to use evidence and to do things that have 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 evidence use, such as integrating analytical staff at all stages of the policy development process.

Box 6: Encouraging better use of evidence in policy making

<i>Increasing the pull for evidence</i>	<i>Facilitating better evidence use</i>
<ul style="list-style-type: none"> • Require the publication of the evidence base for policy decisions • Require departmental spending bids to provide a supporting evidence base • Submit government analysis (such as forecasting models) to external expert scrutiny • Provide open access to information – leading to more informed citizens and pressure groups. 	<ul style="list-style-type: none"> • Encourage better collaboration across internal analytical services (e.g. researchers, statisticians and economists) • Co-locate policy makers and internal analysts • Integrate analytical staff at all stages of the policy development process • Link R&D strategies to departmental business plans • Cast external researchers more as partners than as contractors • Second more university staff into government • Train staff in evidence use

Source: NUTLEY, S., DAVIES, H. and WALTER I., 2002.

Conclusion

The consensus on how to improve effectiveness of aid, reached by hundreds of leaders of governments and civil society organizations from both developing and developed countries, is an historical milestone.

The international agreement on the five key principles, to ensure results in winning the fight against poverty, is a value added for all stakeholders. Monitoring and evaluation is expected to play a strategic role in ensuring such principles are translated into reality. Monitoring and evaluation can do this by providing the evidence needed to take informed policy decisions. In this way, monitoring and evaluation plays an essential role in keeping the promise to improve the life of millions of people around the world.

For the first time, monitoring and evaluation has a clear strategic intent that goes well beyond measuring and tracking, accountability and reporting. The international evaluation community has a clear responsibility to deliver.

Let us keep the promise!

References

- Barnardo's R&D Team. (2000), *What works. Making connections: linking research and practice*, Ilford, p.107.
Available at: < http://www.barnardos.org.uk/shop/acatalog/Catalogue_Other_125.html >
Summary at: < <http://www.jrf.org.uk/knowledge/findings/socialcare/910.asp> >
- BOAZ, A., ASHBY, D. and YOUNG, K. (2002), *Systematic reviews: what have they got to offer evidence based policy and practice?* UK, p.28.
Available at: < <http://www.evidencenetwork.org/Documents/wp2.pdf> >
- BULLOCK, H., MOUNTFORD, J. and STANLEY, R. (2001), *Better policy making*, UK, p.83.
Available at: < http://www.cmps.gov.uk/better_policy_making.pdf
- Cabinet Office, (1990), *Modernising Government*, White Paper, UK.
Available at: < <http://www.cabinet-office.gov.uk/moderngov/whtpaper/index.htm> >
- Cabinet Office, (1990), *Professional Policy making for the Twenty-First Century*, UK.
- Cabinet Office, (2001), *Modernising government*, (Ch. 2, para. 6. Cm 4310).
In: SOLESBURY, W. *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.
Available at: <<http://www.cabinet-office.gov.uk/moderngov/whtpaper/index.htm>>.
- Cabinet Office Performance and Innovation Unit, (2001), *Better policy delivery and design: a discussion paper*, UK, p.21.
Available at: <<http://www.cabinet-office.gov.uk/innovation/whatsnew/betterpolicy.shtml>>.
- DAVIES, 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?*
- DAVIES, P. (2004), PhD, *Is Evidence-Based Government Possible?* Jerry Lee Lecture.
- Department for Environment, Food and Rural Affairs of the UK. (2006), *Evidence Based Policy- Making*.
Available at: < <http://www.defra.gov.uk/science/how/evidence.htm> >.
- European Commission. (2001), *Governance in the EU: a White Paper European Commission: Brussels*. COM, Luxembourg.
Available, with supporting material, via: <http://europa.eu.int/comm/governance/white_paper/index.htm>

GREENE, J. C., BENJAMIN, L. and GOODYEAR, L. (2001), *The Merits of Mixed Methods in Evaluation*. In: Evaluation 2001, pp.25-44.

Independent Evaluation Group (IEG). (2006) *Priority for Government Monitoring and Evaluation Systems*.

Available at: < http://www.worldbank.org/ieg/eecd/priority_for_government_me.html >

Institute of Development Studies (IDS). (2005), *IDS Policy Briefing*, Issue 25, 2005, IDS, UK.

MAJONE, G. (1989), *Evidence, argument and persuasion in the policy process*, USA, pp.190.

National Audit Office (NAO). (2001), *Modern policy making: ensuring policies deliver value for money*, UK. pp.60.

Available at: < http://www.nao.gov.uk/publications/nao_reports/01-02/0102289.pdf >

NHS Centre for Reviews and Dissemination (NHSCRD). (1999), *Getting evidence into practice*. In: *Effective Health Care Bulletin* Sep 5(1), 1999, UK. p.16.

Available at: < <http://www.york.ac.uk/inst/crd/ehc51warn.htm> >

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. 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., 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> >

OECD, (2005), *Aid Effectiveness: Three Good Reasons Why the Paris Declaration Will Make a Difference*. In: 2005 Development Co-Operation Report.

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> >

Performance and Innovation Unit (PIU). (2001), *Better policy delivery and design: a discussion paper*, UK, pp.21.

Available at: < <http://www.cabinetoffice.gov.uk/innovation/papers/betterpolicy.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 >

QUESNEL, J. (2006), *The importance of Evaluation networks and associations*.

In: IOCE (International Organization for Cooperation in Evaluation), *Creating and Developing Evaluation Organizations –Lessons learned from Africa, Americas, Asia, Australasia and Europe*, Peru.

- 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. 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, Evaluation Working Papers ISSUE #5: *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.
- UNDP (2005), *How to guide MDG-based national development strategies*. Available at: < http://mdg-guide.undp.org/?page=glossary_4 >
- 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. (1998), *Evaluation*, Second Edition, USA.
- WEISS, C. H. (1998), *Have we learned anything new about the use of evaluation?* In: American Journal of Evaluation 19(1), 1998, pp.21-33.
- World Bank, (2007), *Evidence-based policy making: challenging interest-group based policy*. Available at: <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALDEVELOPMENT/EXTTOPPSISOU/0,,contentMDK:20587547~menuPK:1441861~pagePK:64168445~piPK:64168309~theSitePK:1424003,00.html>
- World Bank. (2007), *Operation Evaluation Department website*.
- World Bank Operations Evaluation Department. (1998), *Evaluation Capacity Development, Public Sector Performance –The Critical Role of Evaluation*, Selected Proceedings from a World Bank Seminar.

THE RELATIONSHIP BETWEEN EVALUATION AND POLITICS¹

*By Ove Karlsson Vestman, Director,
Mälardalen Evaluation Academy,
and Vice President of the Swedish Evaluation Society;
and Ross F. Conner, University of California,
former President of the American
Evaluation Association and current President of IOCE*

Evaluation is a young discipline that according to Pawson and Tilly (1997) has passed its adolescence. If evaluation is now in its adulthood, it is reasonable to consider whom evaluation should have as its “life partner” or “partners.” The evaluation family traditionally has included good researchers with their ideal of neutral, objective research as the prototype for evaluation and has recognized these partners in the evaluation enterprise with awards and high status. Evaluation work, however, is always couched within a political context, and this reality brings different kinds of partners into the relationship. These partners, including politicians and policy-makers, often make the evaluation family uneasy. There has been a basic conception that evaluation (and similarly research) becomes adulterated when it mixes with politics. Generally the discussion is permeated by a negative view of politics; it conjures up images of trouble, disruption and even violence on the one hand, and deceit, manipulation and lies on the other. It is less common to see a positive or at least neutral view of politics as an important and inevitable part of human life and interaction.

If politics and evaluation are destined to be “life partners” in the adulthood of evaluation, then what forms could the relationship take – marriage, cohabitation or living apart? This chapter will consider some of these possibilities.

1 Reprinted with permission from I. Shaw, J.C. Greene and M. M. Mark (2006). The Sage Handbook of Evaluation. London: Sage Publications Ltd. Additional details about the handbook are available at www.sagepub.com

Definitions of evaluation and politics

“Evaluation” refers to the process of determining the merit, worth, or value of something (Scriven, 1991). The evaluation process involves identifying relevant values or standards that apply to what is being evaluated, performing empirical investigation using techniques from the social sciences, and then integrating conclusions with the standards into an overall evaluation or set of evaluations. The first step in the process, the identification of relevant standards and values to apply to what is being evaluated, has to do with what partners involved in the evaluation see as relevant in the particular case. Should the priority be, for example, on economical, educational, social, ethnic or democratic standards and values? Making these choices is an exercise of power that connects evaluation to politics. That is an interpretation in line with Hammersley (1995) who says that politics has to do with the use of power and that it also concerns making value judgments and taking actions on the basis of them.

According to Caro (1977), evaluation must fulfill two purposes – information and judgment. The former fits well with the research community’s traditional epistemological perspective, whereas making judgments does not. Social research’s aim, traditionally and in a narrow sense, is limited exclusively to producing knowledge but not to producing value judgments or evaluative conclusions. There has also been considerable debate about which models should be adopted for making judgments. One strategy is to treat judgments as technical measurements, in order to avoid involving values with their attendant political implications. It is precisely at this juncture, however, where evaluation and politics are related. Both are concerned with values, value judgments, and value conflicts in public life. The reality is that evaluation, in order to fulfill its second purpose of making judgments, cannot avoid the issue of politics.

Politics – a contested concept

Politics has been defined in many ways. One could say that politics is regarded as an “essentially contested” concept (Gallie, 1956) in that there are controversies about the term so deep that no neutral or settled definition can ever be developed. In effect, a single term (like “politics” or “evaluation,” for that matter) can represent a number of rival concepts, none of which can be accepted as its “true” meaning. For example, it is equally legitimate to define politics as what concerns the state, as the conduct of public life,

as debate and conciliation, and as the distribution of power and resources. On the basis of Lasswell (1936), politics is about who gets what, when and how. The “when” and “how” aspect of politics is put forward by Heywood (2002) who sees politics in its broadest sense as “the activity through which people make, preserve and amend the general rules under which they live” (p 4). The activities are formed into institutions in Dahl’s (1984) definition of politics as: “any persistent pattern of human relationships that involves, to a significant extent, control, influence, power or authority.” (p. 9-10). Politics, however, is not just activities (decisions on allocation of resources, organization of institutions, etc). Easton (1968) argues that politics is the authoritative allocation of values for a society, and that politics essentially is making moral decisions about what is good and what is bad. This definition places politics close to the definition of evaluation that emphasizes evaluation as the production of information together with the production of judgment.

From a narrow to a broad definition

Heywood (2002) presents some illustrative views of politics that can be taken as a point of departure to elaborate the picture of politics. In the narrowest sense, politics can be treated as the equivalent of party politics. Here, politics is restricted to those state actors who are consciously motivated by ideological beliefs and who seek to advance them through membership of a formal organization such as a political party. This view can be expanded to see politics as the art of government. Here, politics is what takes place within a system of social organization centered upon the machinery of government. More broadly, politics can be associated with formal or authoritative decisions that establish a plan of action for the community. This means that most people, most institutions and most social activities can be regarded as being “outside” politics and the policy cycle through which politics and governance takes its form. The politicians are described as “political”, whereas civil servants are seen as “non-political”, as long as they act in a neutral and professional fashion. Similarly, evaluators are taken to be “non-political” figures when they interpret and value the evaluand (a program or a policy, for example) impartially and in accordance with the collected information. Evaluators may be accused of being political, however, if personal preferences or some other form of bias influences their judgments.

According to Heywood this definition can be broadening by taking politics beyond the narrow realm of government and viewing politics

as public affairs. From this viewpoint, politics is understood as an ethical activity concerned with creating a “just society”. Even if one regards institutions such as businesses, community groups, clubs, trade unions, and also evaluation, as “public”, this broader perspective still remains a restricted view of politics in that it does not, and should not, infringe upon personal affairs and institutions. This view is illustrated, for example, by the tendency of politicians to draw a clear distinction between their professional conduct and their personal or domestic behavior. By classifying, say, cheating on their partners or treating their children badly as personal matters, they are able to deny the political significance of such behavior on the grounds that it does not touch on their conduct of public affairs.

Critical thinkers, in particular feminists, have pointed out that this implies that politics still stops at the front door; it does not take place in the family, in domestic life, or in personal relationships, something these and other critical thinkers disagree with. This kind of critique takes us to the broadest view on politics that is also the most radical. Rather than confining politics to a particular sphere (the government, the state or the ‘public’ realm), this view sees politics at work in all social activities and in every corner of human existence. Politics takes place at every level of social interaction; it can be found within families and among small groups of friends just as much as among nations and on the global stage. What makes politics a distinctive activity, distinguishable from any other form of social behavior, is that politics at its broadest, concerns the production, distribution and use of resources in the course of social existence. Politics is power: the ability to achieve a desired outcome, through whatever means. The essential ingredient is the existence of scarcity: the simple fact that, while human needs and desires are infinite, the resources available to satisfy them are always limited. Politics can therefore be seen as a struggle over scarce resources, and power can be seen as the means, through which this struggle is conducted, says Heywood.

Conflict and consensus

From the discussion thus far, it is clear that politics is inextricably linked to the phenomena of conflict and consensus. On the one hand, the existence of rival opinions, different wants, competing needs and opposing interests guarantees disagreement about the rules under which people live. On the other hand, people recognize that, in order to influence these rules or ensure that they are upheld, they must work with others. Hauge, Harrop, and Breslin (1992), for

example, point out that politics does not always involve conflict. They argue that one reason for studying politics is to search out the conditions under which groups can achieve their goals peacefully and effectively. From this view, politics is a constructive and practical subject and one emphasizes its compromising and consensual aspects. Politics relates not so much to the arena within which politics is conducted as to the way in which decisions are made. Politics is more seen as a particular means of resolving conflict, that is, by compromise, conciliation and negotiation, rather than through force and naked power. This is why Crick (1962) portrayed politics as that solution to the problem of order that chooses conciliation before violence and coercion. Crick, who is one of the leading exponents of this view, argues that when social groups and interests possess power, they must be conciliated; they cannot merely be crushed. This view on politics is also based on resolute faith in the efficacy of debate and dialogue. In other words, the disagreements that exist can be resolved without resort to intimidation and violence. Politics is no utopian solution (compromise means that concessions are made by all sides, leaving no one perfectly satisfied), but it is undoubtedly preferable to the alternatives: bloodshed and brutality. In this sense, politics can be seen as a civilized and civilizing force. People should be encouraged to respect politics as an activity, and should be prepared to engage in the political life of their own community.

Evaluation researchers' views on the evaluation and politics links

In the light of these definitions of evaluation and politics, evaluation can be part of the big political process (that is, evaluation in politics) and an aspect of the relationship between the actors involved in the evaluation process (that is, politics in evaluation). Even if evaluation in politics and politics in evaluation are not the most widely discussed issues in the evaluation literature (compared with, for example, models, methods and utilization), several evaluation researchers have dealt with the subject. The discussion below provides some notable examples that are illustrative rather than exhaustive of this discussion.

In the early years, Cronbach and his colleagues (1980) viewed evaluation as essentially a political activity through its influence on political decisions and policy formulation. More recently, one who extensively has discussed the matter is Weiss (1973, 1991). She

points out at least three ways in how evaluation and politics are linked. First, the policies and programs with which evaluation deals are themselves the products of political decisions. Second, because evaluation is undertaken in order to feed into decision making, its reports enter the political arena, where evaluation provides information. Third, evaluation itself has a political stance. Evaluation, by its very nature, makes implicit political statements, such as those challenging the legitimacy of certain program goals or implementation strategies. In this case, evaluation serves as critical inquiry.

The different kinds of information needs in the policy cycle are links that Chelimsky (1987, 1989) underlines in the relationship between politics and evaluation. She argues that evaluators must recognize and accept that politics is involved in evaluation and try to understand the dynamics of the policy cycle and the political process into which the evaluation is fed. The policy cycle consists of agenda setting, problem definition, policy design, program implementation, policy or program impact, and termination. At all stages, there is an information need where program evaluation can serve general audiences and individual public decision-makers. They may need information from evaluation for three very broad kinds of purposes.

- for policy formation – for example, to assess and/or justify the need for a new program;
- for policy execution – for example, to ensure that a program is implemented in the most cost / effective way; and
- for accountability in public decision making – for example, to determine the effectiveness of an operation program and the need for its continuation, modification, or termination. (Chelimsky, 1989, p 75)

Palumbo (1989) also notes that politics plays an important role in evaluation design, process and utilization of results. He comments on the claims that evaluators should not simply be advocates or collaborators of the program managers but of the program and policy itself, as well as of the clients and consumers of the program.

... evaluators may be the only way that the poor, students, offenders, welfare recipients, or mentally ill can influence the policy. These "stakeholders" often are not included in the formulation and implementation of the evaluation. It is in this way that evaluators can represent the public interest rather than specific power holder interest. (p. 38)

Being an advocate for, or at least having an ambition to give, unprivileged stakeholders a voice in the evaluation is one way that evaluators incorporate politics into their works.

Micro- and macro-levels

Greene (2003) shows how evaluation and politics are interwoven from micro- and macro-levels. She starts with the question of what the war in Iraq in the spring of 2003 had to do with evaluation. Her answer is that, in a discussion of evaluation and politics, world events such as war and peace, weapons and diplomacy, oppression and freedom are of central importance. Then she describes how macro politics and micro politics are combined when she meets people in her evaluation work who express concern for relatives in the war; this reality then has effects on the evaluation activities and even how the evaluators' questions (unrelated to the war) are answered. In this way, macro events like the war in Iraq affect the micro work the evaluator does both in conducting the evaluation and reporting the results. This example shows that the evaluator must consider what occurs at both the macro political and micro political levels.

House (2003) provides one more example of this perspective, illustrating how the micro-level view of the role of evaluation in politics has implications on the micro-level choice of an evaluator. He frames a future scenario where evaluation is a tool at the disposal of the powers in force. House describes how evaluators who stand for a perspective that is critical of society will have greater difficulty winning government contracts. Instead, it is the evaluators who are willing to tow the party line who will be hired. Thus, in a sophisticated way, politically-correct evaluators are selected by a process of reverse discrimination whereby one does not blacklist people (which would risk a public debate) but instead "white lists" those one knows are favorable in terms of competence and appropriateness.

How does evaluation influence politics?

The focus so far has been on evaluation writers' perspectives on politics' influence on evaluation. How can evaluation influence politics? This question can be answered from several perspectives. First, from a positivist, rational or social engineer's perspective, evaluation fulfils a rational feedback function within the political system and a steering control function. Evaluation provides the 'rational' and 'unbiased' data that the system needs to determine whether it is on course. Second, from a cultural perspective, evaluation can be understood as one answer to the fundamental need to

be able to associate an organization with meaning and rationality. Evaluations can also fulfill a symbolic or ritual function and can be an answer to the trust that has declined in society today. Those in power and public organizations can use evaluations to recreate legitimacy for a program or operation, according to Hanberger (2003). He mentions that an evaluation can fulfill an enlightening (Weiss, 1977), a conceptual (Peltz, 1978) or a learning function (Preskill and Torres 2000). In addition, evaluations can be used in media debate or in direct meetings with interest parties where the results from the evaluation and possible lines of action are discussed. Such an evaluation function can be described as stimulating public debate.

Stern (2005) distinguishes the following five purposes for evaluation that give a view of how evaluation can have an impact on political decisions for planning, learning, developing and termination of a program.

- o Planning/efficiency – ensuring that there is a justification for a policy / programme and that resources are efficiently developed.
- o Accountability – demonstrating how far a programme has achieved its objectives and how well it has used its resources.
- o Implementation – improving the performance of programmes and the effectiveness of how they are delivered and managed.
- o Knowledge production – increasing our understanding of what works in what circumstances and how different measures and interventions can be made more effective.
- o Institutional and community strengthening – improving and developing capacity among programme participants and their networks and institutions. (Stern, 2005, p. xxvii)

In summary, some evaluation writers have noted that evaluation and politics can be interpreted from a narrow perspective, as the art of government where evaluation is seen first and foremost as a technical instrument to get information and basic data to the decisions making process. Other commentators take a broader perspective that expands the concept of politics to the public arena and thereby to different social institutions, including evaluation. Political- and value-laden aspects are therefore part of evaluation. Finally, in the broadest interpretation of politics, some evaluation writers argue that all aspects of social life, in both the public and private spheres, are inherently political. From this perspective, not only is evaluation

as an institution and undertaking political, but the individual evaluator’s values, background, gender and the like also become part of the explicit and implicit operation of politics in evaluation.

Three positions on how evaluation and politics are related

The examples above show that more and more evaluators are accepting the reality of connections between evaluation and politics. What remains unclear is the inherent nature of these connections and, based on this, the range of possibilities for and the limitations of the evaluation-politics relationship. In this section, we propose and describe a framework to clarify the nature of the connections between evaluation and politics. We also explore the implications of the three different positions that comprise the framework, both for the conduct of evaluation and for the evaluation profession.

The connection between evaluation and politics can be framed in three different ways. These ways, which can be characterized as ‘positions’ or ‘perspectives,’ differ along two dimensions: whether it is possible operationally to separate evaluation and politics, and whether it is desirable conceptually to separate evaluation and politics. In this framework, we have adopted the conception that the two main components of evaluation are providing information (the epistemological component) and providing judgement (the value component).

Table 1: Three positions on the inherent connections between evaluation and politics

Three Positions	Possible to separate evaluation and politics?	Desirable to separate evaluation and politics?
First position	Yes	Yes
Second position	Yes, in providing information; Not entirely when providing judgements	Yes, in providing information
Third position	No	No

The first position holds that it is both possible and desirable, operationally and conceptually, to separate evaluation and politics. The second position maintains that it is possible and desirable to separate evaluation and politics operationally when providing information but not entirely possible to do so when providing judgments, nor it is conceptually desirable. The third position is that it is neither possible nor desirable, operationally or conceptually, to separate politics and evaluation.²

It is important to acknowledge that the three positions are general characterizations and that individual evaluators do not neatly fit into only one position, especially if we consider those with long histories of evaluation work of many sorts and in different contexts. We have made the boundaries more distinct than they are in the complex, pragmatic undertaking that is evaluation. We have done this to highlight the primary differences in the view of the relationship between evaluation and politics among the three positions.

First position – The value-neutral evaluator

The viewpoint from the first position is that politics and evaluation can and should be kept operationally and conceptually apart. The evaluator works independently to provide an objective, neutral assessment of the program, project or policy; the politician then receives this assessment and does with it what he or she decides. This view suits the definition of politics as the art of rational government, where the evaluator is an objective, impartial civil servant. The information function of evaluation should be under the control of the evaluator and be his/her primary activity. The judgment function, based on the information, should be under the control of others, including politicians, program planners and implementers, and the electorate. In this view, evaluation is “social research.”

According to Schwandt (2003), some of those who hold this type of position look at politics as something incomplete and faulty which needs to be held in check to prevent it from poisoning the good relations between people. The cure for these faults is objec-

2 From a logical standpoint, there is a fourth position: that it is not possible to separate evaluation and politics (either the information or the judgment aspects) but that it would be desirable to do so (in both aspects). Because this is not a realistic possibility to guide evaluation work, we have not considered it here. There are some, however, who might argue that serious consideration should be given to this position, because, if it can be shown to be highly desirable, the evaluation community might begin to set in place policies and procedures to bring about the separation. The latter assumes, of course, that the ‘evaluation community’ could and would speak with one voice on this matter. As the three positions described here show, this is unlikely to occur.

tive, impartial, rational and professional officials who are above the temptation to promote their own or selected others' interests, who maintain the public's interests and assert general principles of justice that treats everybody equally. As House and Howe (1999) have noted, this relationship between politics and evaluation neatly fits the representative liberal model of democratic theory (Ferree, Gamson, Gerhards, & Rucht, 2002) in which disinterested, apolitical experts inform public decision making in a detached (i.e., emotion- and value-free) manner, thereby enhancing both the rationality as well as the civility of the debate about a suitable course of action in the free marketplace of ideas.

Against this picture of how politics can become a threat to objectivity, impartiality and rationality, the question to ask is how the evaluator can protect him or herself from political influences. One way to separate evaluation from politics is to emphasize its autonomy in relation to political institutions and to powerful interests in society. Closely connected to this is the idea that power is a source of corruption which evaluation must be insulated against if it is to be conducted effectively.

How can these political influences be minimized? In his winning response to the 1988 AEA President's Problem (Patton, 1988) around the question of evaluation and politics, Robin Turpin (1989) focuses on ways to minimize the political influences in evaluation. Specifically, politics can influence (p 55):

- the selection of the evaluator or evaluation team
- chances of funding
- the selectivity of information given to the evaluator
- the general approach or scope of the evaluation project
- the methods used
- the subject or subject pool selection
- the instruments used or developed
- data analyses
- the interpretation of data
- final recommendations
- information that is disseminated

To “produce good, solid, objective research” (p 55) Turpin suggests that the evaluator should take the following precautions to avoid or minimize political influences (which we have rewritten in minor ways).

- Uncover who wants the evaluation and the motivation behind it
- Uncover all sides of the story by talking to the people involved (not just those officially involved)
- Develop peer review procedures (even for internal, non-funded or routine evaluations)
- Make use of expert panels and/or outside consultants in the whole evaluation process
- Use established scales and instruments whenever possible, and
- Include in the report a ‘limitations’ section that discusses possible political influences and details critical decisions

Although Turpin also notes that politics can have positive effects on evaluation by opening doors to cooperation and information, even these positive effects can extract a cost, often in the form of subtle pressure on the evaluator. “Politics has a nasty habit of sneaking into all aspects of evaluation”, Turpin comments.

The recurring idea that is emphasized is the evaluator as a conscious actor, on guard against undesirable influence and attempts to hinder the evaluation from its task of critical evaluate. The ideal is a professional, disconnected actor, who completes his or her assignment without regard to the more or less explicit desires of the powers that be.

Second position – the value-sensitive evaluator

In the second position on the connection of evaluation and politics, it is accepted that evaluation takes place in a political environment and that evaluation and politics therefore cannot entirely be separated, specifically in the judging component of evaluation. In the operational, information-finding aspects, however, the evaluator can and should stay separate from the political component. For example, Chelimsky (1987) points out the need for evaluators to place themselves in the political context that constitutes the program evaluation and suggests that evaluators must understand the political system in which evaluation operates and the information needs

of those policy actors who utilize evaluations. She says that evaluators must devote much time to negotiating, discussing, briefing, accuracy-checking, prioritizing, and presenting. At the same time, the evaluator takes a professional role for the conduct of the evaluator that is non-political in the narrow definition of politics.

This second position emphasizes the evaluator's role as a professional expert, but it includes two distinctively different ideas on how politics and expertise can be conceptualized. The first idea has a market perspective and reduces the evaluation-politics relation to a technical task, where the profession is defined by the measurement of quality and efficiency. This is in line with the narrow definition of politics as governance that was presented above. The other idea represents a value-committed perspective on the relation that concerns a professional role that makes the evaluation more democratic. This is more in line with the definition of politics as a public sphere.

Evaluation and politics as a market

From a market perspective, politics is reformulated to be primarily a matter of practical problem solving (Amy, 1984). This technocratic view of politics has come to prominence as part of the worldwide spread of neo-liberal discourse. Politics is replaced by rational consumer choice. Here, evaluation becomes a means for quality assurance that measures the performance (efficiency) of practices against indicators of success in achieving the targets. The profession of evaluation is reduced to technical expertise to measure quality and performances through prefabricated schemas and formula. The current emphasis in some counties in the education and health arenas for indicators-based performance management also fits within this characterization.

The movement is known as New Public Management (NPM) and represents a solution to problems in the public sector based on the introduction of management ideas from the private sector. Power (1997) describes what is characteristic of the movement:

Broadly speaking the NPM consists of a cluster of ideas from the conceptual framework of private sector administrative practice. It emphasizes cost control, financial transparency, the autonomization of organisational sub-units, the decentralization or management authority, the creation of market or quasi-market mechanism separating purchasing and providing functions and their linkage via contracts, and the enhancement of accountability to customers for quality of service via the creation of performance indicators. (p. 43)

The citizens are transformed to consumers that make choices on a market of health care, education, social welfare etc. Evaluation is seen as a practice that can guide consumer's choice. The view is that institutional structures for control and "accountability" should be strengthened and that evaluation in the first instance should be defined as a steering instrument for management. Through performance management and measurement and the control of quality, politicians are in a position to demonstrate "value for money" to tax payers. Furthermore, NPM provides a rationale for reducing public sector spending through its support for private solutions rather than politically controlled activities.

Politics and the democratization of evaluation

The other variant of the second position clearly admits that evaluation and politics are not entirely possible to separate, especially when talking of politics in a broad definition that places it in the public sphere. Evaluation is an activity necessarily couched in a political context, and the evaluator must take responsibility for how the evaluation is done not only in regard to the technical aspects but also with attention to ethical aspects and democratic values. This does not mean that evaluation is totally integrated in politics because the evaluator has a distinctive role separated from politics, in the narrower sense of that term, as the provider of relevant, meaningful information.

From this perspective, there is a responsibility for evaluators to make their own professional perspectives on the evaluation visible. The answer to how this could and should take place is given in different forms. Some forms include the evaluator being a facilitator, a critical friend, a dialogue partner, or an educator. In general, the evaluator supports active involvement from stakeholders in the evaluation (Conner, 2005). Special attention is often directed to those who in lack of power to get their problems and questions observed in the evaluation. Here, evaluation is not reduced only to be a technical matter but includes attention to the democratization of the evaluation process, thereby potentially contributing to a larger democratization of the program and its context.

The democratization occurs in the central components in the evaluation process. These components include deciding on the aims for the evaluation (control, development, enlightenment, learning, etc), determining the resources for the evaluation (economy, social and political), and selecting the evaluation questions and methods. Politics, values and power are also apparent in decisions about access

to information and where in the organization the evaluation is centered, as well as whether an internal or external evaluation is undertaken.

Some evaluation models can be connected to this view on evaluation and politics. One of the first researchers to formulate a demand for democratic evaluation was MacDonald (1973, 1977). In his version of democratic evaluation, the starting point is the assumption that power is distributed among interest groups and that the evaluator ought to serve the public's right to know. One of the recent contributions to the field of democratic evaluation is House and Howe's (1999, 2000) deliberative democratic approach. In their view, evaluation process must be based on the full and fair inclusion of all relevant stakeholders and represent the views of socially disadvantaged groups. Therefore, House and Howe are keen to emphasize that the evaluator has a special responsibility to those stakeholders who might not normally be 'heard' (because they are relatively powerless, invisible, unorganized, or for some other reason not likely to be included). To serve the interests of socially disadvantaged groups, the evaluator has to give them a voice in the evaluation.

At the same time, House and Howe reserve the right of the evaluator to make the final pronouncement of the merit, worth or value of the program under consideration. The idea of procedural justice – central to a theory of deliberative democracy – demands that all voices have had a fair hearing and are involved in deliberation. However, this does not mean that the evaluator necessarily takes the side of these less powerful voices. Advocating for the inclusion of those less heard from is not the same as endorsing their interests or points of view. Others, who also urge the evaluator to involve interest groups in an evaluation, designate the evaluator's role as that of consultant for these interest groups (Fetterman, 1994; Patton, 1994, 1996). In this situation, the evaluator becomes a "facilitator" and throughout the evaluation adopts a neutral position with respect to the interests of different groups as they strive to empower themselves as individuals and as a group.

There are several other models of participatory and collaborative evaluations that have strong emphasis on the aim to democratize not only the program context but also society as a whole. Cousins and Whitmore (1998) distinguish between practical and transformative evaluations. Practical participatory evaluation focuses on participation in evaluation. The evaluator assumes responsibility for carrying out technical evaluation tasks, and stakeholders are involved

predominantly in the definition of the evaluation problem, scope-setting activities, and, later, the interpretation of data emerging from the study. In transformative participatory evaluation, the aim has expanded. Here, one strives for more extensive engagement of stakeholders, for radical social change, and for clarifying values that inevitably shape evaluations.

Third position – the value-critical evaluator

In comparison with the first and second positions on evaluation and politics, the third position does not see politics stopping at the private sphere but instead views politics as something integrated in our everyday life. Because of this, there can be no separation between evaluation and politics and therefore no neutral value or operational position taken by the evaluator. The position is associated with a perspective that claims not only that human values are inseparable from descriptions of facts but also that science will benefit from admitting this. With reference to Taylor (1985), Geir (2004, p 197) says that

...values are an intrinsic part of the interpretive process in two ways, individual and common. The interpreter chooses a theoretical framework or conceptual structure in which she understands the phenomenon in question. These frameworks are pre-models of understanding, initially opening some possible connection and closing others. (p 197).

In this view, it is important for evaluators to formulate a theoretical framework for a broader understanding of the program or subject that is evaluated. Evaluation approaches that could be connected with this kind of ideas are characterised to be:

intentionally and directly engage[d] with the politics and values of an evaluation context, in order to explicitly advance particular political interests and values, and often also, to effect some kind of socio-political change in the evaluation context itself. Examples of value-engaged evaluative stances include feminist, empowerment, and democratizing approaches to evaluation. Proponents of these approaches are characteristically informed by ideologically-oriented methodological traditions such as feminism and critical theory. (Greene, 2003).

It is important to note that the borderline between this third position and the “democratic and participatory” variant of the second position is by no means clear-cut. Among those who argue for the desirability of separating some parts of evaluation and politics, as

those in the second position do, are evaluators who also embrace the value-laden quality of human action and thus also of knowledge about human action. What differs between the second and third positions is the relative importance given to the values of social change and transformation.

A broad view on politics and evaluation

From the third position, politics is not viewed as something negative. It is conceptualized in considerably broader terms than only a question of asserting one's own interests and exercising power. Politics is defined as an activity through which we live together and regulate or adapt our goals and efforts. It is also conceptualized as critical reflections on the public good. The basic idea is that it is via citizenship – through people deciding together how they will act and then following through with it – that an individual can achieve his or her full potential. Politics is concerned with taking a stance, being touched and engaged by something, defining right and wrong, good and evil, and acting on one's convictions. With these viewpoints, politics is inherently human, with roots in morals and values (Schuman, 1977).

With morals and values brought into the picture, a number of new critical questions arise concerning who conducts evaluation and for whom, which evaluative questions will be raised, and what judgment criteria will be employed. The stand the evaluator takes on these questions determines the judgment he or she presents. This kind of idea plays a central role in the understanding of how the relations between evaluation and politics are conceptualized. Politics like citizen activity requires both an intellectual and physical arena, a public forum in which people can come together and plan for action. The space provided in voting halls is insufficient; politics requires involvement between elections. One alternative is to go to the streets and demonstrate; others are public enterprises where people meet, for example in pre-schools, schools, and in associations where one has an active interest. Another example of an arena for citizen involvement is evaluation conducted openly with the participation of various interested parties.

Dahler-Larsen (2003) is an evaluation researcher who places the question of evaluation politics on this broader societal level or "res publica". He views evaluation as a creative force in our understanding of society. He looks at evaluation as a cooperative and structuring force in our understanding of society. Evaluation is defined as a practice that describes other practices and that forms our impressions of these by naming the efforts, goals, criteria, standards,

and the like. In this way, evaluation gives prior interpretation of the public efforts and the values that comprise them. Based on Beck's (1994) term, Dahler-Larson notes that we live in a "reflexive modernity" where confidence in progress decreases in concert with the increasing time spent grappling with the problems that these create. According to Beck, the security that has until now been associated with the modern projects' progress has been weakened in the new "reflexive modernity". Instead, "reflexivity" reigns in a double sense. First, the reflexivity is a throwing-back of side effects onto society itself (environmental problems, highway congestion, coordination problems, iatrogenic illnesses, etc.). Second, the reflexivity is an increased moral, ethical and political concern for the handling of these side-effects. One such "side-effect" is reactions to public policies from users and other stakeholders.

These changes in how one looks at the ontological and epistemological foundations for evaluation, and on society in the perspective of new reflexive modernity, have also changed the political framework for evaluation. Society is not the only thing that has become more complex. Evaluations have been given many different functions as well. These functions include some traditional ones, such as the use of evaluation as an instrument for national and local governments to exercise control and as an instrument for society and citizens to receive information and knowledge. A newer function for evaluation includes its use as an instrument for interested parties and organizations to observe and influence.

Evaluation, however, does not simply disseminate results; it also provides a deeper, better understanding of the evaluated object.

Through linguistic designations of "the evaluand", "the points of measurement", "criteria", "standards", evaluation discourse draws attention to certain phenomena and orientations. Hereby evaluation is an interpretation of what the public effort is altogether and in wherein its value consists. (Dahler-Larsen, 2003)

From this point of view, evaluation informs about the merits and value of a program but also has a broader perspective. This type of evaluation informs about a larger framework, with reference to roles and relations. Evaluations become a meta-communication about the character of people and their relations, which in turn are an arrangement of politics in its deep meaning. This does not mean that evaluations always have this impact on our conceptions of the world and ourselves. How strong the impact is depends on a number of contextual factors such as organization, culture and structure.

Summary

In our discussion on the subject of evaluation and politics, we have assumed that evaluation is not an isolated island but instead an enterprise in a political context. This context means that there are multiple actors and institutions with power and interests to influence the evaluation, from the choices of criteria, standards and methods, as well as the choice of an evaluator.

We have described three views on the relationship between evaluation and politics. The first position sees politics as driven to protect its own interests and as harmful to evaluation. In this view, politics is at best a fickle partner, driven by many influences other than information, and at worst an unsavory one. Evaluation can and should be kept apart from it. If an evaluator has to deal with politics, the evaluator must be careful not to be too engaged and not to scrutinize the political influences to decide how to behave. Instead, the evaluator uses professional standards and guidelines to produce objective information, so that if and when the possibility to use it arises, the information is available.

In the second view, in one interpretation, politics is replaced by the idea of the market with rational consumers making choice based on evidence. Here, the political is paradoxically transformed into an outwardly apolitical phenomenon – a style of formalized accountability that becomes the new ethical and political principle of governance (Power, 1997). The role of evaluation in this view is to provide professional technical help to measure quality and to produce quality-assurance. A different interpretation of the second position is to acknowledge the inseparable connections between evaluation and politics in the area of value- or judgment-making and therefore to democratize the evaluation process at critical stages (for example, deciding on the evaluation questions), with special attention to those whose voice may not be easily heard in the public arena. At the same time, however, evaluation is kept separate from politics in the implementation of the evaluation, to avoid biases in the information produced.

The third position views evaluation and politics are inseparable, both in the conceptual and operational aspects. Here, the evaluator accepts that evaluation and politics are connected in many intricate ways and acts accordingly. The evaluator acknowledges and states his or her own ethical and moral standpoints so these are transparent during the evaluation process. Actions such as these place evaluation and evaluators in a more prominent role shaping society and its politics.

Discussion and implications

Each of the three positions (and sub-positions) can be criticized on different aspects. One could question the claims held in the first position that evaluation can be independent from external power. Those who criticize the idea of evaluation's autonomy from external power believe that evaluation easily can become a part of, and work for, the ideological state apparatus in society. Another criticism of the first position focuses on the idea that evaluation is a value-free practice of objective research. Social science has prided itself on being value-free for many decades. However, Scriven (2003/04) notes that this view of social science research is changing as social science becomes more involved with serious social problems, interventions and issues. To be successful in this new arena, social science will have to incorporate evaluation or evaluative elements, he says. A final criticism of the first position is that it is difficult to separate the judgment-making component of evaluation from politics, both on an individual level and on a societal level. Hammersley (1995) presents several arguments why values cannot be insulated from research. One of his arguments notes that, because information and knowledge are always produced within a perspective or framework, the knowledge one prioritizes is also dependent on circumstances in the socio-political context. Another of his arguments focuses on how the researcher's or evaluator's own personal and positional realities (ethnic, gender, economic and the like) play an important role in shaping priorities and interests that can affect an evaluation.

Criticism could also be directed at the second position, with evaluation and politics related conceptually in judgment-making but separated in information-making. The market-oriented variant of this position expects that central values will be based on the needs of the market and expressed by the multiple actors representing different interests. However, only a subset of actors is effectively involved, and the particular subset will shape the normative content of an evaluation, determining the boundaries of the "knowledge base, the scope, and potentially the outcomes of evaluation" (Dabinett & Richardson, 1999, p. 233). The indicators-based performance management focus that is central in the market oriented perspective also carries risks. Four of these risks are that indicators may not measure what they are intended to; that unwarranted attributions of causality for outcomes made be made to indicators; that performance information may be used for purposes for which

it was not intended; and that goal displacement may occur if incentives divert effort from attaining program objectives to meeting the requirements of measuring and reporting (Davies, 1999). Performance measurement systems also decouple accountability from ownership and responsibility, thus assigning to accountability a role in regulation and control and inhibiting shared responsibility among stakeholder-citizens. They “also let the evaluator off the hook, by heavily obscuring their authorship and thereby muting their responsibility” (Greene, 1999, p. 170).

Some criticize the other variant of the second position, focused on democratic approaches to evaluation, because it tends to be connected to the macro politics of society, in that evaluation is expressly positioned as an instrument of democracy and as an advocate for democratic ideals and for change. The explicit ideological stance and political positioning are democratic and the express point of evaluation in these approaches is to render an assessment and judgment of evaluation quality that inherently incorporates democratic standards of judgment and thus serves to advance democratic ideals and values. Above, we mentioned several evaluation researchers that represent these ideas. One more example of this is Mark, Henry and Julnes (2000) who clearly put evaluation in a political discourse of democratic decision-making and also reject the fact-value dichotomy. At the same time, Greene and Walker (2001) note that these authors have

...positioned evaluators and the knowledge they generate apart from the politicized fray of democratic decision-making. From this position, evaluators can use a mix of methods within selected inquiry modes to impartially make sense of the quality of, and the diverse values that accompany, a given social program or policy and then offer that assisted sense-making to those in democratic institutions for their deliberations (p 371).

Against this position, Greene and Walker argue for an alternative view that

does not separate the practice of evaluation from socio-political practices and institutions to which it is designed to contribute or in which it is embedded...Evaluators should not be absolved from the moral and ethical responsibility for the practice choices we make and the knowledge claims we generate. If we wish to claim that a particular social program or policy can indeed contribute to social betterment, we must be responsible for that claim-both as a warranted representation of human experience

and as a defensible valuing of what is “good” and “right” about that experience. (Greene & Walker, 2001, p 371).

What Greene and Walker criticize is the idea of a detached “professional” evaluator that is central in the concept of evaluation and politics held by those working in the democratic variant of the second position. Those working from this perspective also need to address and resolve the problem of identifying and securing comprehensive, representative stakeholder involvement. Furthermore, one could ask how the representatives of groups, sectors or interests are to be chosen, and how the differences in power among stakeholder groups influence their roles in the evaluation. These questions highlight the dilemma facing the evaluator when he/she tries to strengthen powerless stakeholders. One could also ask about the value position that motivates such a decision, and about the influence such an “empower-the-powerless” standpoint is likely to have on confidence in the evaluation among other more traditional, empowered groups (Karlsson 1996, 2001).

Finally, the third position, that evaluation and politics are inseparable in all ways, has some limitations and raises some questions, as is the case with the other two positions. The ethical and moral standpoint that demands a better world, a more equal society and a fight against any discrimination leaves no private zone where less-than-enthusiastic support for these ideas can be hidden. Here, the evaluator cannot, so to say, hide behind a professional role if one chooses not to take a stand on these issues. One could ask if this broad and expanded role for evaluation makes the evaluator more of an intellectual discussant on general political, ethical and moral issues and less of a professional narrowly examining a program in accordance with more specific goals and chosen criteria. Are evaluators trained and skilled to play such a broad, prominent role in societal discussions, and, even if they are, can they reasonably and responsibly fulfill such a broad role? In this more prominent role, what assurances are there that the reasons for the evaluator’s value stance are transparent? How can we know, for example, the extent to which an evaluator’s views are motivated by his/her general personal values rather than by specific factors related to the evaluation? Also, are there safeguards in place that will allow the evaluator to share his/her viewpoints without silencing the views of others who could participate, including those who are often voiceless in the political process? Rather than being the spokesperson for others’ views, maybe the evaluator should work to let them speak for themselves.

An interpretation of evaluation from this broad moral perspective could be that all who work with people in different situations, especially when one has power over others' live, health, education, or security, have the responsibility to reflect actively and systematically about one's own behavior and to be self critical. Here, we can talk about an "evaluator role" that is integrated in every responsible profession, including physicians, teachers, social workers, lawyers and others. This view raises the question of whether there is space for a profession that exclusively deals with evaluation, not as an alternative but as a complement to all other professionals and to their own evaluations and critical reflections on what they are doing.

Implications

What we can learn from our review of the relationship between evaluation and politics is how the relationship between the two is much more complex and difficult to grasp than thought in earlier decades (the 70s, 80s and 90s). Today, we must take a more nuanced view of the evaluator, not simply considering him or her to be a neutral, independent, objective methodologist who presents facts. This older, traditional image can be contrasted with an image from the other extreme that places evaluators (and evaluations) in a political powder keg where various interests and values meet and clash. The better image is probably one in the middle of this spectrum: a professional, skilled, well-trained evaluator working in a context with explicit or implicit political, cultural and personal implications, all of which can potentially exert some influence in the decisions about evaluation questions, methods and results. It is clear that, for better or worse, evaluation and politics are partners. The decisions an evaluator makes are affected not only by issues of science but also by politics and ethics.

What can evaluators do to maximize the benefits of the link between evaluation and politics and minimize its risks? One piece of advice is for the evaluator to watch for the diverse supports and unexpected opportunities that exist in a large, complex context. Another suggestion is that the evaluator be clear about the special skills and perspectives or 'value added' that he or she brings to the situation, in relation to the other participants. These are the anchors around which the evaluator should build. Another suggestion is to have a supportive base in the evaluation profession, an evaluation network or some other professional group. This provides another type of anchor and perspective, when pressures build that the evaluator is not fully in control of.

Although these suggestions mostly focus on the individual evaluator, we also think that there is a need to scrutinize more critically what purpose evaluations can serve. In the wake of increasing uncertainty about how public enterprises can be steered, controlled and developed through democratic decisions, expectations increase about evaluations' ability to solve these steering problems. This has led to evaluation enterprises being viewed as a self-evident requirement at all levels of society. Management and personnel are expected to spend more time finding out about how their efforts are perceived by users and other people who are affected. As a consequence of these increased evaluation efforts, there has been an expansion of administrative systems to handle the information that comes in, which in turn requires more resources. We would argue for an alternative to this expansion of evaluation into a large bureaucratic system, in favor of a shift toward more reflective, critical-focused evaluation as part of every practitioner's work toward a democratic, humanistic ideal that gives marginalized groups a voice.

References

- Amy, D. J. (1984). Why policy analysis and ethics are incompatible. *Journal of Policy Analysis and Management*, 3(4): 573-591.
- Beck, U. (1994). The Reinvention of Politics: Towards a Theory of Reflexive Modernization. In U. Beck, A. Giddens, & S. Lash (Eds.), *Reflexive Modernization* (pp. 1-55). Stanford: Stanford University Press.
- Bridges, E. M., & Groves, B. R. (2000). The Macro- and Micropolitics of Personnel Evaluation: A Framework. *Journal of Personnel Evaluation in Education*. 13 (4), 321-337.
- Caro, F. G. (1977). *Readings in Evaluation Research*. (Second ed). New York: Sage.
- Chelimsky, E. (1987). The politics of program evaluation. In D. S. Cordray, H. S. Bloom, & R. J. Light (Eds.), *Evaluation practice in review, New Directions for Program Evaluation*, No 34, 5-21. San Francisco: Jossey Bass.
- Chelimsky, E. (1989). Linking program evaluation to user needs. (72-99). In J. D. Palumbo (Ed.), *The Politics of Program Evaluation*. London: Sage Publications.
- Conner, R.F. (2005). The Colorado Healthy Communities Initiative. *American Journal of Evaluation*, 26: 363 - 368.
- Cousins, J. B. & Whitmore, E. (1998). Framing Participatory Evaluation. *New Directions for Evaluation*, no 80 (5-23).
- Crick, B. (1962). *In defence of politics*. London: Weidenfeld and Nicolson, cop
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, D. C., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1981). *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*. San Francisco: Jossey-Bass Publishers.

- Dabinett, G. & Richardson, T. (1999). The European spatial approach. The role of power and knowledge in strategic planning and policy evaluation. *Evaluation*, 5, 220-236.
- Dahl, R. (1984). *Modern Political Analysis*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Dahler-Larsen, P. (2003). The Political in Evaluation. *E-journal Studies in Educational Policy and Educational Philosophy*, 2003:1. <http://www.upi.artisan.se>
- Davies, I. C. (1999). Evaluation and performance management in government. *Evaluation*, 5, 150-159.
- Easton, D. (1968). *The political system: an inquiry into the state of political science*. (2nd ed.). New York: Knopf.
- Ferree, M.M., Gamson, W.A., Gerhards, J., & Rucht, D. (2002). Four models of the public sphere in modern democracies. *Theory and Society*, 31: 289-324.
- Fetterman, D. (1994). Empowerment evaluation. *Evaluation Practice*, 15, 1-15.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program Evaluation. Alternative Approaches and Practical Guidelines* (3rd ed.). Boston: Pearson
- Gallie, W. B. (1956). Essentially Contested Concepts. *Proceedings of the Aristotelian Society*, p. 167-198
- Geir, A. (2004). Facts, values and moral education. *Nordisk Pedagogik*, 24 (3), 195-211.
- Greene, J. C. (1999). The inequality of performance measurements. *Evaluation*, 5, 160-172.
- Greene, J. C. (2003). War and Peace...and Evaluation. *E-journal Studies in Educational Policy and Educational Philosophy*, 2003:1. <http://www.upi.artisan.se>
- Greene, J. C., & Walker, K. (2001). Book Reviews in *Evaluation and Program Planning* 24 (2001), 368-371 on Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass.
- Hammersley, M. (1995). *The Politics of Social research*. London: Sage
- Hanberger, A. (2003). Evaluation's hidden politics. *E-journal Studies in Educational Policy and Educational Philosophy*, 2003:1. <http://www.upi.artisan.se>.
- Hauge, R., Harrop, M., & Breslin, S. (1992). *Comparative Government and Politics. An Introduction*. (3rd ed). London: MacMillan Press.
- Heywood, A. (2002). *Politics*. (2nd ed). London: MacMillan Press.
- House, E. R. (1991). Big policy, little policy. *Educational Researcher*, 15(5), 5-12.
- House, E. R. (2003). Bush's Neo-Fundamentalism and the New Politics of Evaluation. *E-journal Studies in Educational Policy and Educational Philosophy*, 2003:1. <http://www.upi.artisan.se>
- House, E. R., & Howe, K. R. (1998). The Issue of Advocacy in Evaluations. *American Journal of Evaluation*, 19 (2), 233-236.

- House, E. R. & Howe, K. R. (1999). *Values in Evaluation and Social Research*. Thousand Oaks, CA: Sage.
- House, E. R., & Howe, K. R. (2000). Deliberative democratic evaluation. In K.E. Ryan and L. DeStefano (eds.), *Evaluation as a democratic process: Promoting inclusion, dialogue, and deliberation*. *New Directions for Evaluation*, no. 85 (pp. 3-12). San Francisco: Jossey-Bass.
- Karlsson, O. (1996). A critical dialogue in evaluation. How can the interaction between evaluation and politics be tackled? *Evaluation*, 2, 405-416.
- Karlsson, O. (2001). Critical dialogue: Its value and meaning. *Evaluation*, 7, 211-227.
- Lasswell, H. D. (1936). *Politics: Who gets what, when, how*. New York: McGraw-Hill.
- MacDonald, B. (1973). Briefing Decision Makers. In E. R. House (ed.), *School Evaluation: The Politics and Process*. Berkeley, CA: McCutchan Publishing Corp.
- MacDonald, B. (1977). A Political Classification of Evaluation Studies. In D. Hamilton, D. Jenkins, C. Kong, B. MacDonald, and M. Parlett, *Beyond the Numbers Game*. London: Macmillan.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass.
- Palumbo, J. D. (Ed.). (1989). *The Politics of Program Evaluation*. London: Sage Publications.
- Patton, M. Q. (1988). Politics and Evaluation. *Evaluation Practice*, 9, 89-94.
- Patton, M. Q. (1996). Utilization-focused evaluation. London: Sage.
- Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice*, 15, 311-319.
- Pawson, R., & Tilly, N. (1997). *Realistic Evaluation*. London: Sage.
- Peltz, D. C. (1978). Some expanded perspectives on use of social science in public policy. In: Yinger, J.M. & Cutler, S.J. (eds.) *Major social issues: A multidisciplinary view*. New York: Free Press.
- Polanyi, M. (1968). The republic of science: its political and economic theory. In E. Shils (ed.), *Criteria for Scientific Development, Public Policy, and National Goals*. Cambridge, Mass., MIT Press.
- Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Preskill, H. and Torres, R.T. (2000) 'The Learning Dimension of Evaluation Use.' In: *New Direction for evaluation*, No 88, winter 2000.
- Schuman, D. (1977). *A Preface to Politics*. Toronto: D. C. Health & Company.
- Schwandt, T. A. (1998). Constructivist, interpretivist approaches to human inquiry. In Denzin, N. K. & Lincoln, Y. S. (eds). *The Landscape of qualitative research. Theories and issues*. Thousand Oaks: Sage.

Schwandt, T. (2003). In Search of Political Morality of Evaluation Practice. *E-journal Studies in Educational Policy and Educational Philosophy*, 2003:1. <http://www.upi.artisan.se>

Scriven, M. (1991). *Evaluation Thesaurus* (4th ed.). London: Sage Publications.

Scriven, M. (2003/2004). Michael Scriven on the Differences Between Evaluation and Social Science Research. The Evaluation Exchange. Volume IX, No. 4, Winter 2003/2004. p 7.

Stern, E. (2005). Editor's Introduction. In E. Stern (ed.), *Evaluation Research Methods*. Vol. I. London: Sage. (p. xxi-xliiii).

Taylor, C. (1985). *Philosophy and the human sciences. Philosophical papers 2*. Cambridge: Cambridge University Press.

Turpin, R. (1989). Winner of 1988 President's Problem. *Evaluation Practice*, 10 (1), 53-57

Weiss, C. H. (1973). Where Politics and Evaluation Research Meet. *Evaluation*, 1, 37-45.

Weiss, C. H. (1991). Evaluation research in the political context: Sixteen years and four administrations later. I M. W. McLaughlin & D.C. Phillips (Eds.), *Evaluation and Education: At Quarter Century* (s. 210-231). Chicago: The University Press.

Weiss, C. H. (1977). *Using Social Research in Public Policy Making*. Lexington Books, Massachusetts.

MONITORING AND EVALUATION, AND THE KNOWLEDGE FUNCTION¹

*by David Parker, Deputy Director,
UNICEF Innocenti Research Centre*

Summary

This paper attempts to situate monitoring and evaluation in the wider context of knowledge management, as an element of organizational learning and performance strengthening. It focuses on the case of UNICEF, within the UN system, which may be seen as representative of that of many other agencies working in the field of human and social development. It begins with a brief overview of the knowledge function, and examines the experience of monitoring and evaluation, pointing to strengths as well as gaps in the context of UNICEF.² An example is then presented of a monitoring system linked to research and policy development in the region of Central and Eastern Europe and the Commonwealth of Independent States. In closing, several suggestions are made for the future.

1 The general presentation in this paper rests largely on a variety of unpublished, internal documents which in turn synthesise and apply a wider formal literature on monitoring and evaluation and on knowledge systems. A number of these other works are referenced in other papers in this collection, to which the reader is referred. The principles presented here and their particular application to UNICEF have emerged through a series of internal discussions, with key contributions by L.N. Balaji, Sam Bickel, Howard Dale, Richard Morgan, Ross Smith and Ian Thorpe, and by Susan Bissell and Marta Santos Pais of the Innocenti Research Centre (IRC). Eva Jespersen and IRC colleagues kindly provided inputs to the discussion of the MONEE Project, including comments on a draft version of that section. All errors and omissions remain those of the author. The views expressed in this paper do not necessarily reflect the perspectives or opinions of UNICEF.

2 Fuller treatments of the field of monitoring and evaluation are found in a wide formal literature as well as organization-specific documentation, including that referenced in other articles in this collection. See, for example, the UNICEF Monitoring and Evaluation Handbook, "Making a Difference?", and the UNICEF Evaluation Strategy, addressing both monitoring and evaluation. A key development has been the evaluation standards, promoted internationally by professional organizations as well as by agencies including UNICEF; see, for example, Russon, C. and Russon, G., eds., (2005).

Introduction

Monitoring and evaluation have long played a dual role in organizational systems, contributing critically both to knowledge generation and learning, and as tools of performance management and the promotion of accountability. Within UNICEF, a review of the evaluation function since the organization's inception, carried out over two decades ago, noted that "... Monitoring and evaluation in the history of UNICEF have been applied to a broad array of purposes with different emphases at different times, different methods and varying auspices..."

These purposes were noted as including:

- Monitoring programme inputs
- Evaluating programme outputs
- Monitoring fiscal management
- Monitoring programme management
- Helping to strengthen national capacity for data collection, monitoring and evaluation
- Undertaking data collection and statistical analysis
- Monitoring progress in relation to global themes.³

This same combination of aims, actions and responsibilities has generally been observed since that time in the definition and organization of monitoring and evaluation function, at headquarters and in the field. At the global level responsibilities for monitoring and evaluation tends to be located in several different offices, including a dedicated office for evaluation, addressing in different respects the data management, learning and performance assessment functions. In the field it is more customary for monitoring and evaluation to be situated within a single office, often linked to the planning role. This diversity in organizational positioning creates particular challenges to define appropriate technical standards, communication channels and working methods to contribute most effectively to the organizational knowledge function.

Achieving an optimum balance between these roles represents an ongoing challenge. In particular, given the critical importance of the performance-oriented function of monitoring and evaluation,

3 Stein, H.D., (1986), *A Chronicle of Evaluation at UNICEF, 1948-1984*.

this dimension is often well-resourced and strategically positioned within organizations. At the same time, the knowledge-related value-added of monitoring and evaluation is often not fully incorporated into knowledge systems, limiting the benefit that an organization takes from the knowledge that is generated through these means.

The knowledge function⁴

The overall aims of a knowledge system for children include:

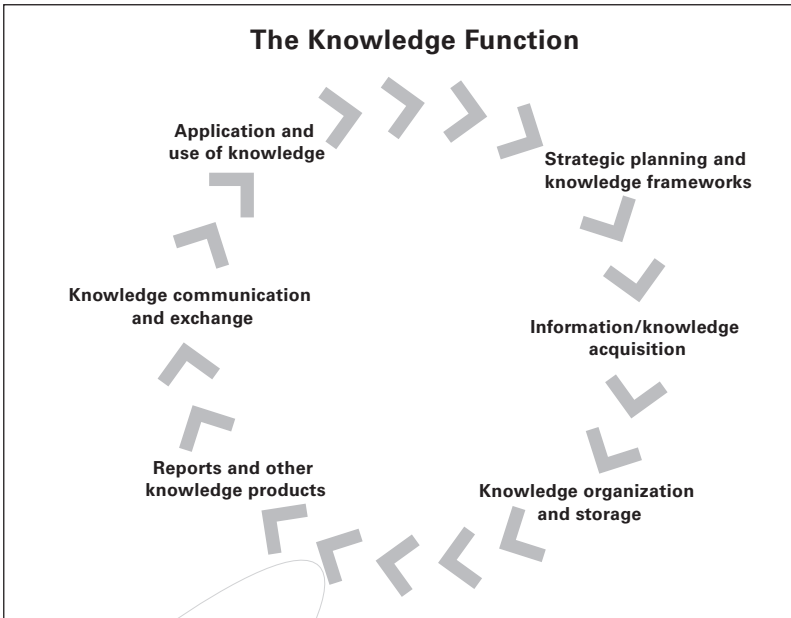
- Generating and refining knowledge for the improvement of children's lives;
- Contributing to improved policies and practices in favour of children at the local, regional and national levels;
- Improving the authority and evidence base to advocate for change and leverage resources for children;
- Learning from experience for more effective programmes;
- Stimulating networking and dialogue among academics and experts, decision makers and practitioners working on children's issues, at the international as well as national levels.

The knowledge function is thus concerned with the acquisition, organization, production, communication and use of knowledge within organizations and beyond their boundaries. Within this, *knowledge management* refers to the management activities supporting all of these steps, which seek to enhance the organization, integration, sharing and delivery of knowledge.⁵

4 The discussion in this section is drawn in large measure from the literature reviews and syntheses of reflections contained in the following unpublished UNICEF papers: Smith, R., (2006), *Improving Knowledge Management in UNICEF – a concept note*; UNICEF, (2006), *Towards an organizational strategy for knowledge acquisition and organizational learning: Annotated outline*, and M. Santos Pais, (2007), *Global knowledge leadership for children*, Innocenti Research Centre.

5 Smith, R., (2006), *Improving Knowledge Management in UNICEF – a concept note*.

The cyclical nature of the knowledge function is depicted in the following diagram:



Briefly to review this framework, a first requirement is for **strategic knowledge planning** to identify knowledge gaps and emerging issues, to prioritize areas in which knowledge is required to be generated or obtained, and to establish frameworks for organizing the knowledge. This step is usually based on reviews of existing knowledge and past experience, alongside a scanning of the environment (over both the short and the long terms), in order to determine priorities. Estimated costs of filling knowledge gaps must also be taken into account. Most development organizations have means for review and planning of monitoring and evaluation activities, because of the resource implications and the long lead times that are needed. Ideally such plans for monitoring and evaluation form part of an overall knowledge assessment, integrating monitoring and evaluation inputs with those from research, policy analysis and other knowledge domains.

6 Adapted from Santos Pais, M., (2007), *Global knowledge leadership for children.*, Innocenti Research Centre.

Some current priorities in the knowledge agenda for children, to which monitoring and evaluation both contribute, include:

- Understanding the child's environment and factors within it affecting child outcomes;
- Assessing the impacts of policies on children;
- Capturing the perspectives of children themselves;
- Understanding factors promoting or constraining social and behavioural change;
- Measuring the impacts of technological innovations and programmatic strategies.

In implementing strategic knowledge plans the principal requirement is for a systematic means of **knowledge acquisition**, drawing from existing sources as well as generating new knowledge. monitoring and evaluation typically plays a major role in acquisition in both these respects, gathering information from a potentially wide range of statistical sources in the area of monitoring, as well as creating new knowledge through surveys, analysis, reviews and field-based evaluations. In the area of monitoring, organizations must typically be of a critical mass to serve as reliable and credible generators of new knowledge, including for the availability (staff or outsourced) of the requisite survey resources and statistical capacity. A large share of monitoring and evaluation activities are carried out in partnership with academic and policy institutions.

The **organization and storage of information** is a crucial management step, to make knowledge accessible for sharing and for analysis. Information, particularly quantitative information, collected through monitoring systems and evaluation exercises is frequently organized within databases and document management systems, for internal use as well as, increasingly, external access. It is somewhat less frequent for data and evaluation findings, especially of a qualitative type, to be integrated within knowledge bases and linked with information from other sources.

The **generation of knowledge products** or outputs takes place in differing ways, but normally involves the production of different types of reports, most of which, in development organizations, rely on monitoring and evaluation data to a large degree. Such products include situation assessments; periodic performance reports; publications of good practices and lessons learned; and, various reports and working papers. Here as well, a challenge exists to integrate

reports on evaluations with information from other sources. Quality assurance mechanisms (such as consultation and advisory group processes, peer review of studies and use of external publication channels), are essential for validation and to maintain credibility.

The **communication and exchange of knowledge** forms a core knowledge function. It is often taken for granted, but there is now increasing attention in many organizations to the importance of a pro-active knowledge communication strategy, and communication plans for different types of knowledge products, involving print as well as electronic media. Audience-specific strategies are frequently needed, particularly to ensure the desired dissemination of monitoring and evaluation-based knowledge to audiences outside the monitoring and evaluation community. The networks of academic, policy and implementing institutions involved in acquiring knowledge become critical to the process of communicating knowledge.

Finally, specific effort is needed to promoting the **application and use of knowledge** by different audiences. Building on the processes of packaging and communication in the preceding phases, information from monitoring systems often serves crucial reference, academic and management support functions. Knowledge from evaluations often needs to be specifically tailored for operational use, e.g., in relation to programme development, scaling up of pilot activities, and adaptation of programme models. Knowledge from both of these sources is widely used in policy advocacy, to argue for improved strategies, laws and resourcing. In all cases, such knowledge has a further use for capacity building, internally and with partners.

Assessment of the experience of dissemination and use of knowledge forms a key input to the adjustment of strategies as the cycle is repeated. A minimal approach to feedback is a review of referencing or citations of reports and articles in the academic and policy areas and in the media. More in-depth, focused studies may also be carried out on the actual utilization of knowledge and its impacts in programming and policy development. Monitoring and evaluation functions may be the subject of specific reviews within organizations.

The MONEE Project: a case study

A monitoring initiative oriented to research and the wider knowledge function is the regional “MONEE” Project - Monitoring Public Policies and Social Conditions in the Central and Eastern Europe and

the Commonwealth of Independent States (CEE/CIS). It was begun in 1992 as a means to monitor the human impact of the political, economic and social transition that has occurred following the fall of the Soviet Union, and the political and economic changes introduced in countries throughout the region. The project is managed by the UNICEF Innocenti Research Centre, working in cooperation with the UNICEF Regional Office for the CEE/CIS. It includes a network of designated focal points from the national statistical offices (NSOs), in the region (some 29 as of 2007), that contribute data and material for the Centre's research and for the public-access TransMONEE database.⁷

The aim of the project has been to track and analyze the situation of children and families, across the region as a whole and for individual countries and sub-regions.

Work to develop the project commenced shortly after the transition began with the break-up of the Soviet Union, and as a result it has been able to generate a consistent, longitudinal base of data that has supported a varied programme of research and analysis. UNICEF country offices have been active partners in the process, contributing perspective, expertise and updated information. The core data collected through the project has been from administrative sources, organized by the NSOs. This has had the strengths of comprehensiveness in relation to the generally wide coverage of administrative services; professional commitment on the part of the participating statistical offices; regular updating; and, broadly consistent definitions and methods applied across the region. Limitations are also those pertaining to administrative data more widely: restricted coverage of some topics and population groups; challenges to validate information against population-based survey information; biases introduced by definitions and methodologies; and, reliance in reporting on national structures which have themselves undergone significant changes during the period.

7 The history and experience of the MONEE Project are documented in a number of sources, most comprehensively in Fajth, G., (2000), *Regional Monitoring of Child and Family Well-Being: UNICEF's MONEE Project in CEE and the CIS in a Comparative Perspective*, Innocenti Working Paper No. 72, UNICEF Innocenti Research Centre, Florence. See also background documentation provided in the 2006 Social Monitor. Regularly updated information is available in the relevant section of the IRC website, www.unicef-irc.org/research. A review of the experience of the project is currently under preparation (Marnie, Sheila, ed., forthcoming 2008), *Fifteen Years of the MONEE Project* (working title), UNICEF Innocenti Research Centre, Florence.

A range of products have been generated by the project since its inception, including:

- Eight Regional Monitoring Reports produced between 1993 and 2001, each addressing a specific theme.
- Innocenti Social Monitors produced in 2002, 2003, 2004 and 2006, succeeding the previous series. The two most recent editions have examined the theme of child poverty, complementing other analysis of this issue carried out by UNICEF and other agencies.⁸
- In-depth studies on specific topics, for example, on children with disabilities in 2005.⁹
- “Country Analytical Reports” produced by the national statistical offices on particular themes each year (e.g., children with disabilities in 2002, supporting the special report on this topic).
- Innocenti Working Papers and other background documents supporting the major publications above.
- Books and publications in professional journals.
- The annually updated TransMONEE regional database, for which the major components have been made publicly available via CD-ROM and the IRC website.
- Examination of selected issues in a series of “features” from the database, separately published in cooperation with the UNICEF Regional Office for CEE/CIS.¹⁰

The TransMONEE database incorporates a wide range of variables which are directly or indirectly relevant to children, on an annual basis from 1989 to the most recent year. It was designed originally around an analytical framework to assess the impacts of socio-economic change; over time it has been expanded to capture a wider range of data specific to children. There are currently over 800 distinct lines of data, organized within eight major categories: (a) demographics; (b) health; (c) education; (d) labour market; (e) social security (retirement; disability, survivorship and occupational injury); (f) family support and child protection; (g) income distribution; and (h) macroeconomics. These data are linked with, and supplemented

8 UNICEF, (2006), *Understanding child poverty in South-Eastern Europe and the CIS.*, Innocenti Social Monitor 2006. UNICEF Innocenti Research Centre, Florence.

9 UNICEF, (2005), *Children and Disability in Transition.*, UNICEF Innocenti Research Centre, Florence.

10 UNICEF, (2006), *TransMONEE 2005*. Innocenti Research Centre, Florence. A 2006/7 edition is forthcoming.

in analyses by, information from other sources including surveys and the databases of other agencies such as the World Bank.

The database is designed to collect data at the national level. Sub-national data were also collected for the Social Monitor 2006. Advances have been made in the content of the database over the years, often linked to the production of Country Analytical Reports. A major recent initiative, for example, has been to expand the scope of the database in the area of social protection.

The network responsible for the TransMONEE database and the research conducted under the MONEE project includes country statistical experts, national and international policy and academics, UNICEF staff at the country, regional and headquarters level, and, indirectly, staff of other international agencies. The core dynamic is the interaction and capacity building which brings together social sector statisticians from different countries in the region, in dialogue with the database manager located at UNICEF Innocenti Research Centre (IRC). The dialogue addresses the availability, definitions and consistency of the data provided, and supports the production of country reports. The presence of a dedicated database manager, drawn from the network, allows the project to release public access data generally within a year of receiving them. Regular use of the data for research purposes at IRC, and the close interaction between researchers and the database manager, further contributes to the quality of the data. Researchers can detect inconsistencies, examine differences vis a vis information from other datasets, identify areas where new data can be sought and support use of the data by academic researchers. This process in turn supports updating of the database template.

The MONEE project has thus been able to develop a strong assembly of data and evidence-based analysis extending over a number of years. On balance, it has been considered successful and influential, highlighting the progress as well as the risks to children and families experienced in the CEE/CIS region, over a period of generally strong economic performance, political development and integration into the global economy and society. In this respect it has Results have included:¹¹

- Contributions to policy discussions on core issues of child poverty and child well-being in the region, keeping children visible in policy debates.

11 These outputs and impacts are examined in detail in Marnie, S., ed., (forthcoming 2008), *Fifteen Years of the MONEE Project*, UNICEF IRC, Florence.

- Provision of new information on issues affecting children which were often, at the time of publication, under-researched and increasing in significance such as HIV/AIDS, migration, and inter-country adoption.
- Inputs to analysis of the situation of children at country and regional level, including through public use of the database.
- Contributions to regional policy reviews and events such as the biennial series of Europe-wide conferences as follow-up to the UN Special Session on Children, which have been held in Berlin, Sarajevo and Palencia, Spain, respectively.
- Inputs to methodological development, for example in the case of differences observed between official statistics and survey estimates of infant mortality.
- Inputs to UNICEF programming in the region. This use however, has been limited by the regional as opposed to country focus of the MONEE project, and the length of the time frame required for research.
- Outside UNICEF, data and products of the project have been used in reviews, studies and sector reports of the World Bank and other agencies; by academics; and by countries themselves for strengthening their own statistical systems.

Through its support of analysis and, of national statistical offices, the MONEE project has played a role in the growth of resources and technical capacity devoted to social development and children's issues in the region, in countries themselves and on the part of UNICEF and other cooperating organizations. The Innocenti Social Monitors, for example, may be seen as having contributed, within the region, to the attention being given to the issue of children living in poverty, including to a policy emphasis on this issue through the UNICEF regional office. The 2004 report addressed "Economic growth and child poverty", while the theme of the 2006 report was "Understanding child poverty in South-Eastern Europe and the CIS". Within the region UNICEF is now undertaking a series of more in-depth country-level studies on child poverty and measures to address it.

A related initiative in recent years concerns the integration and streamlining of the TransMONEE database with other data frameworks. As described above, the database was developed as a means to monitor the situation of children using information from national administrative statistics. As additional information became

available from surveys and a variety of other sources, involving a range of different partners, the challenge emerged of bringing these data together and reconciling information on common topics. A major step in this regard has been to bring the TransMONEE database into the regional DevInfo framework, as implemented by the UNICEF Regional Office for CEE/CIS. This is discussed in other articles in this collection. This effort has had further benefits, for example, of enabling the TransMONEE data to be linked more clearly to the Millennium Development Goals (see www.regionalmdg.org), and facilitating the inclusion of sub-national data.

With regard to the elements of the knowledge function identified above, the MONEE Project may be seen to reflect the following major strengths and limitations:

1. Strategic significance has been maintained through regular consultation and planning with key stakeholders at the national and international levels, within and outside UNICEF.
2. The project has developed an effective means for the acquisition of particular types of information, namely annual, national-level data collected through national statistical offices over the past 15 years. Information has been collected from additional sources to supplement the core database, in the context of specific research needs.
3. The organization and storage of information was, for much of its life, carried out through means specific to the project, particularly through the publicly available TransMONEE database. More recently the use of additional storage mechanisms, such as DevInfo, has proved useful for even wider access (see www.moneeinfo.org).
4. The project has generated several different series of products of high quality. Reports are tailored to different audiences including policymakers, academics, development organizations and the media.
5. There has been a continuous focus on the communication and exchange of knowledge through the project, with national correspondents, within UNICEF and externally.
6. Evidence suggests that there has been good use of project findings at the international level, and to some extent the national level, in the policy arena as well as academically.

Some key factors in this positive experience include:

- *Continuity* in the monitoring framework utilized, the analytical processes and in the network of correspondents;
- *Flexibility* in adapting to changing conditions within the region and to shifting policy priorities in relation to children;
- *Technical credibility* maintained through quality assurance of data, peer review of reports and formal publication processes;
- Focus has been maintained on *capacity building* for monitoring and analysis within the region as well as on knowledge generation;
- In view of the limited financial resources available, the project has moved forward on the basis of *partnerships*, internally within UNICEF and with national statistical offices and other organizations.

Linking knowledge and action

Ensuring that knowledge required is available, and that it is then applied to solving problems, has become an increasing focus of many organizations working in the development field. In the health sector, for example, both 'supply' and 'demand' factors may be seen to influence this dynamic. Recent growth in political mobilization and funding, including from philanthropic sources, coupled with technical developments in such areas as vaccines and laboratory testing, have substantially increased the potential stock of knowledge applicable to health services in the developing world. Such knowledge may derive from a range of sources including basic laboratory research, population surveys and monitoring, intervention field trials and programme evaluations. This supply-side growth has been accompanied by increased demands for knowledge: by national partners, as well as funders and implementing agencies. Organizations in a position to promote and broker knowledge in these different dimensions are substantially increasing their support for systems to generate, strategically present and, disseminate knowledge as a core component of their development activities.¹²

12 See, for example, UNICEF, (2006), *Medium Term Strategic Plan, 2006-2009*, and case studies such as Kerkhoff L. van., and Slezak, N., (2006), *Linking local knowledge with global action: Examining the Global Fund to Fight AIDS, Tuberculosis and Malaria through a knowledge system lens*, and Weber, J., (2004), *An Evaluation of USAID's Evaluation Function, unpublished report, Bureau for Program and Policy Coordination*. Prominent among many web-based resources disseminating reports of monitoring and evaluation work within overall knowledge systems for development work are the Development Gateway portal (www.developmentgateway.org) and the ID21 service (www.id21.org).

In this context, several directions may be identified for enhancing the usefulness of knowledge from monitoring and evaluation sources within organizational knowledge systems:¹³

1. Adopting procedures for **critical review and quality assessment** so that users of monitoring and evaluation knowledge can readily determine:

- relevance of information to different issue areas;
- soundness of the underlying programme or experience from which information is derived;
- validity or applicability of lessons or conclusions drawn;
- accessibility of the underlying data and reference materials.

2. Increased use of **pilot approaches** to raise validity and credibility of knowledge generated through monitoring and evaluation. Pilot projects, intended to test the effectiveness and efficiency of a proposed intervention strategy or programme model, typically involve a higher than usual standard of monitoring and evaluation including baseline measurement, periodic assessments, use of control or comparison groups, and structured variations in the intervention model. The potential yield is considerably richer and more reliable knowledge than from regular development projects. Intervention models must of course be realistically designed, and adaptable to different settings. An important recent example has been the piloting of interventions for Accelerated Child Survival and Development (ACSD), focusing on West Africa.

3. **Strengthened platforms** for the organization, presentation and communication of knowledge including from monitoring and evaluation sources. Electronic media in particular offer opportunities, which are by no means fully exploited, for organizing information into relevant categories, and packaging and communicating it for different types of audiences. Information in the monitoring and evaluation area is increasingly well catalogued and made available to wider audiences. However, opportunities remain to integrate this knowledge, particularly that of a qualitative nature, with information from other sources and to promote meta-analysis of datasets and evaluation findings. New formats for sharing and learning from

13 Further discussion of these issues is found in the sources cited in note 4 above; see in particular UNICEF, (2006), *Towards an organizational strategy for knowledge acquisition and organizational learning: Annotated outline*.

knowledge continue to be developed and refined, calling for investment in their testing and use.¹⁴

4. Finally, in the context of increased volume of development knowledge and expanded channels for its communication, **knowledge impact evaluation** is needed in order to better understand the effectiveness of the diffusion of knowledge, and of its use within and outside organizations.

Recognizing the critical roles played by monitoring and evaluation, these and related initiatives will contribute to a fuller realization of their strategic potential within organizational knowledge systems for the benefit of human development.

References

Fajth, G., (2000), *Regional Monitoring of Child and Family Well-Being: UNICEF's MONEE Project in CEE and the CIS in a Comparative Perspective*, Innocenti Working Paper No. 72, UNICEF Innocenti Research Centre, Florence, 2000.

Kerkhoff, L. van, and Szlezak, N., (2006), *Linking local knowledge with global action: Examining the Global Fund to Fight AIDS, Tuberculosis and Malaria through a knowledge system lens*. Bulletin of the World Health Organization, August 2006, 84(8), pp. 629-635.

Marnie, S. ed., (forthcoming 2008), *Fifteen Years of the MONEE Project* (working title), UNICEF Innocenti Research Centre, Florence.

Ramalingam, B., (2006), *Tools for Knowledge and Learning: A Guide for Development and Humanitarian Organizations*, Oversea Development Institute, London.

Russon, C. and Russon, G., eds. (2005). *International Perspectives on Evaluation Standards*. Special issue of *New Directions for Evaluation*, No.104, February 2005.

Santos Pais, M., (2007), *Global knowledge leadership for children*, Innocenti Research Centre, April 2007.

Smith, R., (2006), *Improving Knowledge Management in UNICEF – a concept note*, unpublished report, Division of Policy and Planning, June 7, 2006, p.1.

Stein, H.D., (1986) *Chronicle of Evaluation at UNICEF, 1948-1984*. UNICEF History Project, March 1986, pp. 7-8.

UNICEF, (1991), *Monitoring and Evaluation Handbook, Making a Difference?*

UNICEF, (2005), *Evaluation Strategy*.

UNICEF, (2005), *Children and Disability in Transition*, UNICEF Innocenti Research Centre, Florence.

UNICEF, (2006), *Medium Term Strategic Plan, 2006-2009*.

14 See, for example, Ramalingam, B., (2006), *Tools for Knowledge and Learning: A Guide for Development and Humanitarian Organisations*.

UNICEF, (2006), *Towards an organizational strategy for knowledge acquisition and organizational learning: Annotated outline*. Unpublished paper, Division of Policy and Planning.

UNICEF, (2006), *TransMONEE 2005*. Innocenti Research Centre, Florence.

UNICEF, (2006), *Understanding child poverty in South-Eastern Europe and the CIS*, Innocenti Social Monitor 2006. Innocenti Research Centre, Florence.

Weber, J. M., (2004), *An Evaluation of USAID's Evaluation Function*. Unpublished report, Bureau for Program and Policy Coordination, USAID, September 2004.

HELPING COUNTRIES BUILD GOVERNMENT MONITORING AND EVALUATION SYSTEMS: WORLD BANK CONTRIBUTION TO EVIDENCE-BASED POLICY MAKING

by Keith Mackay¹, Coordinator, Evaluation Capacity Development,
Independent Evaluation Group, World Bank

More and more governments are working to improve their performance by creating monitoring and evaluation systems to measure, and to help them understand, their performance. These efforts rest on the simple proposition that it is better to have more information than less, in order to support and help guide policy making and management, and to underpin accountability relationships. Thus monitoring and evaluation information can improve the quality of government decisions and the effectiveness of government. A stronger version of this proposition is that monitoring and evaluation is *necessary* to achieve evidence-based policy making, evidence-based management, and evidence-based accountability (Mackay, 2007).

Policy making, especially budget decision making and national planning, focus on government priorities among competing demands from citizens and groups in society. Monitoring and evaluation information can support government's deliberations by providing evidence about the most cost-effective types of government activity, such as different types of employment programmes, health interventions, or conditional cash transfer payments.

Monitoring and evaluation also helps government ministries and agencies manage activities at the sector, programme, and project levels. This includes government service delivery and the management of staff. Monitoring and evaluation identifies the most efficient use of available resources. It can, for example, be used to identify implementation difficulties. 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 help identify good, bad, and promising practices, and this can prompt a

1 The helpful comments of Marco Segone on an earlier draft are gratefully acknowledged.

search for the reasons for this performance. Evaluations or reviews are used to identify these reasons. This is the learning function of monitoring and evaluation.

Finally, monitoring and evaluation enhances transparency and supports accountability relationships by revealing the extent to which government has attained its desired objectives. Monitoring and evaluation provides the essential evidence necessary to underpin strong accountability relationships, such as the accountability of government to the Parliament or Congress, to civil society, and to donors. Monitoring and evaluation also supports the accountability relationships within government, such as between sector ministries and central ministries, among agencies and sector ministries, and between ministers, managers, and staff.

The context

There are many reasons for the increasing efforts of governments to strengthen or create monitoring and evaluation systems, including:

- Fiscal pressures;
- Ever-rising expectations from ordinary citizens that governments should provide more services and at a higher quality;
- Accountability pressures from citizens and parliaments that governments should publicly report and explain their performance;
- The demonstration effect of developed countries: members of the Organisation for Economic Cooperation and Development: which already place considerable emphasis on monitoring and evaluation and monitoring and evaluation systems;
- The encouragement of donors to place higher emphasis on measuring and managing for results: the *results agenda*;² and
- Pressures on donors to demonstrate the results of their own large volumes of aid spending.

2 <http://www.mfdr.org/>

World Bank support for Poverty Reduction Strategies

The World Bank and other donors are now devoting considerable efforts to help countries measure their performance. This support is focused on poorer countries, in particular, those preparing poverty-reduction strategies as part of debt-relief initiatives. These countries are working to achieve the Millennium Development Goals (MDGs) and other policy objectives. The policy making dimension of these strategies places a premium on being able to measure the extent of country success in poverty-reduction efforts. This in turn usually entails a focus on performance indicators which measure progress vis-à-vis the MDGs and other development priorities (Independent Evaluation Group, 2002).

Donor support often focuses on the capacities of national statistical offices and their statistical systems. In practice this has included assistance such as conducting population censuses and household surveys. Most of these poor countries in Africa and other regions have however, found it rather difficult to strengthen their monitoring systems, both in terms of data production and especially in terms of data utilization (World Bank and International Monetary Fund, 2004; Bedi and others, 2006).

Poverty reduction strategies usually focus on the amount of budget and donor resources spent on national development priorities and country progress against the MDGs. These two issues are certainly important, but what is absent from this focus is what Booth and Lucas (2001a, 2001b) have termed the “missing middle”: performance indicators on the intervening steps in the results chain, involving government activities, outputs and services provided and their outcomes; and in-depth evaluative evidence linking government actions to actual results on the ground.

This missing middle relates to the government’s own performance in terms of the results of government spending. That is, the outputs, outcomes and impacts of the government itself. MDGs provide a bottom-line measure of country performance but do not reveal the contributions of the government compared with donors, the private sector and civil society groups such as NGOs. Due to lack of funds, skills and demand, poorer countries tend to rely on donors to conduct evaluations and reviews.

The move towards greater use of programmatic donor lending to governments offers one way to facilitate joint monitoring and

evaluation work by governments and donors. It reduces the scope of project-specific monitoring and evaluation and thus the scope for fragmented donor monitoring and evaluation. In Uganda, for example, the World Bank and other donors provide programmatic budget support to the government (Hauge, 2003). Such support is becoming increasingly common in African and other debt-relief countries. This type of lending requires a focus on “big picture” results or outcomes of development assistance. It also requires a greater reliance on country systems for national statistics and for monitoring and evaluation of government programmes.

World Bank support for government monitoring and evaluation systems

Major donors such as the World Bank have a substantial tool-kit of assistance to help countries in their efforts to create or strengthen their government monitoring and evaluation systems. This includes loans; grants; technical assistance including training; resource materials on good practice; and, support for communities of practice concerned with monitoring and evaluation and *results*.

In the Latin American region alone, the World Bank is currently working with over 12 governments to help them strengthen their monitoring and evaluation systems. This region appears to be at the forefront within the developing world. A particular reason for this appears to be the leading example of advanced countries such as Chile, Colombia and Mexico, which already possess strong national or sectoral monitoring and evaluation systems (May and others, 2006). Even these leading countries are making efforts to further strengthen their systems, both on the technical side, in terms of the quality of performance information and of evaluations, and in particular, on the demand side in terms of achieving more intensive utilization of the monitoring and evaluation information produced by their systems.

World Bank support includes loans to strengthen a government’s public sector management or to strengthen the work of sector ministries. Some of these are designed to build a whole-of-government monitoring and evaluation system, focusing on performance indicators, rapid evaluations and impact evaluations. Some have a somewhat narrower, sectoral focus. Many also support the conduct of rigorous impact evaluations as one component of sectoral activities, particularly in the health, education and social security sectors.

A separate type of Bank loan also supports the strengthening of national statistical systems.

The World Bank has funded a small number of grants for governments working to strengthen their monitoring and evaluation systems. Demand for these grants is high, and competition for them is keen. The amount of these grants varies, but is often in the range of \$0.3m to \$0.5m. They fund activities such as diagnoses of a country's existing monitoring and evaluation systems, provision of monitoring and evaluation training and seminars to raise awareness among senior officials and, sometimes, the conduct of individual evaluations. Some grants fund statistical capacity-building initiatives. A diverse range of countries have benefited from these grants in recent years, such as Brazil, China, Guatemala, Mexico, the Philippines, Uganda and a number of other African countries. In Eastern Europe and Central Asia the countries which have received these grants include Georgia, Kyrgyz Republic, Poland, Turkey and Uzbekistan.

Monitoring and evaluation training is also provided by the World Bank. This training includes one-week introductory courses in monitoring and evaluation, and advanced workshops in impact evaluation offered by the World Bank Institute. The Bank's Independent Evaluation Group also offers its highly-successful 4-week residential course, the International Programme for Development Evaluation Training (IPDET) each summer,³ as well as week-long introductory courses in monitoring and evaluation.⁴ The summer course has had 1,300 participants since it was first offered in 2001, and IPDET alumni are members of a community of practice which stays in touch via an email list-serve.

The World Bank has published a wide range of material on how to build government monitoring and evaluation systems, including country case studies and diagnoses; conference proceedings on this topic; examples of influential evaluations; handbooks on evaluation and performance indicators; and, other guidance material (many of these are listed by Mackay, 2007).

One final area of Bank support is in the growing number of communities of practice. The Bank, and especially the Independent Evaluation Group, has provided financial and other support to a global evaluation association (the International Development

3 <http://www.ipdet.org/>

4 <http://www.ipdet.org/>

Evaluation Association, IDEAS), regional evaluation associations such as the African Evaluation Association (AfrEA) and the Latin America and Caribbean network (ReLAC), and to national evaluation associations in Kenya and Uganda, among others. World Bank support has been provided jointly with donors such as UNICEF; the African Development Bank; the Development Bank of Southern Africa; and, with bilateral donors such as the Netherlands, Norway and the United Kingdom. In partnership with the Inter-American Development Bank, the World Bank has supported the creation of a community of practice among senior officials and evaluation offices in Latin America. Annual conferences are being held for this community (see May and others, 2006).

Collaborative work with other donors

Of course, the World Bank works closely with other development partners, as exemplified by the *Managing for Results* initiative. The 2002 Monterrey conference was a major milestone. Here, the international community agreed that it was important to provide more funding for development, but that more money alone was not enough. Donors and partner countries wanted to be sure that aid is being used effectively; they also wanted to be able to demonstrate the results of that aid. This has led to three subsequent international roundtables: in Marrakech (2004); Paris (2005); and, Hanoi (2007), on managing for development results. It has also led to explicit commitments to support country efforts in this area. These commitments center around specific actions to increase country ownership, harmonization, alignment, managing for results, and mutual accountability for the use of aid.⁵ This work has also led to the preparation of a large volume of resource material in this area, encapsulated in the *Sourcebook on Emerging Good Practice in Managing for Development Results*.⁶ This sourcebook provides a number of country case studies of government monitoring and evaluation systems (at the national, sector, programme and project levels), and of monitoring and evaluation in civil society and the private sector.

World Bank lessons from experience

Many lessons can be drawn from the Bank's efforts to help governments build monitoring and evaluation systems to support bet-

5 See <http://www.aidharmonization.org/>

6 <http://www.mfdr.org/Sourcebook/2-2ndEdition.html>

ter government (see Mackay, 2007 for a much fuller discussion of these issues). The foremost lesson is that **substantive demand from the government is a prerequisite to successful institutionalization**. That is, an monitoring and evaluation system must produce monitoring information and evaluation findings which are judged valuable by key stakeholders, that are used to improve government performance, and that respond to a sufficient demand for the monitoring and evaluation function to ensure its funding and its sustainability for the foreseeable future. For many countries, the reality is an absence of real demand for monitoring and evaluation information, and this can seem to be an insuperable barrier. However, this is not the case. There are steps that can be taken to raise awareness about the potential benefits to be derived from monitoring and evaluation, and thus to strengthen demand. And in particular, there are incentives which a government can put in place to conduct monitoring and evaluation and to use monitoring information and evaluation findings.

These incentives fall into three broad categories: carrots, sticks and sermons (Mackay, 2007). Examples of carrots include: conducting “How Are We Doing?” team meetings to brainstorm ways to achieve objectives and improve performance; awards or prizes to executives who succeed in managing for results; and, staff incentives such as recruitment or promotion to those who are able to demonstrate that they use monitoring and evaluation information in their jobs. Sticks include publicly highlighting bad performance as revealed by monitoring and evaluation information, thus embarrassing poor performers; setting performance (“stretch”) targets; and requiring managers who fail to meet their targets to prepare “exception reports”. Sermons include strong, repeated statements from senior managers or ministers supporting monitoring and evaluation; awareness-raising seminars; and, the highlighting of examples of influential evaluations.

An additional dimension to the demand side is having a powerful champion. That is, a powerful minister or senior official who is able to lead the push to institutionalize monitoring and evaluation, to persuade colleagues about its priority, and to devote significant resources to create a whole-of-government monitoring and evaluation system. Such a champion would usually require close support from a capable ministry that can design, develop, and manage an monitoring and evaluation system. In some countries this has been the finance ministry; in others, the president’s office has taken the lead.

A second key lesson is to **avoid the common danger of over-engineering a monitoring and evaluation system**. This has happened by having an excessive number of performance indicators, or by having multiple and uncoordinated monitoring and evaluation systems (e.g. Uganda). Excessive data collection is not only a waste of scarce monitoring resources; it also can lead to a situation where data providers have little incentive to provide accurate data because they know the information will not be used. Clearly, it is necessary to build reliable ministry data systems to provide the raw data on which government-wide monitoring and evaluation systems often depend. An audit of data systems and a diagnosis of data capacities can be helpful in this situation.

National statistical offices have the capacity to play an important role here, because of their expertise in conducting surveys and censuses, and in data management. Unfortunately, there do not appear to be many examples of national statistical offices in developing countries helping sector ministries and agencies to strengthen their administrative data systems, to collect better data on programme delivery, or on beneficiary satisfaction with government services, or to help them use this information in evaluating programme performance.

Over-engineering an monitoring and evaluation system would reflect a supply-driven view of monitoring and evaluation. This views monitoring and evaluation information as having inherent value such that it should be collected for its own sake. However a cogent argument can be made that the real measure of “success” of an monitoring and evaluation system is not whether it is producing reliable monitoring information and evaluation findings. Rather, it is the extent of utilization of this information. Monitoring and evaluation practitioners thus need to avoid a mindset where they regard monitoring and evaluation information as being inherently valuable. It is hard to persuade a skeptical finance ministry that it should continue to fund an monitoring and evaluation system whose outputs are not being utilized.

A third lesson is that **the structural arrangements of a monitoring and evaluation system are important**. These include arrangements for data verification and auditing. They also include whether evaluations are to be conducted internally within government, or contracted out to academia and consultants; however, there can be a tension between ministry ownership of evaluation findings, and the need for evaluations to be objective and credible.

Related to this is the role of central ministries and of sectoral ministries and agencies in monitoring and evaluation. An overly-centralized approach can dissuade sector ministries from making use of monitoring and evaluation information which has been mandated by central ministries such as the finance or planning ministries.

A fourth major lesson is the **importance of conducting a diagnosis of existing monitoring and evaluation functions**. It is important to understand the strengths and weaknesses of the monitoring and evaluation currently being conducted, on both the demand and supply sides. Such diagnoses are themselves a form of evaluation, and they are useful not just for the information and insights they provide, but also because they can be a vehicle for raising the shared awareness of stakeholders in government, civil society, and the donor community about the importance of monitoring and evaluation and the need to build a new system or strengthen existing systems.

A diagnosis also provides a baseline against which future efforts to institutionalize monitoring and evaluation can be compared. Most countries which have built well-performing monitoring and evaluation systems have had to develop them in a flexible, opportunistic manner, as new opportunities (such as civil service reforms or the introduction of performance budgeting) have emerged, and as roadblocks have developed (such as the departure of a key monitoring and evaluation champion). Moreover, countries which have successfully built monitoring and evaluation systems (such as Chile, Colombia, Australia and the United States, among others), have found that it is a long-haul effort requiring patience and persistence. 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 monitoring and evaluation information among ministries; and to train staff to use monitoring and evaluation information in their day-to-day work.

This experience provides a strong argument for the regular monitoring and evaluation of the monitoring and evaluation system itself, with the unsurprising objective of finding out what is working well, 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 actual extent of utilization of monitoring and evaluation information, as well as the particular ways in which it is being used.

Conclusions

More and more governments in developing countries are coming to understand that sound systems for monitoring and evaluation can help them improve their performance. There are a small but growing number of governments which have succeeded in building monitoring and evaluation systems in support of evidence-based policy making, evidence-based management, and evidence-based accountability. The World Bank and other international donors view this as a priority area, and stand ready to help developing countries strengthen their work in this area.

References

- Bedi, Tara, Aline Coudouel, Marcus Cox, Markus Goldstein, and Nigel Thornton. (2006), *Beyond the Numbers: Understanding the Institutions for Monitoring Poverty Reduction Strategies*. Washington, DC: World Bank.
- Booth, David, and Henry Lucas. (2001a.), *Desk Study of Good Practice in the Development of PRSP Indicators and Monitoring Systems: Initial Review of PRSP Documentation*. London: Overseas Development Institute.
- Booth, David, and Henry Lucas. (2001b.), *Desk Study of Good Practice in the Development of PRSP Indicators and Monitoring Systems: Final Report*. London: Overseas Development Institute.
- Hauge, Arild. (2003), *The Development of Monitoring and Evaluation Capacities to Improve Government Performance in Uganda*. ECD Working Paper No. 10, World Bank, Washington, DC. <http://www.worldbank.org/ieg/ecdf/>.
- Independent Evaluation Group. (2002), *Annual Report on Evaluation Capacity Development*. Washington, DC: World Bank. <http://www.worldbank.org/ieg/ecdf/>.
- Mackay, Keith. (2007). *How to Build monitoring and evaluation Systems to Support Better Government*. Washington, DC: World Bank. <http://www.worldbank.org/ieg/ecdf/>.
- 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*. Washington, DC: World Bank and Inter-American Development Bank. <http://www.worldbank.org/ieg/ecdf/>.
- OECD and DAC. (2007), *Managing for Development Results—Principles in Action: Sourcebook on Emerging Good Practices. Second edition*. Paris: OECD. <http://www.mfdr.org/Sourcebook.html>.
- World Bank and International Monetary Fund. (2004), *Poverty Reduction Strategy Papers—Progress in Implementation*. Washington, DC: World Bank and International Monetary Fund. http://siteresources.worldbank.org/INTPRSP/Resources/prsp_progress_2004.pdf.

TEN STEP TO A RESULTS BASED MONITORING AND EVALUATION SYSTEM

by Jody Zall Kusek, Chief, Global HIV/AIDS Monitoring and Evaluation Group, World Bank¹ and Ray Rist, Advisor, Public Sector Management², World Bank

Summary

An effective state is essential to achieving sustainable socioeconomic development. With the advent of globalization, there are significant pressures on governments and organizations around the world to be more responsive to the demands of internal and external stakeholders for good governance; accountability and transparency; greater development effectiveness; and, delivery of tangible results. Among the stakeholders interested in ensuring that funds used achieve the desired results are governments; parliaments; citizens; the private sector; non-governmental organizations (NGOs); civil society; international organizations; and, donors. As the demands for greater accountability have increased there is an attendant need for enhanced results-based monitoring and evaluation of policies, programmes and projects.

Monitoring and evaluation is a powerful public management tool that can be used to improve the way governments and organizations achieve results. Just as governments need financial, human resource, and audit systems, governments also need good performance feedback systems.

Over recent years, there has been an evolution in the field of monitoring and evaluation involving a movement away from traditional implementation-based approaches toward new results-based approaches. The latter help to answer the “so what” question. In other words, governments and organizations may successfully implement programmes or policies but the key question is, have these programmes or policies produced the actual, intended

-
- 1 The views expressed here are solely those of this author and no endorsement by the World Bank Group is intended or should be inferred.
 - 2 Ray C. Rist is an independent advisor to governments and organizations throughout the world.

results? Have government and organizations delivered on promises made to their stakeholders? For example, it is not enough to simply implement health programmes and assume that successful implementation is equivalent to actual improvements in public health. One must also examine what outcomes and impacts were achieved. The introduction of a results-based monitoring and evaluation system takes decision-makers one step further in assessing whether and how goals are being achieved over time. These systems help to answer the all important “so what” questions, and respond to stakeholders’ growing demands for results.

This paper provides a summary of the book *Ten Steps to a Results-Based Monitoring and Evaluation System* by Jody Zall Kusek and Ray C. Rist. This book, published in 2004 is currently in its fourth printing and has been translated into five languages. It is being used in government offices, NGO’s, and universities across the world, to introduce concepts and principles of results based monitoring and evaluation. We, the authors, have agreed to summarize its contents to support this UNICEF document on the role of monitoring and evaluation in evidence-based policy making.

The Ten Steps model addresses the challenge of how governments in general, but those in developing countries in particular, can begin to build results-based monitoring and evaluation systems so as to provide credible and trustworthy information for their own use and to share with their citizens. The reality is that putting in place even a rudimentary system of monitoring, evaluating, and reporting on government performance is not easy in the best of circumstances. The obstacles for developing countries are greater and more formidable, even as they build experience in constructing more traditional monitoring and evaluation systems. These more traditional systems typically are used to assess the progress and track the implementation of government projects, programmes, and policies.

It should also be acknowledged that it is not a new phenomenon for governments to monitor and evaluate their own performance. For this reason, a theoretical distinction needs to be drawn between traditional monitoring and evaluation and results-based monitoring and evaluation. Traditional monitoring and evaluation focuses on the monitoring and evaluation of inputs, activities, and outputs, i.e., project or programme implementation. Governments have over time tracked their expenditures and revenues, staffing levels and resources, programme and project activities, numbers of participants, goods and services produced, etc. Indeed, traditional efforts

at monitoring and evaluation have been a function of many governments for many decades or longer. In fact, there is evidence that the ancient Egyptians (3000 BC) regularly employed traditional monitoring as they tracked their government's outputs in grain and livestock production. (Egyptian Museum, Cairo, Egypt)

Results-based monitoring and evaluation, however, combines the traditional approach of monitoring implementation with the assessment of results (Mayne and Zapico-Goni, 1997.) It is this linking of both implementation progress with progress in achieving the desired objectives or goals (results) of government policies and programmes that make results-based monitoring and evaluation most useful as a tool for public management. Implementing this type of monitoring and evaluation system allows the organization to modify and make adjustments to its theories of change and logic models as well as its implementation processes in order to more directly support the achievement of desired objectives and outcomes

Why build a results-based monitoring and evaluation system anyway?

A results-based monitoring and evaluation system can help policy-makers answer the fundamental questions of whether promises were kept and outcomes achieved. If governments are promising to achieve improvements in policy areas such as in health care or education, there needs to be some means of demonstrating that such improvements have or have not occurred, i.e., there is a need for measurement. However, the issue is not measurement per se. There is a general need both to document and demonstrate government's own performance to its stakeholders as well as to use the performance information to continuously improve. As Binnendijk (1999) observed:

One key use is for transparent reporting to external stakeholder audiences, on performance and results achieved. In many cases, government-wide legislation or executive orders have recently mandated such reporting. Moreover, such reporting can be useful in competition for funds by convincing a skeptical public or legislature that the agency's programmes produce significant results and provide "value for money." Annual performance reports are often directed to ministers, parliament, stakeholders, customers, and the general public.

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

If information on results achieved is the key, then where does it come from?

Results information can come, essentially, from two sources: a monitoring system and an evaluation system. Both are needed, but they are not the same. The distinction between monitoring and evaluation is made here both for conceptual and practical purposes.

Monitoring can be viewed as periodically measuring progress toward explicit short, intermediate, and long-term results. It also can provide feedback on the progress made (or not), to decision-makers who can use the information in various ways to improve the effectiveness of government.

Monitoring involves measurement: what is measured is the progress towards achieving an objective or outcome (result). However, the outcome can not be measured directly. It must first be translated into a set of indicators that, when regularly measured, will provide information whether or not the outcome is being achieved. For example: If country X selects the outcome of improving the health of children by reducing childhood morbidity by 30% over the next five years, it must now identify a set of indicators that translate childhood morbidity into more specific measurements. Indicators that can help assess the changes in childhood morbidity might include: 1) the incidence of diseases, such as malaria; 2) the level of maternal health; and 3) the degree to which children have access to iodine in water supplies. Measuring a disaggregated set of indicators provides important information as to how well government programmes and policies are working to support the overall outcome. If, for example, malaria incident rates are found to be rising, this indicator can be further disaggregated to measure how many children are sleeping under impregnated bed nets. The government can use this information to step up programmes aimed to educate parents about the importance of keeping children safe from mosquito exposure while they are sleeping.

Understanding the importance of information about whether one's government is keeping promises made or achieving results that are important for various users is a central reason for building a monitoring system in the first place. Key users in many societies who are often left out of the information flow are citizens, NGO groups, and the private sector. The point being that monitoring data have both internal (governmental) and external uses (societal). It is important to note here that information obtained from a monitoring system only reveals information about what is being measured at that time. It can, however, be compared against both past performance and some planned level of present or anticipated performance. Monitoring data do not reveal why that level of performance occurred or provide the likely causality to changes in performance from one reporting period to another. This information comes from an evaluation system.

An evaluation system serves as a complimentary but distinct function to that of a monitoring system within a results management framework. Building an evaluation system allows for a more in-depth study of why results (outcomes and impacts) were achieved, or not; can bring in other data sources, in addition to those indicators already in use; can address factors which are too difficult or too expensive to continuously monitor; and, perhaps most important, can tackle the issue of why and how the trends being tracked with monitoring data are moving in the directions they are. Such data, on impacts and causal attribution, are not to be taken lightly and can play an important role in an organization making strategic resource allocations. Some performance issues, such as long-term impacts, attribution, cost-effectiveness, and sustainability, are better addressed by evaluation than by routine performance monitoring reports.

An additional point to make in this regard is that an monitoring and evaluation system can be designed for, and applicable to, the project level, the programme level, the sector level, and the country level. The specific indicators may necessarily be different (as the stakeholders' needs for information will also be different at each level), the complexity of collecting the data will be different, the political sensitivity on collecting the data may change, and the uses of the information may change from one level to another.

In the end, it is the creation of a system that is aligned from one level to the others that is most critical. In this way information can flow up and down in a governmental system rather than it being

collected at only one level or another, stored and used at that level, but never shared across levels. *Blocking the information from being shared ensures that the linkages between policies, programmes, and projects stay disconnected and uncoordinated.* At each level, performance information is necessary and there should be the means to collect it. While different levels will have different requirements which need to be understood and respected, the creation of an monitoring and evaluation system requires interdependency, alignment, and coordination across levels.

The ten steps to building a performance based monitoring and evaluation system

There is no consensus on how many steps necessarily go into building an monitoring and evaluation system. Holzer (1999) proposes seven steps; an American NGO (The United Way, 1996) proposes eight steps; and Sharp (2001) proposes a model with four areas for measuring performance to provide the data for monitoring and evaluation. We have described elsewhere (Kusek and Rist, 2004) a ten-step approach that we have been using in working with a number of developing countries as they each design and construct their monitoring and evaluation system. We have opted for ten steps (rather than fewer) for the reason that it is important when building such a system to provide sufficient differentiation among tasks. There are so many challenges in building such a system that reducing the ambiguity as to the sequence and activities required at each step can only help.

It is not the intent here to discuss in detail the ten steps, as a more thorough discussion of this is done within the book, itself. Suffice it to also say that while we have labelled each of the following as a "step," we are not implying that there is a rigid sequencing here that allows for no concurrent activities. There are a number of areas where there is the need for concurrent activity that can span over steps and over time. The selection of the word "steps" is more to suggest a focus on discrete components in building an monitoring and evaluation system, some of which are sequential and essential before you move on to others. (Indeed, the last section of this paper will return to Step One for a more detailed discussion on how one begins this process.)

Step One: Conducting a readiness assessment is the means of determining the capacity and willingness of the government and its development partners to construct a results-based monitoring and evaluation system. This assessment addresses such issues as the presence or absence of champions in the government, the barriers to building a system, who will own the system, and who will be the resistors to the system.

Step Two: Agreeing on outcomes to monitor and evaluate addresses the key requirement of developing strategic outcomes that then focus and drive the resource allocation and activities of the government and its development partners. These outcomes should be derived from the strategic priorities (goals) of the country.

Step Three: developing key indicators to monitor outcomes is the means of assessing the degree to which the outcomes are being achieved. Indicator development is a core activity in building an monitoring and evaluation system and drives all subsequent data collection, analysis, and reporting. Both the political and methodological issues in creating credible and appropriate indicators are not to be underestimated.

Step Four: Gathering baseline data on indicators stresses that the measurement of progress (or not) towards outcomes begins with the description and measurement of initial conditions being addressed by the outcomes. Collecting baseline data means, essentially, to take the first measurements of the indicators.

Step Five: Planning for improvements: setting realistic targets recognizes that most outcomes are long term, complex, and not quickly achieved. Thus there is a need to establish interim targets that specify how much progress towards an outcome is to be achieved, in what time frame, and with what level of resource allocation. Measuring results against these targets can involve both direct and proxy indicators as well as the use of both quantitative and qualitative data.

Step Six: Monitoring for results becomes the administrative and institutional task of establishing data collection, analysis, and reporting guidelines; designating who will be responsible for which activities; establishing means of quality control; establishing time-lines and costs; working through the roles and responsibilities of the government, the other development partners, and civil society; and establishing guidelines on the transparency and dissemination of the information and analysis. It is stressed that the construction of an

monitoring and evaluation system needs to clearly address the challenges of ownership, management, maintenance, and credibility.

Step Seven: Evaluative information to support decision-making focuses on the contributions that evaluation studies and analyses can make throughout this process to assessing results and movement towards outcomes. Analysis of programme theory, evaluability assessments, process evaluations, outcome and impact evaluations, and evaluation syntheses are but five of the strategies discussed that can be employed in evaluating a results-based monitoring and evaluation system.

Step Eight: Analyzing and reporting findings is a crucial step in this process, as it determines what findings are reported to whom, in what format, and at what intervals. This step has to address the existing capacity for producing such information as it focuses on the methodological dimensions of accumulating, assessing, and preparing analyses and reports.

Step Nine: Using the findings emphasizes that the crux of the system is not in simply generating results-based information, but in getting that information to the appropriate users in the system in a timely fashion so that they can take it into account (as they choose) in the management of the government or organization. This step also addresses the roles of the development partners and civil society in using the information to strengthen accountability, transparency, and resource allocation procedures.

Step Ten: Sustaining the monitoring and evaluation System within Government recognizes the long term process involved in ensuring longevity and utility. There are six key criteria that are seen to be crucial to the construction of a sustainable system: demand, structure, trustworthy and credible information, accountability, incentives, and capacity. Each of these dimensions needs continued attention over time to ensure the viability of the system.

As noted earlier, there is no orthodoxy that the building of an monitoring and evaluation system has to be done according to these ten steps. One can posit strategies which are more detailed in the number of steps as well as those with fewer numbers (four of which we cited earlier.) The issue is one of ensuring that key strategies and activities are recognized, clustered together in a logical manner, and then done in an appropriate sequence.

Transition and developing countries have notable challenges

The challenges for transition and developing countries in following our “Ten-Step” model or any other model are many. First, in Step Two, it is assumed that governments are likely to undertake a process whereby there will be an agreement on national or sector-wide outcomes. Although developed countries typically undertake a strategic (usually 5-10 years) or a medium-term (3-5 years) plan to guide their government priorities, developing countries can find it difficult to do so. This difficulty may stem from a lack of political will, a weak central agency (such as the Ministry of Planning or Ministry of Finance, or a lack of capacity in planning and analysis. Thus, we continue to emphasize in Step Six that it is important to make sure that traditional implementation focused monitoring and evaluation gets done. That is, tracking budget and resource expenditures, making sure that funded activities and programmes actually occur, and that promised outputs (number of wells dug, miles of road constructed, youth completing a vocational programme, etc.) all exist. The People’s Republic of China represents an interesting example where efforts are being made in this area, especially with their large infrastructure projects (Rist, 2000.) There is no way to move to a results based monitoring and evaluation system without the foundation of a basic traditional monitoring and evaluation system.

To paraphrase from Louis Carroll’s Alice in Wonderland: “It is hard to know where to go if you do not know where you are.” Thus, in Step Four, we describe how the absence of information on the current conditions (baselines) directly hinders policy and resource planning of how best to address something, because it is only weakly documented. The statistical systems in developed countries can give rather precise figures on the numbers of children in rural areas, the number of new HIV/AIDS cases in the past twelve months, or the number of disabled adults. However, in developing countries, such information may or may not be available and with widely varying degrees of precision.

Moreover, many developing countries lack the skill-base residing in government agencies to make collection and use of baseline information possible. One significant hurdle to overcome is the likely certainty that few developing countries will have significant capacity in the workforce to develop, support, and sustain these systems. Typically, few government officials will have been trained in modern data collection and monitoring methods. Further, still fewer will

have been trained in how to interpret different modalities of data analysis. The challenge for development agencies, for international NGOs interested in governance issues, and for in-country universities and research institutes is to provide the needed technical support and training given the rapid turnover in staff, the competing priorities, and the need to rebuild political support and commitment as each new political administration comes into office.

This challenge has been particularly noted in many of the most heavily-indebted countries for whom borrowing from the international community is crucial and for whom subsequent relief from this debt essential. The World Bank and the International Monetary Fund (IMF), do allow for debt relief if these countries can demonstrate a serious commitment towards reform, particularly reforms to promote poverty reduction as outlined in the Poverty Reduction Strategy Papers (PRSPs). One condition for granting debt relief is a demonstrated ability of the country to adequately monitor, evaluate, and report on the reforms proposed.

A second challenge for developing countries is that the governments themselves are often only loosely inter-connected, lack strong administrative cultures, and function without the discipline of transparent financial systems. This results in those in government being uncertain of actual levels of resource allocation, and, of whether the allocation goes where it is intended, and, when it arrives, if it is used as intended to achieve the desired results. Measuring if governments have achieved desired results in such an environment can become an approximation at best. In some countries, for example, the budget process is one where final budget approval into law does not occur until mid-way through, or near the end of, the budget year. Agencies can thus spend well into the budget year without an actual approved budget. This makes it very difficult to introduce a fiscal discipline that includes any concern whether programmes are achieving their intended results.

Third, and based on the two above noted constraints, the construction of a results-based system is hindered when there is no means to link results achieved to a public expenditure framework or strategy. Keeping results information separate from the resource allocation process ensures that budget allocation decisions do not consider past results achieved (or not) by line agencies. Linking the budget process to the knowledge coming from the monitoring and evaluation system begins the process of allocating resources to strategic objectives and targets. If no such link is made, then the

budget process can be supporting project and programme failures just as readily as it is funding those that are successful.

A number of countries still operate with two budget systems, one for recurring expenses and one for capital expenses. Egypt is one such country. The Ministry of Planning has primary responsibility for the capital budget and the Ministry of Finance has responsibility for the recurrent budget. In Egypt, the Minister of Finance is very interested in using the budget process to catalyze the government to focus on improving the performance of the government's programmes and policies. He hopes that implementing a performance-based monitoring and evaluation system will help achieve this. However, with the current situation of two budget systems, he faces a difficult challenge unless he and the Minister of Planning can together ensure that both budgets are used to achieve the Government of Egypt's objectives.

Back to the beginning: first things first

We turn now in this last section to the very beginning of building an monitoring and evaluation system, that is, conducting a readiness assessment. This first step is often overlooked by system designers and we believe merits special emphasis here. Understanding the complexities and subtleties of the country or sector context is critical to the ultimate success or failure in introducing and using an monitoring and evaluation system. (cf. Kusek and Rist, 2001.)

Furthermore, the needs of the end users are often only partly understood by those ready to start the system building process. For all the good intentions to advance the use of monitoring and evaluation information in the public sector, there has been, from our point of view, too little attention given to organizational, political, and cultural factors.

The obvious question is "why?" The answer lies in the lack of sufficient attention to understanding the influence of these organizational, political, and cultural factors on whether the country is "ready" to move to one that is intent on measuring the performance of government programmes and policies. Thus we believe that the first step in the design of a results-based monitoring and evaluation system should be to determine the "readiness" of a government to design and use such a system. If one reads through the literature on building such a system, regardless of the number of steps, the presumption time and again is that, like a runner getting ready to

begin a race, the designer comes up to the starting line, hears the gun, and starts building the system.

Why is it important to begin with a readiness assessment?³

Our experience in conducting readiness assessments prior to assisting in the building of an monitoring and evaluation system points to one fundamental fact: conducting the readiness assessment is like constructing the foundation for a building. It is below ground, not seen, but critical.

A readiness assessment allows those building an monitoring and evaluation system to assess a wide range of contextual factors before any design work. The published literature in this area is rather sparse, but there are several key sources available that stress the importance of studying current organizational capacity in designing an monitoring and evaluation system (Boyle and Lemaire, 1999; Guerrero, 1999; and Mayne, 1997). In this same arena, there are also several diagnostic guides or checklists related to the construction of a strategy for evaluation capacity development (Mackay, 1999). These guides and checklists are often, appropriately, rather technical in nature as they have to provide information on the quality and quantity of available statistical data, on capacity for data analysis and reporting, and on the capacity of the government to generate new data collection procedures. These guides do tend to focus more on the nature of the “supply” of information and data than on the “demand” side. Assessing such supply-side capacity is a necessary part of the design process, but not enough.

The key question we continually stress in this regard is “Why collect the information in the first place?”⁴ Supplying information should be the response to a need, not an end it itself. Thus the demand for information to monitor and evaluate public policy and programmes

3 The use of the term “readiness assessment” here is deliberately used in contrast to the term “needs assessment.” We are of the view that it is no longer a question of whether a government ought to collect and report information on its own performance (i.e., does it need such information), but rather whether it has sufficient institutional capacity and political will to do so (is it ready to initiate such a system?).

4 We are also here framing the issue differently than that of proposing an “evaluability assessment” (cf. Smith, 1989; Wholey, 1987.) The issue here is not to see if the logic and specification of a project or program is sufficiently clear that an evaluation design could be constructed prior to the initiation of the project or program, but to inquire as to whether the particular level or levels of government are in a position to begin collecting, analyzing, and reporting on performance-based monitoring and evaluation data in a continuous fashion so as to inform the decision-making process. In this sense, the emphasis here is less on the program theory of a policy or program than on the operational capacity of a government to initiate such a new function.

comes as the answer to whether desired outcomes and results are, or are not, being achieved. Building an monitoring and evaluation system that provides such information is a profoundly political event. It is not just a set of technical procedures. Thus there are a number of other factors which must also be studied in building the foundation of a sustainable monitoring and evaluation system.

Addressing the following seven questions is, from our experience, critical in ensuring that the key organizational, political, and social dimensions are directly addressed before any monitoring and evaluation system construction begins:

1) What is driving the need for a monitoring and evaluation system within the public sector?

Where the demand for such a system originates, and why, are essential factors in creating a successful and sustainable system. There are internal political and organizational pressures as well as potential external factors for building an monitoring and evaluation system. These pressures need to be acknowledged and addressed if the response is to be appropriate to the demand. Internal demand can come from efforts to push reform in the public sector, for instance, fighting corruption, strengthening the role of the parliament, and expanding the authority of the Auditor General. It can also come internally from political parties in opposition to the sitting government.

External pressures can come from the international aid community, which has been pressing for stronger tracking of the consequences and impacts of its development interventions. They also come, for example, from such international organizations as the European Union and the criteria it is setting for the accession countries or from Transparency International, a global NGO that addresses issues of public sector corruption (cf. Furubo, Rist, and Snadahl, 2002). Still other pressures can come from the new rules of the game which are emerging with globalization, where financial capital and the private sector want a stable investment climate, the rule of law, and the protection of their property and patents before they will commit to investing in a country. The role that external organizations can play in generating pressures for a country to move towards an monitoring and evaluation system should not be under-estimated.

2) Who is driving the need for a monitoring and evaluation system within the organization?

Champions in government are critical to the success and stability of an monitoring and evaluation system. A champion highly placed in the government can give a strong voice to the need for better informed decision-making, and can help

diffuse the attacks of the counter-reformers who have vested interests in not seeing such a system constructed. But if the champion is away from the center of policy making and has little influence with key decision-makers, it will be so much more difficult for an monitoring and evaluation system in these circumstances to take hold. Box 1 below provides a summary of the readiness assessment undertaken in Egypt and the location of a champion (Minister of Finance) to carry the effort forward.

It should also be noted that while the presence of a champion is so important, it is also important to work towards the institutionalization of the monitoring and evaluation system with legislation, regulation, or decree. The need in the end is to not have a system that is personalized or based on charisma, but on the structured requirements in the government to produce quality information.

Box 1: The case of Egypt: slow, systematic moves toward monitoring and evaluation

One of the most important components of assessing a country's readiness to introduce results-based monitoring and evaluation is whether a champion can be found who is willing to take on ownership of the system. Conducting a readiness assessment uncovered significant interest in Egypt on the part of many senior government officials for moving toward a climate of assessing whether or not results were being achieved, or not. The President himself has called for better information to support economic decision-making.

The Minister of Finance was found to be a key champion for the government of Egypt's move to a results focus. This Minister was well versed in the international experience of other countries, such as Malaysia and OECD member countries. The minister underscored the importance of giving increased attention to improving the management of public expenditures by moving forward with a set of pilots to demonstrate how results-based monitoring and evaluation could be used to better manage budgetary allocations. The Minister of Finance will play a key leadership role in any effort to introduce results-based monitoring and evaluation in Egypt.

A number of other senior officials were identified who could play important roles. The First Lady of Egypt, who chairs the National Council for Women, is developing a system to monitor and evaluate efforts across many ministries to enhance the status and condition of women. However, for an monitoring and evaluation effort to be successful and sustainable, there must be a "buy-in" (or a sense of ownership) from line ministers who are responsible for resource expenditures and overseeing the implementation of specific programmes. The team found interest in monitoring and evaluation for results on the part of several line ministers, including the Minister of Electricity and Energy, and the Minister of Health.

The readiness assessment also revealed a high level of capacity in Egypt to support the move toward a results-based strategy. A number of individuals with evaluation training were identified at the University of Cairo, the American University of Cairo, and private research organizations. In addition, the Central Agency for Public Mobilization and Statistics, and the Cabinet Information Decision Support Center have key roles in collecting, analyzing, and disseminating data to be used by both government and non-government researchers and policy-makers.

A key criterion for a successful shift toward results is the development of a well-communicated and executable strategy. The diagnostic identified a fragmented strategy for moving the effort forward. A set of pilots had tentatively been identified, yet there were few, if any, criteria for establishing these as performance pilots. Nor was there a management structure set up within the government to effectively manage the overall effort. The Minister of Finance, however, had begun to define an approach that, if implemented, would provide the necessary leadership to move the effort forward. The minister was definite in his desire to move slowly and to nurture the pilots, learning along the way.

The results of this readiness assessment suggest that the government of Egypt is prepared to take ownership of the effort and to systematically and slowly begin to introduce the concepts of results management. Visible capacity exists that can be drawn upon to sustain the effort. Significantly, there is obvious political support to provide the necessary leadership. (The complete Egypt Readiness Assessment can be found in annex II of the Ten Steps Book).

3) What is motivating the champion? To build an monitoring and evaluation system is to take a real political risk. Producing information in a government on performance and strengthening the basis for accountability are not neutral activities. So, consequently, the question has to be posed as to the political benefits to the champion and to his/her institution in order to be willing to take these risks. One cluster of benefits can come from responding to the pressures. Doing something is better than doing nothing and so letting the pressures mount still further. Another set of benefits can come from being perceived as a reformer in the government. Being seen as a reformer could be a source of political capital. Third, there are benefits in being on the right side of the issue with the international aid community. The calls for reform, for accountability, and demonstrated evidence of impacts are all being made by the aid community. Showing responsiveness to these pressures is not without its benefits. Finally, the champion may be one who is instilled with a sense of public responsibility and who considers that taking on this challenge is important and not something they can walk away from.

4) Who will own the system? Who will benefit? How much information is really required?

For an monitoring and evaluation system to be used, it should be accessible, understandable, and relevant. These criteria drive a need for a careful readiness assessment prior to designing the system. This assessment would consider ownership of, benefits to, and utility for the relevant stakeholders. Further, while these issues are on the demand side, there is a whole spectrum of issues on the supply side to be addressed as well. These are: capacity to collect and analyze data; capacity to produce reports; capacity to manage and maintain the monitoring and evaluation system; capacity to use the information that is produced; etc.

The implications for those who will design the system are that complexity and over designing are constant dangers. There will also be constant erosion in the system that has to be addressed and stakeholders may want to pull the system in too many different directions at once. In addition, little in the political arena will stay the same for long. Such an assessment will also provide important information and baseline data against which necessary capacity building activities can be built into the system. Having said all this, there is still the absolute requirement to collect no more information than is essential. We have found time and again that monitoring and evaluation systems are designed which are immediately in overload. Too many data are being collected too often, and with not enough thought on how or whether they will be used.

5) How will the system directly support better resource allocation and the achievement of programme goals?

Monitoring and evaluation is not an end unto itself. Monitoring and evaluation it is a tool to promote modern management practices and better accountability. The idea in creating such a system is to support innovation, reform, and better governance. This is done by producing useful information that is also transparent, trustworthy, and relevant. It is also our view that treating the creation of an monitoring and evaluation system as a discrete event, unconnected to other public sector and public administration reform efforts, or to efforts at creating a medium term public expenditure framework, or to restructuring of the administrative culture of the government, is not sustainable. In fact, it is quite the contrary. Linking the creation of the monitoring and evaluation system to precisely such initiatives creates interdependencies and reinforcements that are seemingly crucial to the sustainability of the system. The issue for the readiness assessment is whether such linkages are both structurally and politically possible.

6) How will the organization, the champions, and the staff all react to negative or potentially detrimental information generated by the monitoring and evaluation system? It is hard for an monitoring and evaluation system to function in an organization or political climate where there is a great deal of fear and corruption. In such conditions, it is inevitable that an monitoring and evaluation system will at some point produce (even if infrequently) data that can be embarrassing, politically sensitive, or detrimental to those who exercise power. The information can also be detrimental to units and individuals in the organization that have produced the information. Going after the messenger is not an unknown event in organizations. If it is clear from the readiness assessment that only politically popular and “correct” information will be allowed to come from the system, then the system is compromised from the beginning. It will not be seen to be credible by those who are outside the system or by others inside the system. Rather, it will be understood to be a hollow exercise.

In such a setting, building the system carefully, beginning slowly, and trying to find units which will risk the generation of potentially detrimental information about their own performance is perhaps the best that can be achieved.

Consequently, it is good to understand the barriers and obstacles within the organization, whether these are cultural, structural, political, or individual. Not all barriers can be addressed simultaneously in the design of the system. However, not recognizing their presence, not picking the most critical and strategic ones to tackle first, and not taking some initiative to address them is to ensure a level of resistance greater, longer, and more tenacious than would have been necessary otherwise.

7) How will the monitoring and evaluation system link, even in a rudimentary fashion, the project outcomes to the programme outcomes and to sector and national outcomes? It is a key task of the readiness assessment to identify opportunities for, and barriers against, linking information in a vertical and aligned fashion inside the government. In an ideal situation, the project level performance data would feed into, and be linked to, assessment of programmes, which, in turn, would be linked to assessments of sectoral, regional, and eventually national outcomes and targets. Results-based information at any level that is not linked vertically to the information needs at the next level is not useful beyond the restricted information needs at that same level. Choking-off the

flow of information between levels is to ensure that performance-based decisions cannot be made where one level informs decisions at the next. It is also relevant in this context to ascertain if there is a commitment, in the collection and analysis of data, to ensure that there are no levels where data are collected, but not used or shared with persons at that same level. Stated differently, can the system address the need at every level to be both producers and consumers of results-based information.

The objective is to build a system that allows relevant questions to be asked and answered at the appropriate levels. Such a system should be able to take components of information from any level to meet the information needs at the same or any other level. Breaks in that system (much as a chain where links are missing) renders the entire initiative less useful.

Postscript

Building an monitoring and evaluation system is easier said than done. Otherwise, we would see these systems as an integral part of good public management practices in governments and there would not be the need to consider this issue. But the reality is otherwise. There are few such systems (in whole or in part) fully integrated into the public management strategies of developed countries, and still fewer in developing countries. And it is not that governments are not trying, many are. It is just that creating such a system takes time, resources, stability in the political environment, and champions who do not become faint of heart.

This takes us to the significant challenge of sustainability. Indeed, governments willing to use results-based information to assist in the governance of the political system and frame public policies give evidence of some level of democracy and openness. But even in these countries, there is often a reluctance to measure and monitor for fear that the process will present bad news to the leadership and other stakeholders. Presenting one's performance shortfalls to others is not typical bureaucratic behavior. Thus the efforts to build such a system should recognize the inherent and real political limitations. These efforts should start with a simple approach, work with stakeholders (to help them recognize it is their right to be regularly informed on the performance of their government), and continue to stress time and again that information can help improve policy making and public management. The achievement of these modest goals should then be a reason for longer term optimism.

References

- Binnendijk, A. (1999), *Results Based Management*, Prepared for the International Conference on Evaluation Capacity Development, Beijing, China.
- Boyle, R. and D. Lemaire (Eds.). (1999), *Building Effective Evaluation Capacity*. Transaction Books, New Brunswick, N.J.
- Furubo, J., Rist, R., and R. Sandahl (Eds.). (2002), *International Atlas of Evaluation*. Transaction Books, New Brunswick, N.J.
- Guerrero, P. (1999), *Comparative Insights from Colombia, China, and Indonesia*, in Boyle, R. and D. Lemaire, eds., *Building Effective Evaluation Capacity*. Transaction Books, New Brunswick, N.J.
- Holzer, M. (1999), *Public Performance Evaluation and Improvement*, Prepared for the International Conference on Evaluation Capacity Development, Beijing, China.
- Kusek, J.Z. and R.C. Rist. (2001), Making monitoring and evaluation Matter—Get the Foundation Right, *Evaluation Insights*, Vol. 2, no. 2.
- Kusek, J.Z. and R. C. Rist, (2004), *Ten Steps to a Results-Based Monitoring and Evaluation System*, The World Bank, Washington, D.C.
- Mackay, K. (1999), *Evaluation Capacity Development: A Diagnostic Guide and Action Framework*, The World Bank, Washington, D.C.
- Mayne, J. and E. Zopic-Goni (Eds.). (1997), *Monitoring Performance in the Public Sector*. Transaction Books, New Brunswick, N.J.
- Mayne, J. (1997), *Accountability for Program Performance: A Key to Effective Performance Monitoring and Reporting*, in Mayne, J. and E. Zopic-Goni, eds. *Monitoring Performance in the Public Sector*. Transaction Books, New Brunswick, N.J.
- Rist, R.C. (2000), *Evaluation Capacity Development in the People's Republic of China: Trends and Prospects*, in Malik, K. and C. Roth, eds., *Evaluation Capacity Development in Asia*. United Nations Development Program Evaluation Office, New York, New York.
- Sharp, C., (2001), *Strategic Evaluation: Performance Measurement in the Service of Effective Governance*.
- Smith, M.F. (1989), *Evaluability Assessment: A Practical Approach*. Kluwer Academic Publishers, Boston, Mass.
- United Way. (1996), *Measuring Program Outcomes: A Practical Approach*. Alexandria, Virginia.
- Wholey, J.S. (1987), *Evaluability Assessment: Developing Program Theory*, in Bickman, I., ed., *Using Program Theory in Evaluation*. New Directions for Program Evaluation, no. 33. Jossey-Bass, San Francisco, California.





Part 2

The strategic intent of the evaluation function

Enhancing the utilization of evaluations for evidence-based policy making, by Michael Bamberger, former Senior Evaluator, World Bank	120
Country-Led Evaluation, by Marie-Helene Adrien, President, IDEAS and Dennis Jobin, Vice President, IDEAS.....	143
Joint Country-led Evaluation of the policies related to Child well-being within the Social protection sector in Bosnia and Herzegovina, by Azzedina Vukovic, Directorate for Economic Planning, Council of Ministers of Bosnia & Herzegovina, and Debora McWhinney, Deputy Representative, UNICEF Bosnia & Herzegovina	154

ENHANCING THE UTILIZATION OF EVALUATIONS FOR EVIDENCE-BASED POLICY MAKING¹

by Michael Bamberger,
Former Senior Evaluator, World Bank

Concerns about under-utilization of evaluations

There is widespread concern that, despite the significant resources devoted to programme evaluation and its importance in both industrialized and developing countries, the utilization of evaluation findings is disappointingly low (Patton, 1997, chapter 1). This holds true even for evaluations which are methodologically sound. In 1995, the U.S. General Accounting Office (GAO), conducted follow-up case studies on three major federal programme evaluations: the *Comprehensive Child Development Programme*, the *Community Health Centers Programme*, and, *Title 1 Elementary and Secondary Education Act* aimed at providing compensatory education services to low-income students and found that:

“Lack of information does not appear to be the main problem. Rather, the problem seems to be that available information is not organized and communicated effectively. Much of the available information did not reach the [appropriate Senate] Committee, or reached it in a form that was too highly aggregated to be useful or that was difficult to digest.” (GAO, 1995, p. 39).

The GAO’s report helped to explain why “the recent literature is unanimous in announcing the general failure of evaluation to affect decision-making in a significant way” (Wholey, 1970, p. 46), and confirmed that “producing data is one thing, getting it used is quite

1 This chapter draws upon the following publications: Operations Evaluation Department 2004 (Michael Bamberger editor) *Influential Evaluations*; Operations Evaluation Department 2005 (Michael Bamberger *Influential Evaluations: Detailed Case Studies*); Bamberger, Rugh and Mabry 2006 *RealWorld Evaluation Chapter 8*; and on several workshops organized by Michael Bamberger on Evaluation Utilization for the Independent Program Evaluation Network, Tbilisi 2006; the World Bank Independent Evaluation Group, Washington DC 2006; and the International Program for Development Evaluation Training Ottawa 2006, 2007.

another” (House, 1972, p. 412)². Evaluators are also concerned about the related issue of *misuse* of evaluation findings. House (1990), observed, “Results from poorly conceived studies have frequently been given wide publicity, and findings from good studies have been improperly used” (p. 26). In some cases the misuse might be intentional, but in other cases it results from a lack of understanding of how to interpret, and use, evaluation findings.

Regarding evaluations of development programmes, the World Bank Independent Evaluation Group (IEG)³ recently concluded that “for all development agencies, monitoring and evaluation remains the weakest link in the risk management chain.” (SIDA 1999). The Swedish International Development Agency (SIDA), in a recent assessment of its evaluation practices was disappointed to find that most stakeholders never even saw the findings and that few of those who did found nothing very new or useful. SIDA concluded that “for the majority of stakeholders the evaluation could just as well have been left undone.” A former Director General of the World Bank Operations Evaluation Department observed that, the “prerequisite of credibility is missing in the evaluation systems used by most governments, companies and development agencies” (Picciotto, 2002, p. 14).

Defining evaluation utilization

When assessing evaluation utilization it is important to define clearly what is being assessed and measured. For example, are we assessing:

- *Evaluation use*: how evaluation findings and recommendations are used by policy-makers, managers and others.
- *Evaluation influence*: how the evaluation has influenced decisions and actions.
- *The consequences of the evaluation*: how the process of conducting the evaluation, as well as the findings and recommendations, affected the agencies involved, the policy dialogue and the target populations. It is becoming realized that the decision to conduct an evaluation, and the choice of evaluation methodology, can in

2 Several of the examples and citations in this paragraph are taken from Patton, 1997, chapter 1.

3 At the time the two Influential Evaluations reports were published, the Department was called the Operations Evaluation Department (OED) but the name has now been changed to the Independent Evaluation Group (IOG).

themselves have important consequences. For example, the decision to use randomized control trials, or strong quasi-experimental designs, can affect the programme being evaluated. For example, the use of randomization means that selected beneficiaries will include both poorer and better-off families from the target population. However, if beneficiaries were selected by the implementing agency, it is possible that preference would have been given to the poorest families.

Outcomes and impacts can also be assessed at different levels:

- changes at the level of the individual (e.g. changes in knowledge, attitudes or behavior);
- changes in organizational behavior;
- changes in how programmes are designed or implemented, or in the mechanisms used to draw lessons from programme experience;
- changes in policies and in planning procedures at the national, sector or programme level.

Measurement issues

There are a number of measurement issues which must be addressed in the assessment of evaluation use, influence or consequences:

- *The time horizon over which outcomes and impacts are measured:* due to pressures from policy-makers and funding agencies, the evaluator will often be required to make an assessment of outcomes and impacts at a relatively early stage in the programme implementation cycle. This will often result in the conclusion that the intended outcomes and impacts have not been achieved when in fact it was still too early to be able to measure them. This problem can be addressed if an *evaluability assessment* is conducted prior to the launch of the evaluation⁴. In this case the assessment could show that outcomes and/or impacts could not be measured at the point of time if the evaluation were to be conducted then. The evaluation terms of reference should then be revised either to permit the evaluation to be conducted at a later stage or to limit what is to be assessed (e.g. estimating outputs and preliminary outcomes but not the main outcomes or impacts).

4 The purpose of an evaluability assessment is to determine whether the intended objectives of an evaluation can be achieved with the available resources and data and within the specified evaluation time horizon.

- *Intensity of effects*: Many evaluations have only a small impact or influence which may be of very little practical importance. Consequently it is often important to assess the intensity of the evaluation's contribution.
- *Reporting bias*: many policy-makers, planning or implementing agencies may not be willing to acknowledge that they have been influenced by the evaluation findings and recommendations. The complaint is often heard from policy-makers that funding agencies often claim credit for the choice of policies which were being planned by the government before the evaluation was completed. This makes it even more difficult to estimate the influence of the evaluation when clients do not acknowledge its contribution.
- *Attribution*: policy-makers and planners receive advice from many different sources. This means that determining the extent to which the observed changes in policies or programme design can be attributed to the effects of the evaluation is difficult. The changes could equally be attributed to one of the many other sources of information and recommendations to which policy-makers are exposed.

Reasons why evaluations are under-utilized

Lack of a feeling of ownership, or lack of commitment to an evaluation may be an inevitable result in cases where many stakeholders are not consulted about the objectives or design of the evaluation; are not involved in implementation and, have no opportunity to comment on the findings. In many developing countries, access and use is further limited because relatively few reports are translated into the national language of the country studied. Even fewer are available in the local languages spoken by many stakeholders. Civil society frequently shows its frustration at the lack of involvement in the evaluation process.

The main reasons why evaluation findings are under-utilized are:

- **Lack of ownership**: Clients and stakeholders often feel that the evaluation has been designed by funding agencies and so is addressing questions of interest to these agencies rather than the concerns and priorities of the client. In some cases stakeholders feel that they do not have the right to make suggestions on the evaluation design or content or that any suggestions will often be ignored.

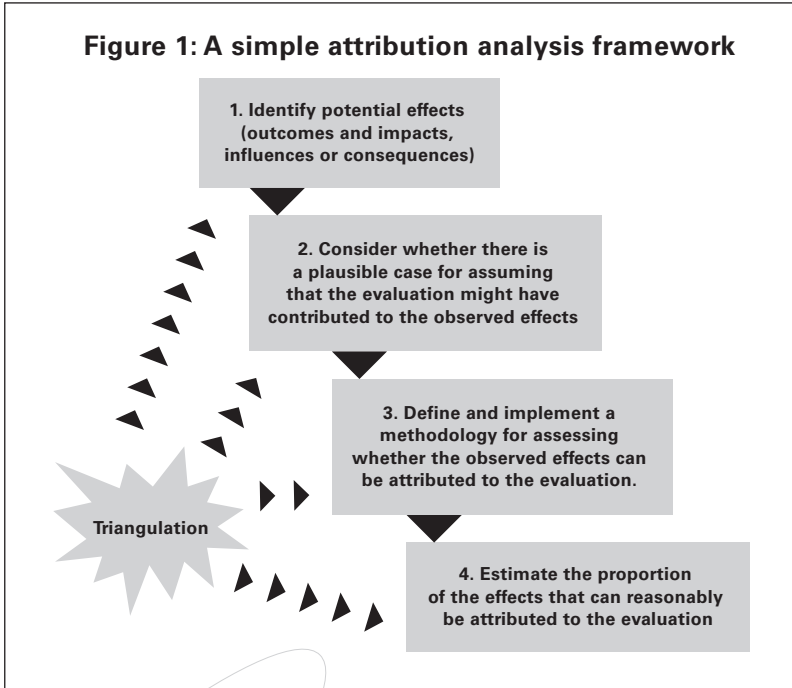
- *Bad timing:* The findings are often not available when they are needed, making them largely irrelevant by the time they are available. At the other end of the project cycle, evaluators often wish to discuss baseline studies and evaluation design at the start of the project which is often when programme management is still struggling to launch the project and when it is far too early for them to have any interest in thinking about what the results will be several years on.
- *Lack of flexibility and responsiveness to the information needs of key stakeholders:* Evaluations are normally conducted according to their own design logic and frequently cannot respond to the immediate information needs and deadlines of the stakeholders.
- *Wrong question and irrelevant findings:* Many evaluations do not ask the questions of concern to stakeholders and mainly provide information on topics of little interest to them.
- *Weak methodology:* The complexities of attributing causality for complex programmes operating in a context with many other actors and exogenous factors, combined with time and budget pressures, and a lack of comparative data, frequently makes it impossible to produce very precise and conclusive evaluation findings.
- *Many evaluations are expensive and make too many demands on overtaxed programme staff:* Even many potential supporters of an evaluation may complain that the exercise requires more resources in terms of funds, staff time and effort than they feel are justified in terms of the new insights the evaluation will produce.
- *Lack of local expertise to conduct, review and use evaluations:* The lack of familiarity with evaluation methods on the part of client agencies and local consultants also limits the utilization of a potentially useful evaluation.
- *Communication problems:* Clients may not be kept informed of progress and initial findings, there may be language problems, or clients may not understand or like the evaluator's presentation style (it may be too technical, too qualitative etc). In other cases the problem may be that the client does not wish to share reports with other stakeholders. Sometimes, in order to ensure "objectivity" of the findings the evaluator may believe that the progress of the evaluation should not be discussed with the client.

- *Factors external to the evaluation*: evaluations are conducted within government or civil society organizational structures. If these are not conducive to the evaluation this may affect implementation or utilization. For example; the country may not have an effective monitoring and evaluation system so that it is very difficult to collect data or disseminate findings, or the government may not accept the funding agency's "evaluation culture" (for example frank identification of problems and wide dissemination and open discussion of the evaluation findings) .

Assessing the influence of an evaluation (attribution analysis)

Figure 1 shows the approach to attribution analysis used to assess the influence of the evaluations included in the World Bank Operations Evaluation Department (OED, 2004, 2005), *Influential Evaluations* publication. This study reviewed evaluations which had already been completed and was conducted on a limited budget. Therefore, it was only possible to conduct a sample survey of stakeholders in one of the eight cases and, extended telephone interviews in one other case. Most of the analysis was based on desk reviews, peer reviews, consultation with the evaluators and e-mail contact with stakeholders. The approach involved four steps:

1. Identify potential *effects* (outcomes, impacts, influences or consequences) of an evaluation. These may be identified by reviewing the terms of reference, reading the evaluation report or, consulting with stakeholders and the team who conducted the evaluation.
2. Consider whether there is a plausible case for assuming that some of the observed effects might be attributable to the evaluation
3. Define and implement a methodology for assessing whether the effects can in fact be attributed to the evaluation.
4. Estimate the *proportion* of the effects that can reasonably be attributed to the evaluation.

Figure 1: A simple attribution analysis framework

Triangulation is used at all stages to compare the consistency of estimates obtained from different sources.

The *Influential Evaluations* report included the following simple and cost-effective methods for assessing attribution:

- Comparing user surveys at two points in time (Bangalore: *Citizen Report Card evaluation*).
- Opinion survey of stakeholders (Bangalore: *Citizen Report Card evaluation*).
- Testimonials from stakeholders on ways in which they were influenced by the evaluation (Indonesia: *Village Water Supply and Sanitation Project* and Bulgaria: *Metallurgical Industry*).
- Expert Assessment and peer review (India: *Employment Assurance Scheme*).
- Following a paper trail (India: *Employment Assurance Scheme*; Uganda: *Improving the Delivery of Primary Education*).
- Logical deduction from secondary sources (Uganda: *Improving the Delivery of Primary Education*).

Examples of effective evaluation utilization

The cases cited in this section are taken from *Influential Evaluations* (World Bank OED 2004 and 2005). Lessons are also drawn from the author's own experience

How are evaluations used?

Evaluations are used, or achieve their influence, in many different ways, not all of which were intended by the evaluator and/or the client. The following are some of the common ways in which evaluations are used:

- *Evaluation is never the only source of advice:* clients receive information and advice from many different sources and the impact of an evaluation will often depend on how well it complements other sources of information.
- *Political cover:* providing political cover or support for difficult or politically sensitive decisions or actions.
- *The credibility of the evaluator:* the credibility and perceived independence of the evaluator can be an important factor when politically sensitive or controversial decisions have to be taken.
- *Identifying winners and losers:* policy-oriented evaluations not only provide findings and advice but can also help identify potential winners and losers and can suggest ways to mitigate negative consequences for losers. Whilst many evaluations emphasize the benefits of programmes or policies, for politicians and policy-makers, it is often equally important to address the concerns of influential groups who may suffer as a result of the programmes.
- *Presenting the big picture:* evaluations can help clients to see their programme within a broader perspective and to understand how the programme is affected by the economic, political and organizational context within which it operates.
- *Understanding the impact on vulnerable groups:* implementing agencies and policy-makers tend to hear from and be influenced by the better-organized and better-off sectors of the target population. An evaluation can make policy-makers aware of the special issues, and of the lack of access often facing the vulnerable and less vocal groups.
- *Providing new knowledge or understanding:* when agencies are receiving funding from new sources (for example the European

Union or a regional development bank), the evaluator may provide important new information on the policies and procedures of these agencies. In the case of the EU, for example, countries may be subjected to new environmental or social policies with which they may not be familiar. Similarly, when an agency is developing new kinds of programmes (for example, community driven development or post-conflict reconstruction), the evaluation team may provide guidance on unfamiliar programme design or implementation methodologies.

- *Catalytic function*: the planning and design of an evaluation, or the review and consultation process, may help bring together people or groups who previously had not worked together. The follow-up action plan may lead to the creation of working groups or consultative mechanisms which institutionalize contacts between these groups.

Types of influence that evaluations can have

Table 1 uses the case studies in the *Influential Evaluations* publication to illustrate the many different types of influence that an evaluation can have. In some cases clients use the evaluation to improve policy or programme design. In other cases it is used to provide political cover for taking difficult decisions. Sometimes the findings of an evaluation also provide objective and independent information with which civil society can more effectively pressure government agencies to improve service delivery. The evaluation process may also serve as a catalyst to bring together a broader range of stakeholders or to help create a commission or working group. The case study on the evaluation of large dams also points out that the process of conducting an evaluation may have consequences that were not anticipated or desired by the agency commissioning the evaluation.

Table 1: The types of effects evaluations can have: illustrated by the Influential Evaluations case studies

1. India: Employment Assurance Scheme	<ul style="list-style-type: none"> • Broad, interagency perspective helped identify duplications and potential cost-savings • The position of the Evaluation Office permitted high level access to the Planning Commission
2. India: Citizen Report Cards	<ul style="list-style-type: none"> • Alerted public service agencies to problems of which they were not fully aware or had not considered important • Provided objective, quantitative data that civil society could use to pressure agencies to improve services
3. Indonesia: Village Water Supply	<ul style="list-style-type: none"> • Made policy-makers aware of the importance of women's participation and the benefits of participatory planning
4. Large Dams	<ul style="list-style-type: none"> • Created political space for civil society by introducing new social and environmental criteria for assessing dams • Launched a dialogue which facilitated the creation of the World Commission on Dams • Discouraged risk-averse funding agencies from investing in large dams (even when there appeared to be a strong social and economic justification)
5. Pakistan: Wheat-flour ration shops	<ul style="list-style-type: none"> • Provided political cover to government to take a politically sensitive decision to eliminate subsidies • Showed how to mitigate negative consequences for influential "losers"
6. Uganda: Education Expenditures	<ul style="list-style-type: none"> • Developed methodology to document what everyone suspected (that approved funds were not reaching schools) • Provided documentation to civil society to pressure for improvements
7. Bulgaria: Environmental Consequences of a Metallurgical Project	<ul style="list-style-type: none"> • Alerted borrowers and the regional development bank to new EU environmental legislation • Showed the company how to avoid fines • Showed the company how to advance the launch date for mineral production
8. China Forestry Policy	<ul style="list-style-type: none"> • Legitimized questioning of the logging ban (which had eliminated millions of jobs) • Promoted more in-depth policy research on forestry • Facilitated the creation of the Forestry Task Force.

Source: World Bank OED (2004), Influential Evaluations.

Some of the discussions on the *Influential Evaluation* report pointed out that, in the majority of the case studies, the main effect of the evaluations was to point out things that had gone wrong and to show how clients and other stakeholders were able to take corrective measures. However, many evaluations can show that a programme is going well and can provide support for its continuation or expansion.⁵ In these cases however, it is often more difficult to demonstrate clearly the contribution of the evaluation since it is providing further support for a decision that an agency was already planning to take.

What difference did the evaluation make?

Table 2 uses the *Influential Evaluations* case studies to illustrate the differences an evaluation can make. The cases illustrate the difference made by the evaluations. These were: cost savings; forcing agencies to take actions; strengthening gender analysis and participatory planning; broadening the evaluation criteria used to assess programmes; discouraging future investments by risk-averse funding agencies; increased efficiency of service delivery; and, facilitating the creation of important policy and planning taskforces or organizations.

Table 2: The differences evaluation can make: taken from the Influential Evaluations case studies

Differences made	Evaluations
Major cost savings	<ul style="list-style-type: none"> • India: Employment Assurance • Bulgaria: Metallurgical Industry • Pakistan: What-flour ration shops
Increased financial benefits	<ul style="list-style-type: none"> • Uganda: Education Expenditures • Bulgaria: Metallurgical Industry
Forced action	<ul style="list-style-type: none"> • Bangalore: Citizen Report Cards • Uganda: Education Expenditures
Strengthened gender and participatory planning	<ul style="list-style-type: none"> • Indonesia: Village Water Supply

5 The evaluation of the Progresa (now called "Oportunidades" conditional cash transfer programme in Mexico is often cited as a case where a very positive evaluation was able to convince a new President to continue a programme started under a previous administration.)

Introduced broader evaluation criteria but discouraged future investments	<ul style="list-style-type: none"> • Large Dams
Increased efficiency of service delivery	<ul style="list-style-type: none"> • India: Employment Assurance • Bangalore: Citizen Report Cards • Indonesia: Village Water Supply
Facilitated creation of important policy and planning agencies	<ul style="list-style-type: none"> • Large Dams • China: Forestry Policy

Source: World Bank, OED (2004) *Influential Evaluations*

Ways to strengthen evaluation utilization

Creating ownership of the evaluation

One of the key determinants of whether an evaluation will be useful and, whether the findings will be used, is the extent to which clients and stakeholders are involved in all stages of the evaluation process. Do the clients feel that they “own” the evaluation, or do they not really know what the evaluation will produce until they receive the final report? How many times have evaluators felt frustrated when the main reaction from the client to a well prepared report is, “This is not what we wanted or expected.” The approach used in the initial scoping phase of the evaluation is critical both for creating a sense of client ownership and also for understanding the client’s information needs (see Point B).

An effective way to enhance the sense of ownership (see Point F), is the use of formative evaluation strategies, which provide constant feedback to key stakeholders on ways to use the initial evaluation findings to strengthen project implementation.

Effective communication strategies

Promoting a positive attitude toward evaluation findings often involves ensuring that clients face “no surprises” (World Bank, IEG 2004). The client should be kept informed of the progress of the evaluation and of preliminary findings as they emerge. In particular, the client should be fully briefed on, and should have a chance to react to, the final conclusions and recommendations before they are presented or made available to others. Clients tend to react more defensively to negative findings if they are sprung on them in a formal meeting with other agencies or, even worse, if they learn the

findings from the press or another agency. As always there is the need to involve clients while maintaining neutrality. This is particularly the case where some negative or sensitive results are emerging which the client may wish to suppress.

Deciding what to evaluate

A successful evaluation will focus on a limited number of critical issues and hypotheses based on a clear understanding of the information needs of clients and how the evaluation findings will be used. The following questions help define the evaluation focus and the required level of precision of the information and the analysis:

- *What do the clients “need to know” and what would they simply “like to know”?* This distinction is critical when deciding whether the data collection instruments can be simplified and information needs reduced.
- *How will the evaluation findings be used?* To defend the programme from its critics? As an initial exploration of whether a new approach seems to work? To present statistically precise estimates of whether the programme has achieved quantitative goals? To estimate the cost-effectiveness of the programme compared to competing programmes?
- *How precise and rigorous do the findings need to be?* As we have emphasized earlier, some digging may be required to determine this. Sometimes the client will state that “a rigorous scientific evaluation” is required. The evaluator may assume this means that a large sample survey is needed to support a pre-test post-test evaluation design with a comparison group. However, the client may only mean that the report must be considered by parliament or the funding agency to have been professionally conducted.

Basing the evaluation on a programme theory (logic) model

A programme theory (logic) model developed in consultation with stakeholders is a good way to identify the key questions and hypotheses which the evaluation should address. It is essential to ensure that clients and stakeholders and the evaluator are “on the same page” with respect to the understanding of the problem addressed by the programme. A common understanding is essential on what the objectives are; how it is expected to achieve these objectives; and, what criteria the clients will use to assess success. In some

cases it is necessary to formulate two or more different programme theory models to reflect the views of different stakeholder groups. This can be particularly important if the evaluation needs to recognize the views of important critics of the programme. Even when conducting an evaluation under extreme time pressure it is essential to find time to involve clients in this process, so as to give them ownership and, so that they have a stake in the evaluation outcomes. There is a great temptation for the evaluator to prepare a logic model describing what they understand the underlying theory to be, and then to present this to the client. Silence or unenthusiastic nods of the head are taken (often wrongly) to signal full understanding and agreement.

Understanding the political context

It is important for the evaluator to understand as fully as possible the political context of the evaluation. Evaluations often address sensitive or even confidential issues so a great deal of discretion and tact is required. The following are some of the issues the evaluator should try to understand:

- Who are the key stakeholders and what is their interest in the evaluation?
- Who are the main critics of the programme, what are their concerns/criticisms, and what would they like to happen? What kinds of evidence would they find most convincing? How can each of them influence the future direction of the programme (or even its continuance)?
- What are the main concerns of different stakeholders with respect to the methodology? Are there sensitivities concerning the choice of quantitative or qualitative methods? How important are large sample surveys to the credibility of the evaluation?

In addition to the sensitivity of the questions, the relationship with the client on these issues must also be treated carefully. While it is important to gain the confidence of the client, it is essential for the evaluator to maintain objectivity and not to be seen as an ally of programme management against their critics.

Timing of the launch and completion of the evaluation

Many well designed evaluations fail to achieve their intended impacts either because they were completed too late (the critical decisions have already been made on future funding or programme directions), or, too early before the questions being addressed are

on the policy-makers' radar screen. Many evaluations had little practical utility because they missed some critical deadlines or, failed to understand who were the key actors in the decision-making process. A report delivered on March 4 may be of no practical use if the Ministry of Finance had already made decisions on future funding on March 3! Similarly, utility will be reduced if the findings did not reach some of the key decision-makers on time (or at all).

Defining the appropriate evaluation methodology

A successful evaluation must develop an approach that is both methodologically adequate to address the key questions and hypotheses and that is also understood by, and acceptable to clients. Many clients have strong preferences with respect to the use of quantitative, qualitative or mixed-method designs. They may also have strong opinions on the merits (or limitations) of randomized control trials and other strong quantitative evaluation designs. Consequently one of the factors contributing to under-utilization of the evaluation may be client disagreement with, or lack of understanding of, the methodology.

Most evaluations are conducted under less than ideal circumstances. Budgets are usually less than required for a rigorous evaluation design, there are often time pressures to complete and present the evaluation findings, and important information (such as baseline data), is frequently not available or of dubious quality. An effective evaluation must adapt to these constraints. The constraints will be assessed by the evaluator but their importance and the client's flexibility (e.g. to delay submission of the report or obtain additional funds) must be discussed and agreed with the client. It will sometimes be found, for example, that the "deadline" for submitting the report is in fact only a deadline for the preparation of an informal status report. Similarly there may be some flexibility in the evaluation budget. However, these constraints and priorities must be fully discussed in a strategy session with the client and the evaluator must never assume that, for example, "the client would not mind waiting a few more weeks to get a better report."

Process analysis and formative evaluation strategies

"An evaluation intended to furnish information for guiding programme improvement is called a formative evaluation (Scriven, 1991), because its purpose is to help form or shape the programme to perform better" (Rossi, Lipsey and Freeman, 2004: 34). Even when the primary objective of an evaluation is to assess programme

outcomes and impacts, it is important to “open-up the black box” to study the process of programme implementation⁶. This is important for the following reasons:

- To explain why certain expected outcomes have or have not been achieved.
- To explain why certain groups may have benefited from the programme while others have not.
- To identify and assess the *causes* of outcomes and impacts. These may be planned or unanticipated, positive or negative.
- To provide a framework for assessing whether a programme that has not achieved its objectives is fundamentally sound and should be continued or expanded (with certain modifications) or whether the programme model has proved not to work, or at least, not in the contexts where it has been tried so far.

In addition to the above reasons, the analysis of the programme implementation process enables the evaluation to contribute to improving the performance of the ongoing programme (formative evaluation). This can be done in two ways. The first is for the evaluator to provide regular feedback and suggestions to programme management and other key stakeholders. The second way is to involve programme staff and other stakeholders in the evaluation so that they learn for themselves what is working and what is not.

Many, but not all, formative evaluation strategies help promote evaluation utilization as stakeholders begin to use the findings, during the process of evaluation, long before even the draft final evaluation reports have been produced. Involving clients at this early stage also means that they are more likely to review the final reports and consider how to use the recommendations.

Evaluation capacity building

While some evaluations are one-time activities which will probably not be repeated, others are likely to continue over a number of years, over different phases, or over subsequent programmes. In such instances, *evaluation capacity-building* (strengthening the

6 The issue of the “black box” (a term widely used in economic analysis) only concerns quantitative pre-test post-test evaluation designs where data is only collected at the start and the end of the project and the project implementation process is not studied. Many authors, including Stufflebeam’s CIPP; Greene’s participatory; Stake’s responsive; Patton’s utilization; Fetterman’s empowerment; Scriven’s goal-free; House’s democratic approaches – do not treat implementation or process as a black box.

capacity of stakeholders to commission, design, implement, interpret and use evaluations), can be viewed both as enhancing the quality and utility of the ongoing evaluation and, as an investment to strengthen the use of findings.

Evaluation capacity building includes both strengthening the technical capacity of evaluators to conduct evaluations and also the capacity of clients and stakeholders to interpret and use the findings of the evaluation. Although evaluation capacity building is often limited to working alongside *evaluation practitioners*, one of the most important components is in fact, to strengthen the motivation and capacity of managers, planners, policy-makers, legislators, funding agencies and public opinion to commission, assess and/or use evaluations. When agencies do not use evaluation findings, one of the contributing factors is often a lack of evaluation capacity in one of the areas described above.

From the perspective of enhancing evaluation utilization the following are some of the key activities to undertake with *evaluation users* (and also with the client commissioning the evaluation):

- Involve key stakeholders and other potential users in the scoping and design phase. The construction of the programme theory model is an important opportunity for capacity building, providing an opportunity to discuss important concepts such as input, output and process indicators and the definition and measurement of impacts.
- Help users understand the logic of the evaluation design, and the trade-offs between the different possible designs, in terms of how the evaluation will be used.
- Invite interested stakeholders to participate in some of the evaluation training programmes or workshops which might be organized primarily for the evaluation practitioners.
- Try to involve all key user audiences in the periodic briefings on the progress of the evaluation.
- Involve users as resource persons when evaluations are organized with other agencies.

Communicating the findings of the evaluation

Many potentially useful evaluations have little impact because the findings are not communicated to potential users in a way which is useful or comprehensible to them. Even worse, the findings may

never ever reached some intended users. Some guidelines for communicating evaluation findings to enhance utilization are:

- Clarify what each user wants to know and the amount of detail required. Do specific users want a long report with lots of tables and charts or a brief overview? Do they want many details on each project site, school, region, or just a summary of the general findings?
- Understand how different users like to receive information. In a written report? In a group meeting with slides or PowerPoint? In an informal personal briefing?
- Clarify if users want “hard facts” (statistics) or whether they prefer photos and narratives. Do they want a global overview, or to understand how the programme affects individual people and communities?
- Be prepared to use different communication strategies for different users. One size usually does not fit all.
- Ensure presentations are pitched at the right level of detail or technicality. Do not overwhelm managers with statistical analysis or detailed discussion of sample design, but do not insult professional audiences by implying that they could not understand the technicalities.
- Ascertain what the preferred medium is for presenting the findings. A written report is not the only way to communicate findings. Other options include: oral presentations to groups, video, photographs, meetings with programme beneficiaries or visits to programme locations. Sometimes attending a meeting in the community in which residents talk about the programme can have much more impact than a written report.
- Make sure the communication is in the right language(s) when conducting evaluation in multilingual communities or countries.

Developing a follow-up action plan

Many evaluations present detailed recommendations but have very little practical utility because the recommendations are never put into place, even though all groups might have expressed agreement. What is needed is an agreed action plan with specific, time-bound actions, clear definition of responsibility, and procedures for monitoring compliance. Many government and international agencies have standard procedures to monitor the implementation

of evaluation recommendations. For example, the World Bank Independent Evaluation Group and many other agencies keep a log of all recommendations included in its evaluations, management response to these, and agreed actions. Periodic follow-ups to report on the status of the agreed actions also take place.

The definition of a follow-up action plan is an effective way to promote utilization of the evaluation findings. Some of the steps, as stressed above, include:

- A key strategy is to ensure client and stakeholder “buy-in” to the evaluation process so that there is willingness to review, and where there is agreement, to implement the evaluation findings.
- The evaluation report must identify the key issues on which decisions must be taken and follow-up actions agreed. However the external evaluator needs to be cautious about presenting specific recommendations so as not to discourage users from taking ownership of the action plan. In preparing the evaluation report the evaluator, in consultation with the clients, must decide whether it is better to:
 - o Present a list of issues but not propose specific actions;
 - o Present a number of follow-up options but not recommend which one is best;
 - o Present specific recommendations on follow-up actions. This may be appropriate when discussing technical issues (for example which financial management package is compatible with the computer systems used by the agency).
- The action plan must be developed by the interested organizations with the evaluator as a technical resource and possibly facilitator. It is sometimes better for the evaluator not to participate in the action planning meetings so as to give more feeling of ownership and freedom of action to the agencies themselves.
- Often the evaluator can help develop measurable indicators and timetables to monitor progress. One of the evaluator’s key contributions is to ensure that the action plan is actually developed before she or he leaves.

Conclusion: Special challenges in enhancing the utilization of evidence-based evaluations

There is a growing demand for more rigorous, “evidenced-based” methodologies for assessing the effectiveness of development assistance. Patton (2005), defines “evidence-based” evaluation as the use of methods to ensure that the impact of the intervention “has been subjected to at least some empirical validity”. While most evaluators would probably agree that claims concerning the effectiveness or impact of a programme should be supported by the best empirically verifiable evidence, the problem is that there are major disagreements as to what comprises the “best”, or even good evidence. One school of evaluators has argued that randomized control trials (RCTs),⁷ are usually the best form of evidence. For example, the U.S. Department of Education recommended, in 2005, that priority should be given in all educational evaluations to the use of RCTs⁸. The Poverty Action Lab⁹ at MIT offers international training programmes advocating the use of RCTs, and the Center for Global Development advocates the increased use of RCTs and strong quasi-experimental designs¹⁰. Supporters of RCTs have coined the term the “gold standard” to refer to this approach.

The debate over whether RCTs are (i) the best way to evaluate impacts; (ii) one of several equally valid design options, or (iii) an inappropriate or even unethical way to assess the effectiveness of

-
- 7 The essential characteristic of a randomized control trial is that all individuals, households, communities or organizations who satisfy project eligibility criteria are randomly assigned to the project or control groups. This is often done through a public lottery to ensure a transparent selection process. The great methodological advantage of this approach is to eliminate the different kinds of selection bias that occur when participants are either self-selected or are selected by the project agency.
 - 8 For an example of this debate see <http://www.eval.org/doestatement.htm> which gives the response of the American Evaluation Association to the proposal by the US Department of Education to give priority to randomized control trials. This also includes links to the original document.
 - 9 See www.povertyaction.com for examples of randomized control trials and information on the Poverty Action Lab program.
 - 10 A strong quasi-experimental design (QED) uses a pre-test/post-test design with a project and comparison group. The difference is that with QEDs separate selection procedures are used for the project group (usually self-selection or selection by the implementing agency) and the comparison group. This presents the problem that there may be some systematic differences between the two groups (for example participants may be more motivated or have more years of education) and that what are assumed to be effects of the project may be due, at least in part, to these initial differences between the two groups.

development programmes, has strongly divided the evaluation community¹¹. On the positive side the insistence on more rigorous and independently verifiable standards of evidence for assessing project impacts has required many agencies to define more carefully what their programmes are intended to achieve and how their effectiveness can be measured. The Center for Global Development's May 2006 report, *When will we ever learn? Improving lives through impact evaluation*,¹² has served to focus this debate by pointing out the economic and social costs of poorly designed evaluations leading to wrong conclusions on whether programmes are achieving their development objectives (and by implication whether they should be continued or terminated). This report has encouraged many development agencies to give greater priority to a more rigorous assessment of the effectiveness of their development interventions. At the same time critics have challenged the report's conclusion that experimental and strong quasi-experimental designs normally offer the most valid way to assess development impacts.

This debate has created additional challenges for ensuring the utilization of evaluations (in addition to all of the problems we have already discussed!). When technical debates among evaluation specialists are used to decide evaluation design, rather than consultations with intended users, as to what methods would be most useful to answer their priority questions, some developing country agencies feel disempowered. The fact that many "summative" evaluation designs¹³ do not produce their findings until the project is completed, serves to reinforce that feeling of disempowerment. At that stage, the information is often of little practical use to managers and policy-makers because it is too late to use it to correct problems during project implementation. In addition, decisions on the termination, continuation or expansion of the programme have

11 In a July 2006 presentation to the International Program for Development Evaluation Training (unpublished but available from the author) Michael Bamberger argued that while RCTs and strong quasi-experimental designs are potentially very powerful, experience to date suggests that strong evaluation designs have probably been used in less than 10 per cent of impact evaluations in developing countries and that RCTs have probably been used in significantly less than 5 per cent of evaluations.

12 <http://www.cgdev.org/content/publications/detail/7973>

13 A *formative* evaluation design is intended to provide constant feedback to program management to help detect problems and to find ways to improve project implementation and ensure accessibility to all sectors of the target population. In contrast, a *summative* evaluation is intended to estimate the effects (outcomes or impacts) of a project intervention and to determine to what extent the observed changes can be attributed to the project intervention. Very often a summative evaluation will not provide any feedback to managers and policy-makers until the end of project.

often already been made. There are also issues concerning the consequences of RCTs for the programme being evaluated. The need for a small number of precisely defined quantitative indicators of inputs, outputs and impacts, and for a standard treatment that does not vary over the life of the project, may limit the flexibility of a programme to adapt to changing circumstances. Such flexibility is essential in many development contexts.

The debate over what constitutes a good evaluation design presents several challenges for evaluation utilization.

Firstly, it is important to ensure that clients understand the debate and what is at issue. The evaluator must help them to understand that, despite what they may have been told by funding agencies or consultants, there is no consensus among evaluators (or development agencies), on what constitutes the “best” evaluation design. Clients should be made aware of the fact that all evaluation designs have strengths and weaknesses; that different designs will be more appropriate in different circumstances; and, that mixed-method designs, combining both quantitative and qualitative methods, will almost always be more robust than single-method designs. The most appropriate design in a given circumstance depends on a number of considerations including: the purpose of the evaluation; the required level of precision; budget and time constraints; data availability; and, not least, the methodological preferences of the client and other key stakeholders.

Secondly, it is essential to ensure that clients are fully involved in the decisions on evaluation design and that they feel ownership of the evaluation.

Thirdly, one of the roles of the evaluator may sometimes be to provide guidance and perhaps moral support to clients who feel pressured by funding agencies to use evaluation methodologies with which they do not feel comfortable; which they do not fully understand; or, which they believe are not appropriate for their needs.

If the evaluator can address these challenges the debate over evidenced-based evaluation can result in a more active involvement of clients and stakeholders in the choice of the most appropriate evaluation design and, ultimately, in enhanced utilization of the evaluation findings.

References

- Bamberger, M; Rugh, J; and Mabry, L. (2006), *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Sage Publications.
- Best, J. (2001), *Damned lies and statistics: Untangling numbers from the media, politicians and activists*. Los Angeles. University of California Press.
- General Accounting Office. (1995), *Programme Evaluation: Improving the Flow of Information to the Congress*. GAO/PEMD-95-1. Washington. D.C. GAO.
- House, E. (1972), *The Conscience of Educational Evaluation Teachers College Record* 73(3): 1972, 405-14.
- House, E. (1990), *Trends in Evaluation*. *Educational Researcher* 19(3): 1990, 24-28.
- Operations Evaluation Department. (2002), *Memorandum to the Executive Directors and the President on the Annual Report on Operations Evaluation*. June 10, 2002.
- Operations Evaluation Department. (2004), *Influential evaluations: Evaluations that improved performance and impacts of development programmes*. Independent Evaluation Group. Washington D.C. The World Bank. (www.worldbank.org/ieg/ecd)
- Operations Evaluation Department. (2005), *Influential evaluations: Detailed case studies* Independent Evaluation Group. Washington D.C. The World Bank. (www.worldbank.org/ieg/ecd)
- Patton, M. Q. (1997), (Third Edition). *Utilization-focused evaluation*. Thousand Oaks. California: Sage Publications.
- Patton, M.Q. (2005), *Best Practices in Mathison, S. (Ed.). Encyclopedia of Evaluation*. Thousand Oaks, California. Sage Publications.
- Picciotto, R. (2002), *International Trends and Development Evaluation: The Need for Ideas*. *American Journal of Evaluation*, 24 (2), 2002. pp/ 227-234
- Rossi, P; Lipsey, M & Freeman, H. (2004), *Evaluation: a Systematic Approach* (7th Edition). Thousand Oaks. CA, Sage.
- Scriven, M. (1991), *Evaluation Thesaurus* (4th ed.). Newbury Park, CA: Sage.
- Swedish International Development Agency (SIDA). (1999), *Are Evaluations Useful? Cases from Swedish Development Cooperation*. Department for Evaluation and Internal Audit.
- Wholey, J.S. et al. (1970), *Federal Evaluation Policy: Analyzing the Effects of Public Programs*. Urban Institute. Washington D.C.
- World Bank. (2004; 2005) *Operational Evaluations Department, Influential Evaluations*.

COUNTRY-LED EVALUATION. LESSONS LEARNED FROM REGIONS

by Marie-Helene Adrien, President, IDEAS and
Dennis Jobin, Vice President, IDEAS

Introduction¹

Country-Led Evaluation (CLE) is a relatively new concept, and one that reflects the world's growing recognition of the importance of a nation's self-determination in its own development.

In recent years, the field of development evaluation has evolved considerably. Contemporary discussions and events in the international development arena have broadened the scope and design of evaluation, from an earlier, narrower focus on projects or programmes to broader assessments that encompass policy, policy coherence, and development outcomes. Consequently, newer evaluation methodologies now consider many factors including gender equity, social justice, environmental sustainability, and participation.

At the same time, there has been increasing pressure to make evaluation central to a country's own development process and more relevant and meaningful to the people whose lives are affected by development interventions. The field of evaluation is being reshaped by the evolving context of international aid, and particularly, by the emerging recognition that effective development assistance requires that donor agencies "respect partner country leadership and help strengthen their capacity to exercise it."²

Conventional forms of evaluation, typically mandated and funded by development agencies, are now being challenged by emerging independent forms of assessment which put the recipient country in the driver's seat. The rationale for CLE is clear, but the question now is how to do it. What are the obstacles to Country-Led Evaluation? What needs to be done to support it?

1 This paper reflects the positions of the two authors only.

2 *Paris Declaration on Aid Effectiveness*, High Level Forum, Paris, February 28-29 2005, p. 2

IDEAS

Since its inauguration in 2002, the International Development Evaluation Association (IDEAS) has worked to make evaluation more relevant and useful. In keeping with its mission – “to improve and extend the practice of development evaluation by refining knowledge, strengthening capacity, and expanding networks for development evaluation, particularly in developing and transition countries” – IDEAS has focused on rethinking, reforming and reshaping development evaluation. In a concerted attempt to make evaluation more central to development and the eradication of human poverty, IDEAS has promoted the notion that rights and equality, justice and freedoms, peace and security are all legitimate dimensions of evaluation. It has focused on finding ways to make the outputs of evaluation more meaningful to the people whose lives are affected by projects, programmes, and broad policy interventions. This has involved thinking through standard methods and approaches, and encouraging new ones that make the practice of evaluation more rigorous and more participatory. This move to make evaluation more central has also meant grappling with the issues of ownership and governance of evaluation.

In order to better understand these issues, IDEAS conducted two regional workshops on Country-Led Evaluation (CLE) – one in the Central and Eastern European Region (Prague, Czech Republic, June 2006), and one in the Africa region (Niamey, Niger, January 2007). These workshops brought together representatives from government, academia, the private sector, and the NGO community, who had something to share about Country-Led Evaluations.

The specific aims of the regional conferences were to:

- obtain views on how participants define Country-Led Evaluation, and its design and purpose;
- provide a forum for sharing regional experiences to foster networking and knowledge sharing;
- identify factors that enable or hinder CLE;
- identify lessons learned from country and regional experiences;
- encourage discussions of how to develop the capacities required for CLE.

This paper provides some insights about what CLE is and its importance for development effectiveness, and summarizes the find-

ings from the two CLE workshops. (For detailed reports from the workshops, please see CLE Report – Prague (IDEAS, 2006) and CLE Report – Niamey (IDEAS, 2007) on the IDEAS website: www.ideas-int.org/)

Country-Led Evaluation

While the OECD-DAC³ does not yet have an official definition of Country-Led Evaluation (CLE), several development agencies have come to define it as “evaluations in which the country leads the evaluation by determining which evaluations will be done, and is responsible for steering and managing them.”⁴ CLE fosters country ownership of the development process, and reflects a paradigm shift in the delivery of development aid which is illustrated in the table below.

<i>The Changing Paradigm of Development Aid</i>			
ORIENTATION	PAST	PRESENT	FUTURE
Management and control	Donor's Audit	Donor Led Evaluation	Country Led Evaluation
Focus	Money	Policy	Institutions: good governance
Instruments	Projects	Programmes	Partnerships
Unit of Account	Inputs	Outputs	Results & Outcomes
Emphasis	Donor driven	Joint sponsorship -Swaps	Country ownership
Dominant discipline	Engineering Education	Macro economic and finance	Multidisciplinary

Source: Jobin, Denis (Forthcoming) "A Transaction-Cost-Based Approach to Partnership Performance Evaluation" Evaluation: The International Journal of Theory, Research and Practice, U.K.

-
- 3 The Organization for Economic Development and Cooperation's Development Assistance Committee
- 4 Country-Led Evaluations. A discussion note prepared by WB/OED, UNDP/EO and IOB. March 2003.

Enabling factors and barriers to Country-Led Evaluation

Introduction

Traditionally, most evaluations have been funded, and led, by development agencies to meet their own requirements. Over time, however, it has become clear that this approach undermines recipient country ownership of development results and significantly increases the transaction costs of development partners. In contrast, country ownership of evaluation has been found to favour development effectiveness by increasing the use of evaluation information and the efficient use of evaluation resources. Along with ownership comes the need for recipient country accountability and responsibility, and the need to learn from experience in order to improve performance. The challenges of CLE are those of a true partnership: what conditions are required for recipient countries to take on this new role and, for development agencies to abandon their own monitoring and evaluation (M&E) systems in favour of the systems of the developing country partner? In the following sections we explore some possible answers to these questions.

Good governance

Good governance is essential to the development of CLE. It has many 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.⁵

While some developing countries and countries in transition have seen the emergence of champions who have led government reforms toward good governance, others have been less successful.⁶

Clearly, development agencies' aid strategies must accommodate the context and specific needs of the region or country where their development interventions take place. But in terms of CLE, trust is paramount. In order to support CLE, development agencies must feel confident that they can rely on the evaluation information that

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

6 See: <http://www1.worldbank.org/publicsector/Reforming.pdf>

is generated by Country-Led Evaluations. Hence, the importance of good governance.

Monitoring and evaluation capacity

While the ability to initiate, manage and fund evaluations at the national level is essential for genuine CLE, developing country partners face a number of challenges in achieving this goal. Firstly, most countries do not have the financial resources to create or sustain an internal market for evaluations capable of nurturing a national monitoring and evaluation system. Secondly, most developing countries, challenged by weak coordination within and between national government departments and by the lack of consistency in monitoring and evaluation approaches and methods, have not developed an adequate national framework for monitoring and evaluation. Thirdly, there is a dearth of human capacity, particularly in evaluation skills and knowledge, and more training in evaluation methods and approaches is needed. Finally, most existing evaluation standards were developed by donor countries, and developing country partners need to develop national evaluation standards that reflect their own context and culture.

Providing developing country partners with opportunities to steer, manage and conduct evaluations (*learning by doing*), appears to be an effective way to support capacity development. Such opportunities build national monitoring and evaluation capacity and also improve the quality of evaluation information. A first step on the road to true CLE might be the experience of joint evaluation. There have been cases reported where joint evaluations (conducted by development agencies and developing country partners), have strengthened the capacity of both partners and have provided better quality and more useful information for decision making.⁷

A culture of evaluation

Whilst greater national government commitment to “managing for results” is essential to CLE, a range of other factors also help to create the environment for CLE. These include:

- an active civil society which demands government accountability and advocates for transparency in the use of public funds and, which supports the development of evaluation capacities through public education;

7 CLE Prague Report. Joint evaluations, UNICEF, Jean Quesnel, pg 15

- evaluation champions who provide active leadership for CLE⁸;
- national evaluation associations;
- building trust and a common vision between development partners – through harmonization and alignment of development efforts.

Participants in both IDEAS workshops noted that civil society and the private sector have important roles to play with regard to CLE. Indeed, the participation of both civil society and the private sector support the independence of evaluations and thus increase their credibility.

The emergence of national evaluation associations also seems a promising way of developing and supporting an evaluation culture.⁹ National evaluation societies can play a central role in building the capacity of individual members through training, advocating for their professional interests, developing evaluation standards and norms, and providing opportunities for sharing knowledge and experience through networks.

Development agency commitment to country ownership

Recent agreements in the international development community show that development agencies are committed in principle to improving development effectiveness, and provide a significant incentive to support a shift toward greater country ownership of development processes, and hence CLE. Indeed, in March 2005, the *Paris Declaration on Aid Effectiveness* reaffirmed international development agency commitments made in Rome in 2003 (*Rome Declaration on Harmonization*). These are:

- **Ownership:** partner countries exercise effective leadership over their development policies and strategies and coordinate development actions.
- **Alignment:** development agencies base their overall support on partner countries' national development strategies, institutions and procedures.

8 CLE Report - Niamey; p.10 cases in Egypt and Niger

9 CLE Prague Report, pg. 16

- **Harmonization:** development agencies' actions are more harmonized, transparent and collectively effective.
- **Managing for results:** managing resources and improving decision-making for results.
- **Mutual accountability:** development agencies and partners are accountable for development results.

While these high level donor commitments and guiding principles would seem to support the notions of country ownership and development initiatives which focus on results, tangible results remain elusive.

Somewhat paradoxically, one key finding from the CLE workshops in both Africa and Eastern Europe is that the market for evaluation remains largely donor-driven. In a context in which limited national resources are available to conduct evaluations, development agencies still have a significant role in deciding *why*, *when* and *what* to evaluate, largely because they control the funds for this.

Nevertheless, development agencies can and are beginning to play a strategic role in fostering the development of CLE. In Eastern Europe, for example, the European Union (EU) community has focused its aid strategy on building monitoring and evaluation capacity by:

- establishing and/or strengthening evaluation units in member states' departments responsible for managing European Union (EU) structural funds
- financing evaluation capacity building efforts under EU programmes
- providing guidance on evaluation manuals, working documents, guidelines for common understanding, etc.;
- facilitating evaluations at EU level;
- producing evaluation guides.¹⁰

10 CLE Report – Prague, page 15, and http://www.ideas-int.org/Documents/Mairate_paper.doc

Lessons learned in Eastern Europe and Africa

Eastern Europe Region

IDEAS first CLE workshop was held in the Czech Republic (Prague, June 19-20, 2006). The main lessons learned coming out of this workshop for Eastern Europe were:

- Development cooperation must be a two-way partnership. While focusing on recipients' needs, we cannot forget the motivations and interests of development agencies. We have to identify the mutual benefits of and the responsibilities for the CLE approach.
- Development agencies have an important role to play in fostering CLE, by creating or enhancing the demand for CLE, enforcing standards, and developing guidelines.
- In general, there is a weak evaluation culture in Eastern Europe, perhaps due to the political and historical institutional framework. The regional experiences shared at the workshop called for more networking, developing evaluation capacity, and a broad dialogue rather than a radical step towards a centralized evaluation system.
- Due to the regional political background and its recent history, the role of civil society is of particular importance to CLE. Civil society can serve as a bridge between a government, development agencies, and a "country" in its broadest meaning. Government and public institutions ought to be open to dialogue, disseminate the evaluation results and, make the processes more transparent, democratic and credible.
- When considering capacity building for CLE in Eastern Europe, it is important for donors and development agencies to distinguish between countries in transition and developing countries. Although they face similar problems, their experience in evaluation differs substantially. For instance:
 - o recent EU member states (such as the Czech Republic, Hungary, Lithuania, Poland and Slovakia) have had similar experiences in the evaluation of EU structural fund projects and in the evaluation of their own Official Development Assistance (ODA) projects and programmes. As a result, they have more mature CLE capabilities;

- o EU accession countries (such as Bulgaria and Romania) have made some progress in EU-driven CLE capacity building, although they are at a less mature stage;
- o other partner countries (such as Armenia, Azerbaijan, Bosnia and Herzegovina, and Moldova), while not directly affected by the EU integration, have had some experience in building their evaluation capacities.

In Eastern Europe, countries which are members of the European Union have more and better support for capacity building of CLE than non-member countries.

Africa Region

The CLE workshop in Niamey, Niger (January 16, 2007), held under the umbrella of 4th AfrEA conference, invited participants from across the African continent to share their experiences with regard to CLE. The following points summarize the key lessons learned from the African experiences.

Developing a demand, a framework, and a culture for evaluation

- CLE will become a reality only when there is an internal demand for evaluation. Since the concept of CLE is based on the assumption that beneficiary countries will take the lead in evaluation, the final objective must be to encourage the demand for evaluation by these countries.
- CLE will become effective when evaluation is accepted and used like other administrative functions in project and programme management.
- The definition of an institutional framework for CLE is essential, with roles and responsibilities clearly identified. The culture for CLE will be stronger if this framework is supported by solid policies for planning, monitoring and evaluation.
- A significant element in the success of CLE will be the strategic shift from projects that are designed and controlled by development agencies to longer-term programmes that are designed and carried out by nationals.
- Training civil servants in monitoring and evaluation (M&E) concepts and approaches, and providing follow-up support in their daily work, is necessary to reinforce a culture of evaluation.

Participation

- The use of participatory approaches to evaluation enhances local ownership and utilization of the evaluation results.
- It is imperative to ensure that key stakeholders (governments, development agencies, civil society, etc.), participate in national CLE processes and contribute to making them coherent and effective.
- The purpose and scope of a CLE should be negotiated between development agencies and local stakeholders early in the process (at the initiation stage).

Planning and implementation for Country-Led Evaluation

- Development interventions should be aligned with national priorities and integrated into a national monitoring and evaluation system, thus avoiding the development of parallel efforts.
- For successful CLE, the local context must be considered during both the preparation and implementation phases of CLE.
- It is useful to plan pilot phases of CLE in a way which aids documentation of results and identifies lessons learned which can inform future CLE processes.
- It is important to plan for monitoring and evaluation, as the information generated is essential to good decision making.

Conclusion

Country-Led Evaluation is a timely and reasonable idea which is supported by recent thinking, in the international development community, on how to improve the effectiveness of development aid.

It is now widely recognized that development initiatives work better when recipient countries identify their own priorities and the needs to be addressed, when they participate in the planning and execution of initiatives, and when development agencies and recipient countries are jointly accountable for the results. However, while development agencies have committed to support country ownership of development processes, there remain huge gaps in the capacity of most developing countries to take on the responsibility for such processes.

Country-Led Evaluation can provide valuable information for improving the performance of development initiatives, and consequently

for improving the quality of life of those targeted by such initiatives. Like all other development processes, however, effective evaluation requires capacity, support and funding, at every level, from individual skills and knowledge, to departmental cultures for evaluation, to national frameworks for monitoring and evaluation, to regional networks for sharing knowledge and experience. Perhaps most important of all, it requires that stakeholders recognize the value of evaluation and demand it. Building this understanding and demand is one of IDEAS' primary concerns.

Part of the IDEAS' mission is to advocate, share knowledge, and support networking for CLE. Therefore, we will continue to advocate for the ongoing efforts made by countries and development agencies to develop evaluation capacities, and we will continue to share regional CLE experiences and knowledge through the IDEAS website and other fora such as our Biennial. Finally, we hope that by convening groups and individuals, at the regional level, we will foster networking that will help each country learn from the experience of others in Country-Led Evaluation.

References

DAC Working Party. (2003), Country-Led Evaluations: A discussion note prepared by WB/OED, UNDP/EO and IOB, March 2003. Available at:

http://www.oecd.org/secure/pdfDocument/0,2834,fr_21571361_34047972_31743020_1_1_1_1,00.pdf

Development Co-operation Directorate (DCD-DAC). (2005), Paris Declaration on Aid Effectiveness: High Level Forum, Paris, February 2005. 28-29 . Available at:

http://www.oecd.org/document/18/0,2340,en_2649_3236398_35401554_1_1_1_1,00.html

Gwin, C. (1999), The New Development Cooperation Paradigm ODC Viewpoint Overseas Development Council. Available at <http://www.ciaonet.org/pbei/odc/gwc01.html>

Jobin, Denis. (Forthcoming, 2007), A Transaction-Cost-Based Approach to Partnership Performance Evaluation, Evaluation: The International Journal of Theory, Research and Practice, U.K.

Picciotto, R. (2002), Development Cooperation and Performance Evaluation: The Monterrey Challenge Washington DC: World Bank. OED Working Paper.

Available at:

[http://nweb18.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/8AB016849FCE796085256BC20065B2D7/\\$file/Monterrey_Challenge.pdf](http://nweb18.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/8AB016849FCE796085256BC20065B2D7/$file/Monterrey_Challenge.pdf)

UNESCAP What is good governance? Available at: <http://www.unescap.org/pdd/prs/ProjectActivities/Ongoing/gg/governance.asp>

JOINT COUNTRY-LED EVALUATION OF THE POLICIES RELATED TO CHILD WELL-BEING WITHIN THE SOCIAL PROTECTION SECTOR IN BOSNIA AND HERZEGOVINA

*by Azzedina Vukovic, Directorate for Economic Planning,
Council of Ministers of Bosnia & Herzegovina
and Debora McWhinney, Deputy Representative,
UNICEF Bosnia & Herzegovina*

Background

The Directorate for Economic Planning, DEP (previously the Economic Policy and Planning Unit (EPPU), which was transformed into a core body of the Council of Ministers of Bosnia and Herzegovina (BiH) in 2006, has shown its commitment to the Paris Declaration (PD) principles, particularly in working “to establish results-oriented reporting and assessment frameworks that monitor progress against key dimensions of the national and sector development strategies” (Indicator 11). The EPPU/DEP oversaw the preparation of the Medium-Term Development Strategy (MTDS) and is responsible for monitoring its implementation and for evaluating results achieved.

UNICEF also works in accordance with similar PD principles, namely to “link country programming and resources to results and align them with effective partner country performance assessment frameworks”. UNICEF’s programme cycle is a highly consultative one that requires formalised discussions and approval from government in the preparation of each Country Programme (usually 5 years in length), as well as on an annual basis. In addition, UNICEF works with all of its partners to evaluate the results achieved, mid-way through the Country Programme Cycle, in the form of a Mid-term Review (MTR). UNICEF BiH undertook, in 2007, an MTR with an increased focus on evaluating the results of work in the area of policy advocacy and partnerships for children’s rights.

UNICEF’s Regional Office (RO) in Central and Eastern Europe/ Commonwealth of Independent States (CEE/CIS), has been supporting Country Offices and their partners in enhancing overall capacities for monitoring and evaluation. In 2006, the International Devel-

opment Evaluation Association (IDEAS), contacted the UNICEF Regional Monitoring and Evaluation Advisor in search of a partnership to organize a Regional Workshop on Country-Led Evaluations. In addition to financial and technical support, the UNICEF RO encouraged Bosnia and Herzegovina (BiH) to consider applying this methodology since the EPPU/DEP was due to begin preparation of a new Medium-Term Development Strategy in 2007 (to cover 2008 – 2013), and UNICEF BiH planned a mid-term review.

A senior member of EPPU/DEP, and key UNICEF programme staff, attended the IDEAS conference in June 2006. This initial capacity building exercise ensured not only that information on the specific methodological approach was transferred, but that EPPU/DEP and UNICEF were also provided with an opportunity to identify both separate and common priorities. It was during this conference that an initial agreement on the scope of a CLE in BiH was reached.

Early discussions took place around whether this CLE would constitute an 'ex-ante' evaluation from the government's perspective. These discussions took place during the preparation for the creation of a new Development Strategy for Period 2008-2013 and first Strategy of Social Inclusion. These documents would form the basis of the National Development Plan of BiH, which was one of the documents required in the process of EU integration. In 2004, EPPU/DEP created a mechanism to monitor the progress in implementation of MTDS measures. This mechanism focused on governmental and parliamentary bodies at the level of the State and of the individual Entities of Bosnia and Herzegovina. An 'ex-ante' evaluation was not carried out at that time primarily because the process of monitoring and evaluation was developing and there was insufficient capacity for such a task. Although the human resource capacity within EPPU/DEP, and in other governmental institutions, was subsequently increased, the capacity to conduct a proper 'ex-ante' evaluation remained insufficient. As a result, EPPU/DEP considered this UNICEF-supported initiative as an important one since it provided an opportunity to initiate the development of a systematic process of ex-ante evaluation. Activities undertaken to date have involved a large number of representatives of various ministries and have shown that there is interest in developing the capacity for CLE and ex-ante evaluation. These activities have also pointed out that most participants were not familiar with either approach.

For its part, UNICEF discussed the approach internally and decided that the CLE should play an important role in informing the mid-

term review, particularly with regard to future strategic positioning in the area of evidence-based policy making.

Country-Led Evaluation is a relatively new phenomenon and increase the focus on participation, ownership and governance of the evaluation function. UNICEF has been applying participatory approaches to policy development for a number of years and saw the Country-Led Evaluation as a chance to support the strengthening of evaluation capacity within the government, as well as to enhance the analysis of UNICEF-supported activities related to evidence-based policy making in the area of child well-being. As a result, it was proposed to term this a “joint” Country-Led Evaluation as a way of highlighting the partnership between UNICEF and EPPU/DEP without detracting from the central role of government leadership and ownership of the process.

To ensure an independent and objective evaluation process and result, it was decided that an external team would be needed to guide the evaluation alongside EPPU/DEP and in consultation with UNICEF. Terms of reference (TORs) for the CLE were prepared jointly and the two-fold purpose of the evaluation was defined as follows:

- A. To provide an ex-ante evaluation for the BiH EPPU to inform and structure the production of the strategic social sector documents in 2007, including:
 - o recommendations to address the weaknesses of the system in reaching its developmental objectives; and,
 - o recommendations on policy development criteria, as well as indicators for monitoring and evaluation of social policy implementation process.
- B. To inform UNICEF’s Mid Term Review and the evaluation of the UN Development Assistance Framework (UNDAF), and to assess UNICEF’s contribution to the BiH social protection sector, including:
 - o Recommendations on UNICEF’s capacity to contribute to the development of evidence-based policies; and,
 - o Recommendations regarding the development of a more structured and coherent approach to policy development and implementation.

In addition to these broad goals, the TORs included four main objectives:

1. To evaluate the effectiveness, and ease of implementation, of selected child and family focused policies, as defined within the *BiH Medium Term Development Strategy's Social Protection Chapter*. The evaluation would thus provide an ex-ante evaluation for use by the BiH EPPU to inform and structure the production of the strategic development documents in 2007;
2. To evaluate the effectiveness of elements of the UNICEF BiH policy support activities in contributing to development of evidence-based, child-focused policies in the social protection sector;
3. To assess the implementation of Paris Declaration targets by national stakeholders and donors, including the establishment of country-led monitoring and evaluation systems; and,
4. To develop and document the CLE participatory methodology for BiH for i) its further application in the evaluation of development goals in BiH, and ii) to contribute to the development of CLEs within the international evaluation community.

The detailed Terms of Reference were publicised as a request for bids in December 2006. Ten bids were received from both individuals and companies, and a short-list was agreed upon by all parties. The short-listed candidates/companies were asked to submit a more detailed proposal. These submissions were further reviewed and a company with considerable evaluation experience was selected.

BiH capacity to evaluate national development priorities

The Office for Monitoring and Implementation within EPPU/DEP was transformed into the *Sector for Preparation of BiH Development Documents, Analysis of Social Inclusion and monitoring and evaluation*. This created significant core capacity within the Council of Ministers for the development of evaluation processes. This sector will also be in charge of coordinating the preparation of a new *National Development Strategy* for 2008-2013, as well as the first *Social Inclusion Strategy*. Following the preparation of these documents, the sector will be responsible for monitoring the implementation of these strategies. The methodology for country-led and ex-ante evaluations are therefore of crucial importance for

the sector, as well as for implementation by other governmental bodies.

A large number of local and international experts, representatives of relevant ministries at state and entity level, as well as from civil society organizations were involved in the process of developing the MTDS. As this was the first time that BiH had prepared such a comprehensive development document, some improvements were needed. In particular, prioritisation of various measures in the *MTDS Action Plan* was lacking. This weakness was realized very soon after implementation began and one of the most significant corrective measures identified was the need for capacity building in the area of evaluation of national development priorities. A number of training sessions were organized for EPPU/DEP personnel, as well as for members of the working groups which supported the monitoring and implementation of MTDS.

This enhanced capacity led to improvements in the preparation of a revised MTDS document; better prioritization of measures; and, the definition of national priorities according to the three main goals of the MTDS. Support from the international community in this process was significant but, the number of international experts involved in the creation of the revised MTDS was significantly reduced in comparison to the creation of the original MTDS. This fact, as well as better prioritization of measures, showed that BiH had made important progress in the area of evaluation of national development priorities.

In spite of this significant progress, BiH still has insufficient capacity for evaluation processes at all levels of government. Strengthening these capacities must continue in order to achieve effective monitoring and evaluation of national developmental goals.

The EPPU/DEP will disseminate the evaluation results widely among stakeholders in different Ministries and at different levels of government. This serves to raise awareness of the evaluation itself, the methodology used, and the leadership shown by EPPU/DEP in the area of evaluations. Further, the methodology for conducting ex-ante evaluations, which was developed during this process, will be applied to the process of preparing a new MTDS and *Social Inclusion Strategy*.

UNICEF support to monitoring and evaluating national development priorities

The joint CLE further strengthened the existing partnership between UNICEF and EPPU/DEP in the area of strengthening national evaluation capacities. On-going, project-based collaboration was deepened through UNICEF's commitment to providing opportunities for DEP to enhance and increase its leadership and skills in the evaluation of development goals. Whereas monitoring the implementation of development goals is extremely challenging in BiH, given the decentralised administrative structure and de-linked policy development and implementation functions, this joint evaluation process provided DEP with an opportunity to exhibit further leadership on the evaluation of development goals. This is critical in general, but particularly so given that the process of preparing a new *Medium Term Development Strategy (2008–2013)* and *Social Inclusion Strategy* began in 2007.

UNICEF's commitment to consultative, participatory and transparent approaches to programming was further enhanced by the choice of methodology and consultants which prioritized local ownership and experience. The combination of Reference Group meetings; thematic workshops; individual interviews; and, a case study allowed for information gathering through a range of modalities. The thematic workshops proved particularly successful in engaging a range of actors in a dialogue on issues related to Objectives 1 and 2 respectively.

Given that UNICEF BiH's current Country Programme (2005-8) was designed according to a human rights-based approach, it was important to evaluate the relevance, effectiveness, efficiency, sustainability and result of initiatives targeted at different, but inter-related levels: policy, services and community. In order to maximise the impact of the evaluation of UNICEF's contribution to the development of evidence-based, child-focused policies using different approaches, UNICEF selected the following projects as targets for the evaluation in Objective 2:

- i. Data collection, research and policy analysis on children and women (Iodine Deficiency Disease Strategy Development).
- ii. Inclusive basic services (Human Rights-based Approach to Child Protection).
- iii. Community-based activities (Participation Action Research Groups).

The detailed analysis of these three interventions will be used to directly inform the Mid-term Review (MTR) of UNICEF's Country Programme (2005-8). The answers given to the following questions will also provide UNICEF with considerable information on which to base its decision-making regarding future strategic approaches, leadership areas and activities in the context of the MTR:

- How effective and relevant is the UNICEF research, data collection and child rights monitoring programme to the development and implementation of programmes and policies in the child protection sector in BiH?
- How effective are the partnerships and coordination mechanisms established between community/service delivery/policy development (as defined in the structure of the UNICEF Country Programme), within UNICEF's programme of support to the social protection of children?
- How relevant are UNICEF's policy support projects in relation to the national priorities, as well as to the international development agencies priorities?
- To what extent are the UNICEF-defined results harmonised with the nationally-defined results in the social protection/child protection field?
- What is the relevance and efficiency of UNICEF's policy support activities for direct duty-bearers and rights-holders in communities?
- What were the barriers to implementation of the UNICEF-supported policy measures at various levels?

Answers to these questions provide a complex picture of the results of UNICEF's efforts to promote national ownership, participatory approaches and capacity development in the area of child well-being.

Main achievements

1. Enhanced accountability and responsibility on the part of the EPPU/DEP with regard to monitoring and evaluation for the MTDS, particularly in the area of child-related policies

As has been described, the CLE provided a strategic opportunity for EPPU/DEP to demonstrate increased leadership in the field of monitoring and evaluation of national development strategies. This

leadership role had already been defined given the work done in the past on the creation of the MTDS, however, leadership in the context of BiH is rarely straightforward. The decentralised nature of the governmental structure means that while EPPU/DEP developed the MTDS in a participatory manner, they were not responsible for its implementation as this lay with lower levels of government. Therefore, the importance of strengthening EPPU/DEP's accountability and responsibility in monitoring and evaluation of national developmental policies in relation to other ministries and governmental bodies cannot be underestimated.

The EPPU/DEP's leadership in the CLE was strategic as that same year, 2007, they began the process of preparing a new MTDS, Social Inclusion Strategy and National Development Plan. Their ability to apply the lessons learned "by doing" in the joint CLE process would prove to be particularly valuable at this time.

2. Inter-relatedness of objectives in the CLE serving multiple strategic intents.

The terms of reference for the joint CLE were multi-faceted. Rather than simply evaluating the effectiveness, relevance, efficiency, sustainability and impact of one specific policy area, the decision was made to combine an examination of child and family-focused policies as defined by the MTDS with an evaluation of the effectiveness of the contribution of certain UNICEF-supported interventions to evidence-based, child-focused policies as defined in the same document. This dual approach allowed for an evaluation of governmental and UNICEF interventions both individually and, more importantly, the interaction between them. The focus on evidence-based policy making in the area of child well-being, allowed the highlighting of functional and accountability gaps affecting implementation and monitoring by both the government and UNICEF. Strengths and weaknesses of the policy development and monitoring and evaluation approach taken by both EPPU/DEP and UNICEF were also identified.

Further objectives related to the implementation of Paris Declaration targets by national stakeholders and donors, as well as the critical component of documenting the methodology used in the joint CLE for its further application in BiH. This is an innovative addition to existing studies, functional reviews and processes in the country. With EPPU/DEP leadership and strengthened capacity, similar processes could be carried out in the future.

Lessons learned

The Joint CLE in Bosnia and Herzegovina was among the first experienced worldwide. For this reason, special attention was given to learning lessons during the process, and to document them. Three main lessons have emerged:

1. Ensure common understanding and buy-in regarding the role of the Reference Group (RG) at the earliest possible stage.

A Reference Group was created with the objective of “ensuring the perspective of key stakeholders in social and development outcomes in BiH are included in the evaluation design, analysis and recommendations.” The following Terms of Reference for the RG were created and disseminated to all members:

- The CLE RG is to meet at least twice during the CLE process:
 - o At the initial inception meeting the RG is to provide feedback on: the design and implementation of the CLE; the relevance of CLE within wider reform process and the Paris Declaration; potential use of CLE as a capacity development tool; and, on mobilization/advocacy potentials of the CLE.
 - o To give feedback on the draft evaluation report, proposed follow-up actions, and dissemination and use.
- Interested members of the RG will be contacted during the CLE process for individual consultations, feedback and information dissemination.
- The CLE RG will propose a final dissemination plan and will provide specific feedback on the future steps in CLE findings analysis and use.
- The CLE RG will serve as the forum to discuss ongoing social sector/social outcomes evaluations implemented by other institutions and agencies and will potentially act as RG for these evaluations under leadership of government institutions.

In preparing for the first Reference Group meeting, insufficient time was given to informing the RG members of the CLE process and content. Further, the first RG meeting consisted of a number of presentations, many of which were quite abstract and overly emphasised the prominence of UNICEF-supported interventions. Also, EPPU/DEP saw the potential to use the Reference Group for

further capacity building in the area of evaluations. This element was introduced to the group before a clear understanding of the role of the group in this particular CLE process had been reached. These factors led to fairly negative reactions from many of the participants and fairly weak commitment to the Group as neither the information presented nor the purpose of the Group were clearly defined.

Some of these issues were addressed in the meeting itself, such as clarifying that the main aim of the CLE was to evaluate governmental policies in the area of child well-being, rather than to focus primarily on UNICEF (supported) activities. Information on the CLE was shared with RG members following the meeting, but the lack of clarity about roles continued until the end of the process. The tactic used by EPPU/DEP, UNICEF and the consultancy company was to clarify the process at each step, as well as to concentrate on the substantive information in order to gather feedback and comments. However, this also proved challenging as there were very few core RG members, since the individuals nominated by their Ministries for each of the three meetings were often different.

The Reference Group plays an important role in CLEs, not only in the process itself, but in the strengthening of national capacities in the area of evaluation. However, adequate preparation of the individuals to be members of the Group is critical, as is a clear explanation and discussion of the purpose of the CLE prior to initiating presentations.

2. Ensure that roles and responsibilities of key stakeholders are clear and understood.

The joint CLE in BiH proved to be a complex one given the range of actors involved. Given the nascent evaluation capacity in the country, UNICEF and EPPU/DEP felt that it was important to engage a strong consultancy company with experience in this area to oversee the process, provide strong technical inputs and ensure independence/objectivity.

A consultancy team was built, but suffered early on from varying levels of understanding of the task, differences of commitment and, ultimately, the resignation of one of the key members. The consultancy company responded quickly and effectively to this event, but the definition of roles within the team remained unclear and somewhat contentious throughout the CLE process, particularly with the local consultants. UNICEF emphasised from the outset the need to hire strong local consultants and liaised with the consultancy com-

pany in this process. However, as individuals who knew UNICEF's work were hired, contact often went through UNICEF, albeit informally, rather than through the consultancy company. UNICEF was required to play a negotiating role on a few occasions.

In addition to relationships within the consultancy team, there was also the additional challenge of the BiH administrative structure with respect to ensuring adequate and appropriate representation of individuals from various levels of government. Some key actors (e.g. Deputy Ministers of Social Welfare in both Entities) were asked to participate in the Reference Group and the three thematic workshops, and were asked for individual interviews. In other cases, representatives from Cantons, the Entities and Brcko District were invited along with NGOs, members of statistical institutes and research agencies, in order to gather experiences from the most relevant range of actors.

The lesson learned is that it is critical to allow sufficient time for the development of the consultancy team; definition and clarification of roles; and, agreement on the form that the main elements of the CLE will take. The more time that the consultancy team can spend in country, the more smoothly this process will run.

3. Identify functional weaknesses at the outset and work to mitigate them.

The weaknesses inherent in the administrative system in BiH (e.g. high level of decentralisation and multiple layers of government; weak vertical communication; lack of data in some key areas; little connection between macro goals and local planning; limited implementation capacity; etc.) were on the one hand the subject of the evaluation (Objective 1) and also the cause of some key limitations.

This challenge was overcome by utilising a variety of evaluation methods: thematic workshops; individual interviews; case studies; and, focus groups, with individuals working at various levels of government. The lack of data in certain areas made it impossible for firm conclusions to be made, but experiential evidence and testimonials were used to the greatest extent possible.

Conclusion

The EPPU/DEP and UNICEF are equally committed to supporting participatory policy development processes. The joint CLE has served to provide EPPU/DEP with further opportunities to strengthen their leadership and ownership of the evaluation function in the country, particularly as it relates to child well-being in the context of national development strategies. EPPU/DEP exhibited leadership in all stages of the process, from development of the TORs to suit their specific information and capacity development needs, to the adaptation of the role of the Reference Group and strategic decision-making. UNICEF was able to reaffirm its commitment to the Paris Declaration principles by exhibiting its commitment to country-led evaluation processes. The joint CLE also allowed UNICEF to enhance its strategic influence as an advocate for children's rights in the policy context and to gain recommendations on ways to improve its programming for the remainder of its Country Programme Cycle.

The CLE provided UNICEF with a rich source of information on the effectiveness of selected activities in contributing to the development of evidence-based, child-focused policies in the social protection sector. This information will prove critical for the process of re-aligning the UNICEF Country Programme and for future leadership areas.

The joint CLE in Bosnia and Herzegovina can be considered a success given its adherence to the Paris Declaration principles of ownership; alignment; managing for results; and, mutual accountability. The evaluation capacity of the government has been enhanced, as has the ownership by the EPPU/DEP over this critical process.



Part 3

The strategic intent of data collection and dissemination

The strategic intent of data collection and analysis. The case of Multiple Indicator Cluster Surveys (MICS), by Daniel Vadnais and Attila Hancioglu	168
The strategic intent of data dissemination. The case of DevInfo, by Nicolas Pron, DevInfo Global Administrator, Division of Policy and Planning, UNICEF Headquarters	185
Using DevInfo as a strategic tool for decision making. Achievements and lessons learned in Moldova, by Mohamed Azzedine Salah, Deputy Representative, UNICEF Moldova	195
Using DevInfo to support Governments in monitoring National Development Strategies. The case of the Republic of Serbia, by Dragana Djokovic-Papic, Statistical Office of the Republic of Serbia, and Oliver Petrovic and Vladica Jankovic, UNICEF Serbia	200
Using DevInfo as a strategic tool to facilitate local communities' empowerment. The case of the Municipality of Pirot, by Vladan Vasic, Mayor of Pirot, and Oliver Petrovic and Vladica Jankovic, UNICEF Serbia	203

THE STRATEGIC INTENT OF DATA COLLECTION AND ANALYSIS. THE CASE OF MULTIPLE INDICATOR CLUSTER SURVEYS (MICS)

By Daniel Vadnais and Attila Hancioglu

MICS: A chronological overview

Population-based social surveys were originally implemented in the first half of the 20th century. In the area of demography and health, the first surveys focused on fertility and family planning issues. From the 1960s to the mid-1980s, three global household survey projects were implemented. These provide an international body of comparative information. The first was the Knowledge, Attitudes and Prevalence (KAP) survey on family planning. This was followed by the World Fertility Survey (WFS) and finally, the Contraceptive Prevalence Survey (CPS). Each survey overlapped with the previous one. Not until 1984, with the implementation of the Demographic and Health Survey (DHS) programme, were maternal and child health issues systematically added.

Household surveys have been used in situations where vital registration data of good quality (such as those on birth or death registration) are lacking. They are also used to give countries with no reliable or regular census data the necessary information they need for planning purposes, for setting up programmes and for implementing policies for the well-being of their populations. With few exceptions, the technical assistance provided to countries, to carry out household surveys, has largely been funded and offered through international agencies. This has made it possible for countries facing financial problems; absence of infrastructure; lack of human technical capacity; wars; conflicts; or natural disasters, to continue to obtain necessary statistical information on a regular basis.

In 1990, participants in the World Summit for Children adopted a set of goals to promote the rights and welfare of children. At that time, it was recognized that many countries lacked the capacity to accurately measure progress toward these goals. UNICEF responded by developing the Multiple Indicator Cluster Survey (MICS). The MICS is an international household survey initiative designed to assist

countries in filling data gaps for monitoring human development in general and, in particular, the situation of children and women. Since 1995, MICS has been conducted every five years. Each round of surveys builds upon the last and offers new indicators, to monitor current priorities, in addition to monitoring trends. During the third and latest round of MICS (referred to as MICS3) which began in 2005, UNICEF added several new indicators to track progress toward the Millennium Development Goals (MDGs), and other major international commitments. Almost half of the MDG indicators are collected through MICS, making them one of the largest single sources of data for MDG monitoring.

MICS3, carried out in more than 50 countries in 2005 and 2006, is generating data representative of close to one in four children (23 percent) living in developing countries or almost two in five children (38 percent) if India and China are excluded¹. Since the initiation of the programme, nearly 200 MICS have been implemented in approximately 100 countries. MICS is generally carried out in countries around the world where recent data from other household surveys, such as DHS, is not available. In fact, to prevent duplication of efforts and resources, UNICEF discourages countries from implementing MICS if other recent and comparable information on children and women already exists. MICS and DHS surveys use very similar methodologies for data collection and analysis, and are largely harmonized in terms of content. Currently, DHS surveys produce comparable information on close to three-quarters of the MICS indicators.

When combined, MICS and DHS surveys provide reliable and up-to-date data for a majority of developing countries. A complete list of countries where MICS surveys have been implemented since 1995 is given in the table below².

1 The number of children (aged below 18) living in MICS3 countries is about 450,000,000 out of approximately 1,959,000,000 living in developing countries. If we exclude India and China, this number is 1,165,300,000. In percentage terms, this means that MICS3 has collected information on 22.8 percent of children living in developing countries (or 38.3 percent if we exclude India and China). This information comes from: *The State of the World's Children 2008*. UNICEF, in press.

2 For a detailed list of surveys carried out under the DHS programme since 1984, go to: <http://www.measuredhs.com/aboutsurveys/>

MICS surveys: 1995-2007***Eastern and Southern Africa***

Angola	Malawi
Botswana	Mozambique
Burundi	Rwanda
Comoros	Somalia
Ethiopia	Swaziland
Kenya	Tanzania
Lesotho	Zambia
Madagascar	

Western and Central Africa*

Burkina Faso	Guinea-Bissau
Cameroon	Liberia
Central African Republic	Mali
Chad	Mauritania
Côte d'Ivoire	Niger
Equatorial Guinea	Nigeria
Gabon	Sao Tome and Principe
The Gambia	Senegal
Ghana	Sierra Leone
Guinea	Togo

Central and Eastern Europe and Commonwealth of Independent States*

Albania	Moldova
Azerbaijan	Montenegro
Belarus	Serbia
Bosnia & Herzegovina	Tajikistan
Croatia	Turkey
Georgia	Turkmenistan
Kazakhstan	Ukraine
Kyrgyzstan	Uzbekistan
Macedonia	

<i>East Asia and the Pacific</i>	
China Indonesia Korea DPR Lao PDR Mongolia	Myanmar Philippines Thailand Vanuatu Vietnam
<i>Latin America and the Caribbean</i>	
Belize Bolivia Cuba Dominican Republic Ecuador Guyana	Jamaica Panama Suriname Trinidad & Tobago Venezuela
<i>Middle East and North Africa</i>	
Algeria Djibouti Egypt Iran Iraq Lebanon Libya Morocco Oman	Occupied Palestinian Territory Palestinians in Lebanon Palestinians in Syria Sudan Syria Tunisia Yemen
<i>South Asia</i>	
Afghanistan Bangladesh India Maldives	Nepal Pakistan Sri Lanka

**Zaire and Yugoslavia conducted a survey in the first round of MICS*

What does MICS offer?

MICS is the largest source of statistical information on children. It produces statistically sound, internationally comparable estimates of social indicators such as those required for monitoring the goals and targets of the Millennium Declaration; the World Fit for Children Declaration and Plan of Action; the goals of the United Nations

General Assembly Special Session on HIV/AIDS; and, the African Summit on Malaria. Some important features which the MICS programme has to offer are:

- Randomly selected and representative data at the national, urban/rural and sub-national level.
- Statistically sound data often unavailable through other data collection tools.
- Useful information on household living conditions.
- Disaggregated data by sex, education, wealth and residence useful to highlight geographic, economic and social disparities.
- Quality statistics emanating from individual interviews with women of reproductive ages, 15 to 49 years.
- Information on children under five collected from mothers or primary caretakers.
- Flexibility to suit country specific requirements (through the use of modules).
- Comparability of data over time and across countries as the methodology, sampling procedures and questionnaires used throughout the world are standardized.

The third round of MICS has added information on child protection issues not readily available in household surveys, such as those on child discipline, child labour, early childhood development, and child disability.

MICS data are widely used for international reporting requirements. They are also critically needed for policy advocacy at the national level to improve the lives of children and women. Country-specific data is added to www.childinfo.org as it becomes available.

Strengths and limitations of MICS

The strengths and limitations of MICS have a direct impact on how they can be used to provide information for social policy planning and budgeting, and for monitoring and evaluating programmes and policies for children and mothers. MICS produces quality data that can be used as evidence by local, national and regional decision makers for the implementation of policies and programmes for targeted population groups, such as women and children. The surveys are also very useful for monitoring and evaluating existing policies and programmes in order to highlight progress and challenges.

In the absence of quality civil registration systems and/or reliable routine data collection tools, MICS provides countries with an ideal tool to collect much needed statistical information. Yet, several countries are faced with shortcomings due to a lack of funding, mismanagement or inadequate staffing. Personnel may be inadequately trained to meet global standards and requirements for the acquisition of quality data. In fact, many societies still do not have in place a dependable universal health management information system with a vital statistics structure. These factors often make population-based studies the only means for collecting accurate data for the use of policy and decision-makers.

MICS is an invaluable resource as it:

- Helps track progress on government commitments to attain national and international goals.
- Can be used to assess the financial needs required for the provision of services such as building more schools, day care facilities, health clinics, youth centres, and the like.
- Can assist countries in filling data gaps for monitoring human development in general and the situation of children and women in particular.
- Offers data that can be used to calibrate or validate information from routine reporting systems granted these systems provide complete and reliable information.
- Is able to produce high quality, in-depth information on causes, correlations and consequences of a variety of social, demographic and health processes, since data collection is conducted with well-trained interviewers and with in-depth, well-structured questionnaires.
- Contributes to increasing a country's institutional and human capacity to conduct household surveys, to analyse data and to disseminate statistical information.

Key limitations of the MICS are:

- It cannot be implemented in a quick turnaround. The first two rounds of MICS were quickly implemented and disseminated. With the adoption of more sophisticated survey tools and approaches in the third round, as well as more emphasis on standardization of survey reports and an intense review process, MICS now requires more than a year for the data to be properly analyzed and made publicly available.

- It can be difficult to administer when a large number of modules are added to the core survey or when data has to be collected at a level below sub-national. These two decisions often end up significantly increasing the implementing costs.
- It is confined to populations living in households, thus excluding institutional and non-household populations (such as individuals living in elderly or orphan homes, street children, sex workers, children in detention, etc).
- It can be a relatively expensive data collection tool (per unit of observation) to put into place. MICS tends to be less costly than DHS as UNICEF can rely on its own regional and country presence, all over the world, to offer technical assistance and regular follow-up.

In a recent article published in *The Lancet*, Boerma and Stansfield argue that “National surveys can be costly, and substantial investments have to be made to ensure data quality. Yet a comparative analysis of different sources of health data showed that household survey costs per capita are often lower than other health data sources. The main limitations of surveys include the inability to disaggregate at local level and to provide information at short time intervals. Household surveys also can be misused to obtain information on topics for which there are no valid and reliable questions or tests”.³

In an effort to reduce the implementing costs of their surveys, MICS encourages countries to only select a limited number of thematic modules.

Frequency of MICS surveys

Until now, MICS has typically been implemented at the country level every four to six years. Starting with the fourth round of MICS, UNICEF is now working on a plan to offer technical assistance every three years to countries interested in collecting data more frequently. This should increase the countries’ ability to capture rapid changes in key indicators related to child survival. It will also provide more rapid feedback, to policy-makers and other users, to better fine-tune policies and programmes on behalf of children and mothers. This is especially important in those countries where child survival is the priority issue and fast changes may be taking place in the determinants of early age mortality.

3 Boerma T, Stansfield S. K. Health statistics now: are we making the right investments? *Lancet* 2007; 369: 779-86.

The strategic intent of MICS in data collection and analysis

Thanks to its modular approach, MICS is well suited to provide countries with the relevant amount and type of information they need to put into place, or to monitor and evaluate, nationally or sub-nationally relevant policies and programmes.

While relying on a standardized approach, MICS provides enough flexibility for countries to add modules which can help them answer questions specific to their socioeconomic, cultural or geographic circumstances. In the third round of MICS, for instance, modules were offered on security of tenure; early childhood development; sources and costs of supply of anti-malarials, or of oral re-hydration salts and antibiotics; child discipline; domestic violence; disability; or maternal mortality.

Conversely, MICS gives countries the possibility to remove questions and/or sections of the core questionnaires which are irrelevant to them. For example, in case where countries have implemented recent research covering similar topics, or in cases where specific topics are irrelevant to those countries (for instance, in places where malaria is not a problem).

Highlighted below are other specific advantages offered by MICS, with an emphasis on countries from the Central and Eastern Europe/ Commonwealth of Independent States (CEE/CIS) region.

MICS helps strengthening national statistical capacities

In addition to yielding valuable information, the actual planning and field implementation of a survey such as MICS, and the use of the results, contribute to strengthening current and future national monitoring capabilities. MICS emphasizes the importance of involving, at all levels of the survey implementation, personnel from national institutions, such as statistical offices; medical and public health schools; education and training institutes; and, university departments in statistics and social sciences.

For these reasons, MICS is typically carried out by government organizations, with the technical support and financial aid of UNICEF and its partners. UNICEF facilitates its assistance and training through a series of regional workshops (limited to four workshops per region) during which all countries in a region gather for five to

seven days at a time. These workshops discuss questionnaire content; sampling and survey implementation; data processing; data quality and data analysis; and, report writing and dissemination. This is complemented by the provision of user-friendly documentation and on-going technical support from headquarters' staff, on all steps of the survey, from fieldwork to the presentation of data. During the third round of MICS, more than 300 experts from participating countries were trained in survey methodology worldwide.

MICS, as an alternate data source

Although MICS data on childhood mortality may not be free of problems, it has been increasingly used in the countries of the CEE/CIS region for collecting infant and under-five mortality estimates. This information is used to complement official data from administrative sources as well as to further assess the quality of mortality data obtained through routine data collection.

Several countries in the region are considered to have been underestimating the levels of child mortality for two basic reasons: the use of the Soviet definition of live births and the under-registration and misreporting of infant deaths.

The discrepancies between data from vital registration systems and survey data have given rise to concerns regarding the reliability and international comparability of infant mortality data from the vital registration systems in some countries of the region, especially in Central Asia and the Caucasus. Without a reliable routine registration system for vital statistics, policy-makers and analysts lack a regular source of timely information on infant and child mortality. Such information would allow them to evaluate the effectiveness of current policies, the quality of health management and provision of care, and to identify challenges and inequalities and draw up appropriate policy measures and actions aimed to tackle them.

The Inter-Agency Group for Child Mortality Estimation, which includes UNICEF; WHO; the World Bank; the United Nations Population Division; the Latin American and Caribbean Demographic Centre (CELADE); Harvard University and the US Bureau of Census, scientifically combines all available mortality estimates in a country (such as vital registration systems, censuses and surveys) to produce harmonized current estimates and trends. Together, data from these various sources are analyzed and their quality is scrutinized. Such an approach minimizes the errors associated with each individual estimate and helps harmonize trends across time. MICS data is vital for this exercise.

Using the Inter-Agency methodology, infant mortality in the countries of Central Asia was estimated to be much higher than indicated by the vital registration systems.

In Turkmenistan, for example, the 2005 infant mortality rate was estimated to be 81 per 1,000 live births while data from the vital registration system showed a level of 14 per 1,000, or almost six times less. Uzbekistan had a similar situation with a rate of 57 per 1,000 in 2005 compared to 15 per 1,000, or about four times less, by the vital registration system.

Summarizing the key differences between both definitions, Charyeva, V.R., Samarkina, E.Y. and Sullivan J.M. wrote in the 2000 Turkmenistan DHS final report:

“The most important difference is for pregnancies ending at a gestational age of less than 28 weeks. The Soviet protocols classify such pregnancies as miscarriages (even if signs of life are present at the time of delivery) unless the child survives for seven days. On the other hand, the World Health Organization defines a birth showing any sign of life (i.e., breathing, beating of the heart, or movement of voluntary muscles) as a live birth, irrespective of the gestational age at termination of the pregnancy (WHO, 1993).

A second difference between the Soviet protocols and WHO’s definition concerns pregnancies ending at 28 or more weeks of gestation. According to the definition of the Soviet protocols, these events are classified as live births if the child breathes and as still-births if breathing is not evident at delivery. The World Health Organization defines these events as live births if any sign of life is present at delivery (i.e., breathing, beating of the heart, or movement of voluntary muscles) and otherwise as still-births”.

The problems of under-registration and misreporting of infant deaths by parents and/or medical officials are still frequent in the region either because of lack of knowledge, or for deliberate actions. Parents, for instance, may be discouraged from registering the birth and death of their child for financial reasons (e.g. registration fees, cost of transportation to reach administrative offices). Health officials, on the other hand, may still want to conceal infant deaths for fear to be held responsible or to be accused of negligence in countries where child mortality has traditionally been used as one of the indicators to evaluate the performance of health facilities.

The implementation of household surveys will not necessarily resolve the problems countries in the CEE/CIS region face in terms of accurately measuring child mortality rates. What it can do, though, is stimulate a national debate on the reasons why these discrepancies exist, and what should be the strategy to improve the situation.

MICS plays a key role in the monitoring of global and national commitments made to children

As mentioned earlier, MICS was originally designed to assist countries in filling data gaps for monitoring human development in general and the situation of children and women in particular. Starting with the third round of MICS, a number of indicators have been added to those already focusing on the World Summit for Children and the Millennium Summit. These new indicators stem from other international commitments made to children, ranging from those such as in the 2001 United Nations General Assembly Special Session (UNGASS) on HIV/AIDS and the 2002 World Fit for Children, to the 2005 Abuja Targets for controlling and gradually eliminating malaria. In addition to providing data for reporting on progress towards global and national commitments, MICS also responds to the needs of interagency monitoring groups such as those working on MDG indicators related to water and sanitation; malaria; AIDS; immunization; and, child survival. Indicators can be added as long as they are internationally approved and do not compromise the data quality of other indicators.

MICS has expanded over time to cover emerging topics while consistently monitoring, since 1995, several key indicators on nutrition; child mortality; child health; environment; education; and reproductive health.

MICS fills data gaps

UNICEF developed MICS with the aim of assisting countries in filling data gaps for monitoring human development in general and the situation of children and women in particular. A good example of this can be found in the 2007 issue of the State of the World's Children, in which UNICEF presents the most recent data available (mainly from MICS and DHS) on the percentage of households consuming iodized salt. The findings clearly show that the region of Central and Eastern Europe/Commonwealth of Independent States and Baltic States lags behind all other regions of the world, ranking just below South Asia.

In order to ensure the virtual elimination of iodine deficiency disorders (the world's single greatest cause of preventable mental retardation, and which causes goitre and cretinism and significantly raises the risks of still-births and miscarriages) it is necessary to collect data to identify sub-national areas most at risk.

Whereas data is available from salt producers, it does not help decision makers to plan policies and programmes to improve the actual consumption of adequately iodized salt at the household level. In fact, there are many regions of the world which produce salt in large quantities, yet the populations living in their vicinity often do not consume the salt because it is reserved for export, or if they do, the salt may not be properly iodized. That is why salt consumption data from household surveys such as MICS and DHS is so useful.

Through the use of a rapid testing kit, results can immediately tell researchers if the salt used by a household is within the internationally agreed standardized cut-off point (15 parts per million) or not. Analysis of the data can also inform researchers if the iodized salt intake is distributed evenly throughout a country or not.

MICS helps measuring new or emerging issues

MICS surveys serve as a natural media for collecting data on issues which are either new to the development agenda or which have recently recaptured people's attention. Due to their particularities, censuses, registration and observations through sentinel surveillance sites are not the most appropriate tools to measure new indicators or to collect certain types of data, for at least two reasons:

- (a) some types of information, such as those on attitudes of the population on disciplining children, on early child development, and similar psycho-social processes, are best collected through household surveys;
- (b) in some cases when a new, emerging issue is of concern to policy makers and the like, MICS is an excellent tool to test new approaches before they are incorporated into larger-scale, more expensive data collection tools such as censuses.

In the third round of MICS, for instance, there are several modules and indicators which have been included to provide information at the global level for the first time. A good example is the module on child discipline, for which mothers/caretakers of children age 2-14 are asked whether the child has been subject to various forms of discipline, either physical, psychological, violent or non-violent,

by other members of the household, and whether the mother/care-taker thinks that children should be physically punished. Such information is very difficult, if not impossible, to collect through censuses or observation.

MICS sheds light on disparities

Household surveys are useful in providing detailed data on disparities, such as in wealth, education, age, sex, urban or rural residence or other variables. When used with appropriate sampling strategies, such surveys can also assess possible correlations between these disparities.

Until the 1990s, the majority of DHS surveys, as well as other similar household survey programmes, were not in a position to produce information on disparities based on wealth, income or expenditure. Information on the availability of household possessions and its correlation with other demographic and health outcomes was relatively unknown, or was simply explained by other characteristics, such as employment or education, which were thought to be closely correlated to, or good proxies of, wealth. With the advent of the DHS wealth index in the 1990s (which used information on household assets and amenities to construct an index of wealth) it became possible to divide the household population into quintiles and produce separate estimates of demographic and health outcomes for the poorest 20 percent of the population. Researchers are now able to show, for instance, the close relationship of poverty with mortality, nutrition, and health outcomes. MICS and DHS surveys now produce data for the five wealth quintiles of household populations, and produce data useful for highlighting rich-poor differentials.

MICS can also be used to show differences in demographic and health outcomes, for various socio-cultural groups, therefore identifying vulnerable population groups in the society and making it possible to develop strategies to extend the benefits of services to all those eligible. In the third round of MICS, several countries have designed their survey to allow for this type of information. Serbia, for instance, has used an innovative sampling approach to oversample the Roma population (which would otherwise be impossible to do because of the low proportion of the Roma population in the general population) for the production of separate estimates of this excluded group. The Serbian MICS has found that among Roma population, under-5 mortality is five times higher than the national average. It also found that while 20 percent of the Roma children are stunted (an indicator of chronic malnutrition), the correspond-

ing figure for the rest of the population was five percent. Based on these results, it is hoped that there will be more emphasis on inclusive policies which specifically target excluded children.

The role of data in evidence-based policy-making: the case of selected national MICS surveys around the world

The role of data in evidence-based policy-making is illustrated from around the world by the following examples.

TAJIKISTAN: The MICS survey conducted in Tajikistan in 2005 was the second MICS to be implemented in the country. The survey raised concerns over the status of children despite recent progress made in the country's economy and in poverty reduction. Of particular concern is the nutritional status of children in Tajikistan with high levels of chronic malnutrition compared with those from other countries in the region. Recent trends in secondary school enrolment rates are also disturbing and there are signs that the quality of education and its relevance are declining. MICS findings also show that girls remain less likely than boys to enrol in secondary education. Based on the survey findings and results of the 2000 World Bank's TLSS (Tajikistan Living Standard Survey), UNICEF published a Child Poverty Study measuring both material and non material child poverty. The study provided evidence-based information that child poverty is greater than adult poverty. Both MICS and the Child Poverty Study findings were shared with the PRSP (Poverty Reduction Strategy Paper) Monitoring Unit of the government of Tajikistan at the time the government was developing its second PRSP (2007-2009). The PRS provides a unique opportunity to develop a much more comprehensive approach to investing in human capital, especially in children and young people, linking objectives to a clear budgetary framework, and ensuring that the next generation is equipped to contribute to national progress and sustained economic growth. MICS data provided direct value in helping to identify policy options and investments needed to address poverty reduction, especially from non-income/social aspects of poverty.

THAILAND: In the first MICS survey to be completed in the third round, Thailand stands out as a unique country in a number of respects. In addition to providing national and regional estimates on MICS indicators, the survey set out to produce indicators at the provincial level, and therefore, required a very large sample (more

than 40,000 households). For many of the indicators, not only was Thailand producing information at the national level for the first time, but it was also producing provincial estimates for many more indicators. It is also worth mentioning that Thailand relied on a well-defined dissemination plan right from the early stages of the survey process. Combined with the MICS large sample size, the strategy allowed Thai survey analysts to produce and disseminate estimates down to provincial level. These have been instrumental in influencing local policy makers by highlighting disparities throughout the country.

MALAWI: in 2004, Malawi conducted a DHS survey and two years later, a MICS survey was implemented. A two-year period is usually insufficient to detect significant changes, however, the rationale for Malawi policy-makers to undertake a MICS was that the DHS had only produced estimates at the regional level, whereas they needed data at the district level. Sampling specialists were able to draw a large sample capable of producing estimates for selected indicators at the district level. This decision also resulted in higher survey costs. As in the case of Thailand, it is hoped that Malawi's efforts will turn MICS data into policy messages, and actions. Early dissemination plans should help make that a reality.

IRAQ: Household surveys may be the only option to collect information for monitoring the development of countries where censuses are not possible and registration systems are no longer functioning, such as in Iraq. The last census in that country took place in 1997 and the registration system, which did not have universal coverage even before the recent conflict, is currently not functioning. Contrary to what many would expect, Iraq is today a country where there is an abundance of household surveys, such as the Iraq Living Conditions Survey (ILCS), the Iraq Child and Maternal Mortality Survey (ICMMS), a World Health Organization Survey, etc. In 2006, Iraq decided to carry out a MICS survey with the technical support of the DHS programme. Although the final results of the survey were not available at the time of writing this article, the Iraq MICS preliminary findings seem to indicate that the mortality levels are in synchrony with those obtained through the ILCS. These were previously considered to be too low. They are also much lower than the mortality estimates previously made available through the ICMMS. In fact, for years, the ICMMS results were used as the mortality estimates for Iraq, and were extrapolated to estimate mortality levels until 2005. MICS data on mortality are sure to open new avenues for discussion in, and outside, Iraq.

The role of MICS data analysis in providing further evidence-based information

By leading and playing an active role in interagency monitoring groups, UNICEF advances data analysis through the development of new methodologies, indicators, and tools. In that context, MICS data play a crucial role in UNICEF's capacity to properly monitor and evaluate the situation of children and women worldwide. UNICEF has developed new methodologies to measure standard and new indicators on water and sanitation, maternal mortality, low birth weight, and under-five mortality. With its local and international partners, including DHS, UNICEF develops joint estimates and harmonizes global monitoring efforts. This includes ensuring the comparability of estimates of key indicators on malaria, HIV and AIDS, and other priority issues, many of which are collected through MICS surveys. UNICEF maintains a series of global databases on key indicators. The databases, updated annually through a rigorous process facilitated by the vast network of UNICEF field offices, are found at www.childinfo.org, where all MICS-related information is presented.

Further analysis is necessary to comprehend changes over time, to identify the reasons why some population groups are more disadvantaged or at risk than others (youth, women, poor, etc.), to look at some possible associations between indicators, to better assess the quality and representativeness of the information, or for several other reasons.

For instance, the analysis of MICS survey data can help policy-makers to better understand the reasons why previously observed childhood mortality declines have recently stagnated in some parts of a country, and not in others; what is the causal relationship between household structure and socio-economic status on children's health outcomes; how childhood malnutrition has changed over time; where are the areas of a country with the lowest levels of breastfeeding; how does access to water and sanitation facilities vary by wealth, residence, education, and other stratifiers.

References

- Boerma, T., Stansfield, S.K., (2007), *Health statistics now: are we making the right investments?* *Lancet* 2007; 369: 779-86.
- Charyeva, V.R., Samarkina, E.Y. and Sullivan, J.M., (2000), Chapter 9, *Infant and Child Mortality*, in *Gurbansoltan Eje Clinical Research Center for Maternal and Child Health*

(GECRCMCH), Ministry of Health and Medical Industry (Turkmenistan), and ORC Macro. (2001), *Turkmenistan Demographic and Health Survey 2000*. Calverton, Maryland, USA: GECRCMCH and ORC Macro.

United Nations Children's Fund (UNICEF), (in press). *The State of the World's Children 2008*.

World Health Organization, (WHO), (1993), *International classification of diseases and related health problems*, tenth revision. Geneva, Switzerland: WHO.

THE STRATEGIC INTENT OF DATA DISSEMINATION. THE CASE OF DEVINFO

by Nicolas Pron, DevInfo Global Administrator, UNICEF Headquarters

Overview

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 48 MDG indicators) and user-defined indicators. DevInfo is compliant with international statistical standards to support open access and widespread data exchange. DevInfo is distributed royalty-free to all Member States and UN agencies for deployment on both desktops and the web. The user interface of the system and the contents of the databases supported by the system include country-specific branding and packaging options which have been designed for broad ownership by national authorities.



Innovations achieved

DevInfo 5.0 has evolved from a decade of innovations in database systems which support informed decision-making and promote use of data to advocate for human development. A major innovation of DevInfo 5.0 is the introduction of data and metadata standards

to encourage open access and use of data across multiple organizations, platforms, and systems. DevInfo has adopted international standards in the areas of indicators (SDMX ISO/TS 17369:2005), data sources (DDI/Dublin Core) and digital maps (ISO 19115:2003). See www.devinfo.org for more information and www.devinfo.info for online databases.

Lessons learned for scaling-up

The DevInfo initiative is being implemented under the endorsement of the UNDP in collaboration with more than 20 UN agencies. More than 10,000 professionals have been trained in the use of DevInfo for improved statistical literacy and database administration (approximately 60% government and 40% UN professionals). More than 80 national statistics organizations and other agencies have officially launched DevInfo database adapted to user-specified requirements. There are a number of UN Agencies which have published adaptations of DevInfo – namely UNICEF, ILO, UNHCR, UN Habitat, UNHCR and UNFPA. For the second consecutive year, the UN Statistics Division has published the official UN data on MDG indicators in an adaptation of DevInfo, called *MDGInfo*. *MDGInfo 2006* has been prepared to accompany the *Millennium Development Goals Report 2006*, presenting the most up-to-date country-level statistics available in the UN (as of July 2006), for the global monitoring of progress achieved towards the MDGs since 1990.

The DevInfo database technology has been used to launch the Central and Eastern Europe/Commonwealth of Independent States (CEE/CIS) *Regional MDGInfo* (see www.regionalmdg.org). This database was officially launched at the Palais des Nations in Geneva on 27 April 2007



by Paolo Garonna, UNECE, Deputy Executive Director, Shahnaz Kianian-Firouzgar, UNICEF Deputy Regional Director for Central and Eastern Europe and the Commonwealth of Independent States, and Jafar Javan, Chief of Policy Support and Programme Development at UNDP's Bratislava Regional Centre.

Strategic intent for data dissemination using DevInfo

The vision, which 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"*.

DevInfo is enabling the UN system in realizing this vision as a general purpose database system designed for the compilation, dissemination, presentation and advocacy of human development indicators.

A common database system for evidence based planning

DevInfo is being used by UN Country Teams to support the Common Country Assessment (CCA) process using the latest available data. The system is also being used to setup and monitor key indicators of the UN Development Assistance Framework (UNDAF). 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 have been 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 a 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 sub-groups 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 rather than 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 meta-data. This assists planners to assess all of the available data related to the current situation, weigh alternatives and plan ahead as effectively as possible.

A common database system for increased access to information

The DevInfo database system serves as a common UN database technology platform for the collation, dissemination and presentation of human development indicators. The technology has been specifically designed to support governments in MDG monitoring. The MDG goals and targets are imbedded in the system linked to the 48 MDG indicators in a goal monitoring framework. In addition to the MDGs, the system can be adapted to include additional user-defined indicators linked to national monitoring frameworks. By serving as a common database, DevInfo can be used to add value to national statistics systems by complementing existing databases and bridging data dissemination gaps. The DevInfo common database can also be used as an advocacy platform by UN agencies to engage both government and civil society in policy choices for human development.

At the global and national level, DevInfo serves as a central repository of all data on human development to provide support for programming and decision making. DevInfo is used to make the latest statistical information available to a broader audience. The database can also be used to make lower level data more broadly available to researchers and organizations that help evaluate the effectiveness of methods and interventions.

A common database system for results based monitoring

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.

Within UN agencies, such as UNICEF, the system has been customized to monitor key performance indicators of the Medium Term Strategic Plan (MTSP Info).

A common database system to disseminate information to a broad audience

Human development data is disseminated by the UN system using various media: information sheets, booklets, CD-ROM, intranet and internet. In particular, the DevInfo database system provides content for publication on two global sites: www.devinfo.org and www.devinfo.info. Several DevInfo national websites have recently been launched to disseminate national and sub national data. Data can also be provided to international and government partners on demand, through online information dissemination services.

DevInfo data can be disseminated through data exchange utilities to other existing database systems, particularly database systems supported by the UN system. This paves the way for greater integration between existing Human Development databases from various government institutions and UN agencies, based on important emerging International Standards.

Use of DevInfo in strategic decision-making

DevInfo implementation reports from countries around the world provide a wealth of information on experiences and insight on the role of DevInfo in harmonizing monitoring systems, determining development priorities, supporting programming activities and guiding strategic decision-making in general.

An important success factor learned from these experiences is that the development and implementation of any DevInfo adaptation should be guided by the needs of intended users. These intended users should be identified from the beginning and include those persons with decision-making functions. The role of the UN should not be over-emphasized, but instead, national ownership should be encouraged, complemented with UN support in capacity-building and technical assistance.

Other guiding principles for the successful implementation of DevInfo include:

- DevInfo has an important role to play in a harmonized monitoring framework by encouraging agreement among different

stakeholders on indicator names and definitions, metadata and data sources.

- National ownership and strategic partnerships and linkages are important elements to further the use of DevInfo by decision makers at country level.
- Alignment with national priorities, relevance, reliability, and other qualities of the database technology are essential for maximizing the utility of DevInfo for decision-making purposes.
- Awareness-raising activities are effective in the dissemination of the database and in highlighting the value of the database among decision-makers.

DevInfo's role in supporting harmonized monitoring systems

Several countries have shared their experience on DevInfo's role in a harmonized framework for monitoring a range of plans and national priorities. In Bosnia and Herzegovina, for example, the various participating agencies discussed and jointly detailed the list of indicators to be included in the database. Similarly, in Nepal, the process of developing *NepalInfo* has played a key role in the coordination and harmonization of statistics in the country by requiring key ministries and partners to agree on the data to be contained in the system. Lesotho shares a similar experience with the harmonization of national statistics, and highlights the important role of DevInfo in the standardization of metadata. In Tanzania, the development of the DevInfo database has promoted the standardization of indicator definitions, time periods, units and metadata, ultimately increasing the confidence in the quality of the database. Once a database is recognized as harmonized, users give more recognition to its utility. In Mali, the utility of the database for the Coordinator of the *Poverty Reduction Strategy Paper* (PRSP), was increased by the fact that it was agreed upon by all ministries.

National ownership, partnerships and linkages of the system

Government ownership of the system is vital to the effective use of DevInfo by decision makers. Costa Rica selected a strategic implementing government partner who is responsible for, and who has assumed ownership over, the system and thus, is developing it further, promoting it, and most importantly, sharing the information contained therein. In Egypt, a Memorandum of Understanding

transferring ownership of the database was signed with the government agencies in charge of data collection, processing, analysis and dissemination. Furthermore, a major issue is the DevInfo adaptation's linkages 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, and is embedded in the monitoring system for the *National Strategy for Growth and Reduction of Poverty*. In addition, *TSED* is linked to a database used by the local government to collate and analyze statistics. The two databases complement each other as they serve different purposes. Cambodia provides a clear illustration of strategic linkages. The *Statistical Literacy Project* has partnered with the *CAMInfo* initiative to conduct joint nation-wide training 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 hoped 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 the Government, the EU, UNDP and UNICEF, but most important has been government ownership and their commitment to maintain and use it. Following this successful example, DevInfo is now being rolled out throughout the Eastern Caribbean. In Papua New Guinea, a Memorandum of Understanding was signed between the Department of National Planning and Monitoring and the UN system to monitor localized MDGs in *PNGINFO*, the national adaptation of DevInfo.

DevInfo is recognized in many countries as a powerful advocacy tool for mobilizing society and government. It is very important for the DevInfo initiative to form partnerships with the stakeholders in the areas of advocacy and communications.

Data quality and national priorities

The content and quality of the database can be a determining factor for whether DevInfo is used simply as a data repository or, to its full potential, for decision-making. An important basic consideration for ensuring that the database is relevant for decision-making is its alignment with national development priorities, plans and procedures. 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*, *CCA*, *National Human Development Report*, *Situational Analysis of Women and Children (SAWC)* and many others. 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 Indicator*; *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/program formulation and budgetary processes. Others have also used the database for reporting, proposal writing and presentations. Similarly, Malawi's *MASEDA* contains indicators for monitoring the country's development strategies, MDGs, and the UNDAF monitoring and evaluation matrix, supplemented by indica-

tors from other relevant areas such as governance. In Cambodia, *CAMInfo* 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.

The interest in linking MDG data with project data appears in several countries. Indonesia's *Aceh Nias Info* is a good example. Additional indicators suggested to make databases more relevant in certain contexts were governance, crisis and environmental indices and data on disaster vulnerability.

In order to be used consistently, the database needs to be perceived as reliable and remain relevant. Therefore, certain attributes of the database are considered crucial for its success. The database must be updated regularly in order for the data to be useful for current analysis and planning. This implies a structured, well-defined and documented data collection policy which ensures the continuous availability of such data. Several countries highlight the need for the data to be reliable and accurate and for there to be a national consensus on the content of the database in order to ensure it is trusted and relied upon and that its use is generalized. Malawi illustrates the need for the data to be valid and at an increased level of disaggregation, and for the system to reflect data integrity and proper metadata. In Moldova, *DevInfo Moldova* provides decision makers with reliable data that can help them adjust the design of social policies and promote dialogue with other stakeholders engaged in the poverty reduction strategy.

The inclusion of data at the sub-national level is often considered a key factor in enabling decision makers to use the database for situation analysis, determining priorities, monitoring progress and guiding decision-making processes in general. The experience in Serbia, for example, shows how the availability of *DevInfo Serbia* (which contains relevant data at the sub-national level), has contributed to decentralization and allowed the authorities in the various regions to monitor implementation of Regional MDG Plans of Action. In addition, the database informs the municipalities' budget allocation process. Similarly, in Thailand, *TPD Info* contains sub-national data and so plays a crucial role in monitoring the MDGs and

national development plans, at national and provincial level. It also serves as a dissemination tool for data to guide decision-making at the sub-national level. In the case of Moldova, inclusion of sub-national data in the database has yielded valuable inputs for policy analysis and reports which show disparities between districts. In Indonesia, such a database has facilitated planning at the provincial level. In order to ensure the relevance of *MASEDA*, Malawi is ensuring that it is part of a program aimed at strengthening monitoring and evaluation systems at national level, as well as at sector, district and community levels.

Advocacy and awareness-raising

Society in general, and decision-makers in particular, can be effectively made aware of the availability of the system through national and sub-national launches and dissemination activities. Thailand has focused on a dissemination strategy which includes development of public information materials. In particular, it aims to reach provincial governors and line ministry officials who can benefit from using *TPDInfo* in their decision-making. In Vietnam, *CiaB Info* is being used to inform pro-poor planning and decision-making processes in Cao Bang province. District and province authorities are in charge of the development and maintenance of the database. The database has been developed in Vietnamese in coordination with the national DevInfo adaptation, *VietInfo*.

Continuous advocacy and communication campaigns can help maintain awareness of the availability and usefulness of the DevInfo system. As experienced by Colombia and the Dominican Republic, for example, bulletins, newsletters and other material can be prepared periodically using DevInfo in order to illustrate its potential. Continuous capacity-building activities for DevInfo users and administrators will help keep in-country capacity up-to-date.

For more information on the implementation of DevInfo, visit www.devinf.org/worldwide for updates organized by country with references to the most recently published databases and the current database focal points.

USING DEVINFO AS A STRATEGIC TOOL FOR DECISION-MAKING. ACHIEVEMENTS AND LESSONS LEARNED IN MOLDOVA

by Mohamed Azzedine Salah,
Deputy Representative, UNICEF Moldova

Context and Challenges

Moldova is a small, landlocked, densely populated country located in Central Europe. Since its independence, the country has carried out extensive reforms of its public, economic and social frameworks guided by a vision of integration with the European Union (EU). Moldova has managed to put in place the basis for a transition to a market economy. In 2000, the government finalized a *National Strategy for Sustainable Development* (NSSD) which was the basis for an *Economic Growth and Poverty Reduction Strategy Paper* (EGPRSP) for 2004–2006. In 2003, Moldova also implemented a *Mid-Term Expenditures Framework* (MTEF) to better prioritize national resources.

Thanks to these measures, the country's overall human development has improved compared with the deteriorating situation throughout most of the 1990s. Despite these significant efforts, however, Moldova remains one of the poorest nations in Central Europe ranking 115th on the 2006 Human Development Index. The UN Common Country Assessment (CCA) indicated that inequities have increased recently with 24% of the population still living in persistent poverty. Due largely to the inability of the existing data system to highlight inequities and chronic disparities, decision-makers have not been able to develop and implement the strategic plans needed to address them. This was the legacy of a long period during which the local data collection system changed very little from Soviet times. Until recently, the system was characterized by a low demand for, and poor supply of, qualitative data. Investment in statistics focused on improving supply and little attention was paid to generating demand for data and its use in planning.²

Although the debate on the use of data for policy decision-making was identified as a main concern in 2000, the issue gained greater

prominence in 2003 with the launch of the second generation of structural reforms and the implementation of the EGPRSP. It then became clear that decision-makers would increasingly need data for measuring, and objectively reporting, the impacts of policies. The challenge for the Government was to reshape the existing monitoring and evaluation system in such a way as to make it more relevant, more useful for taking action and, more in line with basic international standards.

Building national capacity in monitoring poverty reduction

Based on their comparative advantage in building national capacity for equitable and sustainable economic growth, UNICEF and UNDP were asked by the Government of the Republic of Moldova to support the development of a monitoring and evaluation system. This needed to be able to generate and process data on human development not only for policy making but also for public use and for advocacy.

To this end, in 2004 a joint UNICEF-UNDP project entitled *Support for Strategic Policy Formulation, Monitoring and Evaluation in the Republic of Moldova* was developed with the main objective of assisting Moldova in strategic planning, and in monitoring and evaluating the EGPRSP (the plan for achieving the Millennium Development Goals (MDGs)). DevInfo was identified as a major instrument for this project.

DevInfo technology was officially offered to Moldova in 2005. The Ministry of Economy and Trade (MoET) in charge of overall EGPRSP coordination used DevInfo's standard package to build local capacity. The national team first developed local databases using DevInfo as the preferred software, and then they facilitated discussion among stakeholders to ensure the database met user needs. To avoid multiple databases, it was agreed to use DevInfo Moldova as the single tool for monitoring both the MDGs and EGPRSP. In October 2005, a draft version of DevInfo Moldova was built which included the Moldovan MDG framework. In 2006 DevInfo Moldova incorporated two monitoring frameworks and an up-to-date mapping tool which included first and second sub-national levels. In 2007 a new DevInfo web version was in place, making it possible for DevInfo Moldova to be available for open web tests starting in February 2007, (www.mec.gov.md) and an official web launch took place in May 2007.

Achievements in building decision-making capacity

Although the project has not ended, it had, by May 2007, achieved its main goal. That is, the equipping of the General Division of Macroeconomic Policies and Development Programmes at MoET with a sound institutional framework which allows the participation of a wide range of stakeholders in formulating, monitoring and evaluating policies to reduce poverty.

The national DevInfo database contains a set of indicators and tools which are regularly updated and used for various decision-making purposes. Users can make comparisons over the past 5 years. Although improvements are needed to ensure compliance with international requirements, the MoET database is able to provide 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 the two databases (EGPRSP and MDG), the team avoided duplication in collecting statistics and increased the reliability of reporting. They also avoided the conflicts which traditionally occur in maintaining statistical data systems.

With the objective of improving national capacity in decision-making, MoET developed two different types of comprehensive, analytical reports which are also DevInfo based. 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. This took advantage of measures to increase local consumption and tax revenues for the government. The EGPRSP helped to convince the Ministry to invest more in the social sectors (up 21 per cent in 2006). The *2005 Poverty and Policy Impact Report*¹ provides an overview of national development and includes detailed analyses on child poverty and on poverty in rural areas. These reports do not replace economic evaluations and public expenditures reviews. They do however provide useful information for decision-making since they contain analyses which indicate those elements which influenced programme results and, how the programme elements interacted among themselves. The reports are produced through an exclusive and nationally owned process where staff from MoET interacts with key decision-makers in line ministries. Because they provide objective analyses of local

1 The Poverty and Policy Impact Report 2005, April 2006, www.scred.moldova

realities, they are also used by external donors.² MoET organizes an annual event which is 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 are now fully institutionalized and used for strategic planning including by teams developing the *National Development Plan (NDP) 2008–2011*. The reports are posted on the government website and are used by a wide range of stakeholders.

As the DevInfo database continues to foster evidence-based assessments, it is gradually playing a role in facilitating a common understanding among the government, civil society organizations (CSOs) and development partners. Data analyses and maps are used as platforms for the national dialogue on poverty reduction. As information is easily accessible, DevInfo is used to produce a bulletin on EGPRSP implementation which is published regularly in Moldovan newspapers and posted on government websites. This bulletin leads 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 presents civil society's view of development in Moldova. The main purpose of the Report is to play a role in decision-making and, in particular, to influence the content of the new NDP for 2008–2011.

The DevInfo database is expected to gradually reduce the costs of data dissemination. Moldova has massive amounts of statistical data traditionally disseminated through paper publications or through Acrobat documents on the web. No searching, no browsing and no presentation features are available. The launch of the web version (www.devinformd.com) will not replace the existing documents but will help data users to access a modern, full-featured data dissemination system.

Lessons learned

The EGPRSP database is the first comprehensive and up-to-date socioeconomic database on the situation of human development in Moldova for use by government institutions, the donor community and civil society counterparts. It was created by the same unit that developed the EGPRSP report. Including the EGPRSP analysts in the first steps of establishing the database was essential given the influence they have in designing poverty reports.

2 Moldova Poverty Update, World Bank, June 12, 2006

Building consensus among stakeholders and keeping the first version of the database small and supportive are important factors in the success of the project. Given the complexity of the institutional framework of the national statistical system, the freedom to maneuver to revise the poverty monitoring and evaluation system was quite limited. The project team thus agreed to gradually develop the system, and instead of reinventing indicators, DevInfo used existing indicators which allowed the MDG database to be established simultaneously.

This experience represented a change in the relationship between the providers and the users of statistics in policy choices. The National Bureau of Statistics was always considered as the main data producer. Now there is a growing consensus that the gap between producer and user should gradually be reduced. As a first step, the policy making functions within ministries were reviewed, and they were assigned more responsibility for data management. The staff of these units will be involved in the review of indicators using DevInfo as the preferred database. They will be able to measure the effects of policies in their own respective departments. These improvements in statistical competence will prevent data being ignored and the incorrect reading of statistics.

Summary

After many years without quality socioeconomic analyses from local institutions, Moldova has tried a new approach oriented towards measuring poverty and using the data for making effective policies. MoET is currently meeting the demand for data through the UNICEF-UNDP joint programme *Support for Strategic Policy Formulation, Monitoring and Evaluation in the Republic of Moldova* which uses DevInfo technology. As the importance of data grows, along with the development of the information society, it is expected that the DevInfo system will be used on a larger scale for analysis; for decision-making; for improving public information; and, for international comparisons. This widening use will ensure the long-term sustainability of the DevInfo database and, of the monitoring system established, thanks to DevInfo technology.

USING DEVINFO TO SUPPORT GOVERNMENTS IN MONITORING NATIONAL DEVELOPMENT STRATEGIES.

THE CASE OF THE REPUBLIC OF SERBIA

by Dragana Djokovic-Papic, Statistical Office, Republic of Serbia and Oliver Petrovic, UNICEF Serbia

Background

As early as 2004, DevInfo was introduced in Serbia. Today, the DevInfo database is run at the national level by the Statistical Office of the Republic of Serbia (SORS) with full participation of line ministries and institutions. DevInfo is installed, as a monitoring and planning tool, in the Poverty Reduction Strategy Unit; the Council for Child Rights; and, in many ministries and institutions. Selected data are also presented for public use on the Republic Statistical Office website¹ so that international and local institutions, community-based organizations, school teachers, and media can monitor and study the situation in the country.

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). Both databases are strong tools for monitoring and planning at central and local level.

1 The official SORS web site is: <http://www.statserb.sr.gov.yu>

Use for policy development and child rights monitoring

The first version of DevInfo technology, called ChildInfo, was officially offered by UNICEF to the Republic of Serbia in 2004. The DevInfo task force was formed at the national level, with the aim of supporting database development and use. The first database, developed in 2004, focused on monitoring the *National Plan of Action for Children*. The system continued to develop smoothly and today DevInfo is a key tool for NPA monitoring which is accepted and used by the Council for Child Rights². *Local Action Plans for Children*, established in 16 municipalities in Serbia, are also monitored by DevInfo system.

Since 2006, DevInfo has been used for planning and development of national MDGs including the national MDG report. The Government has officially declared that DevInfo will be used for monitoring the MDGs. The MDG national team is using DevInfo extensively for preparation of reports and for planning purposes.

DevInfo is being used by the PRS unit for PRS monitoring and reporting. The second progress report, prepared jointly by the Poverty Reduction Strategy (PRS) unit and line ministries, uses data provided by the DevInfo database. The PRS Unit has also supported the use of DevInfo by the Joint Project and the Strategy for Sustainable Development teams, which are both bodies of the Government of Serbia.

As soon as DevInfo was adopted, many line ministries (e.g. Ministry of Housing; Ministry of Environment; and, Ministry of Social Welfare), received training on the database use. They all use DevInfo for planning and presentations on their respective sectors. Various types of institution and individuals also use DevInfo, including the Office of the President of the Republic of Serbia; the Institute for Social Sciences; and, the Institute for Education Quality and Evaluation.

In 2006, the gender statistics database was created within the DevInfo application and has since been used by SORS and the United Nations Development Programme (UNDP), to prepare the publication: *Men and Women in the Republic of Serbia*. The data will be monitored continuously to track changes in the position of women in the Republic of Serbia; to support analysis and reporting for the Serbian Parliament; and, for international publications. SORS

2 Governmental body responsible for the monitoring of NPA

will continue the development of the database and provide support for its use. The UNICEF publication: *The State of the Children in Serbia* also draws on DevInfo.

Sustainability

At the end of 2006, two years after the direct support by UNICEF, DevInfo was declared as a tool of particular interest for the Republic of Serbia and so became part of the regular programme of SORS. A Group on social indicators and analysis was formed as part of the Statistical Office of the Republic of Serbia's Department for Social Standards and Indicators. The unit consists of four people, paid for by the State, who have undertaken the task of further development and maintenance of the DevInfo database at the national level.

At local level, a DevInfo coordinator has been nominated in each of 16 municipalities where the Local Plan of Action for Children is being implemented. The coordinators are paid for by each municipality and are responsible for developing and maintaining the database.

Future directions

Future activities to support the application of DevInfo include:

- customizing the Millennium Development Goals and Poverty Reduction Strategy databases to include nationalized indicators;
- publishing a standardized national report twice a year to share the latest figures and trends for main indicators with key national decision makers and the media;
- continuing regular publication of *The State of Children in Serbia*;
- creation of a web-based DevInfo system which will enable broader and more interactive access by the public to various functions of DevInfo;
- planning and promotion of a special strategy to promote the use of DevInfo among experts, and, in particularly, civil society and the media.

USING DEVINFO AS A TOOL TO FACILITATE LOCAL COMMUNITIES' EMPOWERMENT.

THE CASE OF THE MUNICIPALITY OF PIROT

by *Vladan Vasic, Mayor of Pirot, Republic of Serbia;*
Oliver Petrovic, UNICEF Serbia and
Vladica Jankovic, UNICEF Serbia

The story which follows is about DevInfo's significant contribution to local development, inclusion and participation, in Serbia.

Following the *National Plan of Action for Children*, 16 (out of 167) municipalities in Serbia initiated *Local Plans of Action for Children* (LPA). These are strategic documents to define and guide optimal child development in local settings. One of the first municipalities to take up this challenge was Pirot. Pirot is a small municipality in the remote area of south-eastern Serbia, surrounded by mountains and quite isolated from the rest of the country.

In mid-2004, a Memorandum of Understanding was signed between UNICEF and the local authorities in Pirot. An important part of the agreement was the introduction of DevInfo as a tool to enable strategic planning and monitoring at the local level. Recognizing the importance and potential of DevInfo, the local LPA team, led by the Mayor, started developing the DevInfo database in parallel with the LPA project. The LPA project clearly defined two strategic objectives: (1) establishment of a local LPA implementation team and, (2) establishment of a DevInfo database for monitoring LPA implementation.

Key members of the local DevInfo task force, in addition to the DevInfo coordinator, are the representatives from the health, education and social welfare sectors. The main task of the team was:

- defining measurable indicators for agreed goals (the team also proposed redefinition of non-measurable goals);
- defining reliable data sources;
- agreeing the methodology and process of data collection; and,
- proposing additional research needs.

Additional research, conducted to obtain scarce data, particularly on marginalized families, revealed a number of significant findings which led stakeholders to modify their understanding of the situation within the Pirot municipality. For example, many Roma and disabled children were found to be outside the social system and missing from local statistics. Consequently, strategic plans were redesigned to ensure that these children enjoy their full rights to social services.

In general, DevInfo is used for monitoring the LPA implementation in Pirot. Team members from different sectors send data to the DevInfo coordinator. The coordinator maintains the database and, as agreed with the Mayor, produces a report twice a year which is sent to the Mayor and the Municipality Parliament.

DevInfo is also used to monitor implementation of other projects related to the LPA. Over recent years, with support from UNICEF and the Pirot Municipality, DevInfo has been used by local institutions and NGOs working on social inclusion of the most marginalized children.

Inclusion of young people's views is a significant part of the LPA process. The youth are considered as a partner in conducting monitoring, analysis, and decision-making on issues related to children. Once a year, during the "Children's Week", a roundtable takes place with local authorities. Here, children use DevInfo to present data, analyse what has been achieved and, identify future priorities.

The availability of data makes it possible to identify the need for social change so that appropriate action can then be taken.

DevInfo highlights important social trends which may otherwise be overlooked. The first DevInfo report was quite shocking in terms of the number of children left out of the system. Together with local situation analysis DevInfo helped to reveal that social services had overlooked many children, in particularly Roma children and those with disabilities. For example, no single Roma child living in a Roma settlement was enrolled in any facility for early childhood education. Today, there are 50 Roma children participating in pre-school education programmes for Roma, organized by the Roma NGO. There is still a long way to go. Currently, only one Roma child attends the regular pre-school institution. DevInfo also revealed to the community that most of the children who are attending the so called "special school", for children with some form of disabilities, (created during the socialist era), are Roma. Experts reviewing the cases realized that there was no explanation for this other than prejudice.

Many of the Roma children attending special school had no mentally disability. They simply lacked basic skills such as knowledge of the Serbian language and numeracy. Immediate action taken reduced the number of Roma children attending "special school" by 50% in the current school year. The DevInfo report also showed that only 20% of Roma children continued education following completion of elementary school. Now a local team of Roma representatives and educational experts are working to prepare the ground for continuing education of Roma children.

A further use of DevInfo is for review of the child budget allocation. As a result, from 2005, investment for children was increased 7 fold in just two years. In addition an increasing demand from the local population for a 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. Capacity building of child-worker professionals is also ongoing. Additionally, a new pre-school was built which tripled the access to early childhood education, raising it to 90 percent.

The process has actively involved local media from the very beginning. Today, the media are essential partners who regularly promote DevInfo and the LPA. Together with children, the media monitor the implementation and fulfilment of the goals. Local experts have, through the local media, promoted the LPA and highlighted DevInfo as the tool to monitor the LPA's , promoting its key features, importance and application. As a result, the LPA has become a common goal for the whole population of Pirot and they are watching closely the implementation of the LPA.

DevInfo is also an important tool when preparing donor proposals. Local experts and NGO representatives are today using DevInfo to develop funding proposals, and to apply for funding from the municipal budget and local companies. Several applications have already been approved and implemented.

Public access to the DevInfo database helps transparency and empowers the local community to stand up for the betterment of their society. For instance, the amendments to the *Law on Financial Support of Families with Children* adopted in 2006 imposed restrictions on access to child allowance. From the DevInfo database, it became evident that the number of beneficiaries of child allowance in Serbia had dropped. Even the poorest families were deprived of this right. In collaboration with UNICEF, the local community took

the initiative to raise the issue, citing the LPA provisions for children. This resulted in the adoption of a further amendment to this Law (effective as of 1 January 2007) which revoked the extremely strict conditions for child allowance. Subsequently, the DevInfo data has shown that the number of beneficiaries started to rise again.

As illustrated by these examples, DevInfo has been shown to be a strategic tool which helps to bring about positive social changes based on the evidence available. The Pirot Municipality has plans to further improve the function and use of DevInfo. In particular, to improve cooperation with institutions dealing with children. Since the mechanism for their cooperation and coordination is not institutionalized, data flow is sometimes slow. A process of decentralization and delegation of responsibilities to local municipalities and, local mechanism for cooperation, need to be established locally. Investment is needed to increase the limited technical and human capacities at local level. The previous successes encouraged the municipality to extend the use of DevInfo and to start monitoring other relevant issues, such as youth employment, housing problems and agricultural development. Public access will be extended through development of the LPA and DevInfo web site, and the installation of DevInfo for public use in the Municipality Hall.

Networking with other LPA municipalities in Serbia, and establishment of data flow with the DevInfo group at the National Statistical Office, will certainly strengthen the use of DevInfo at the local level and further mobilize the community to improve the situation for children in Pirot. DevInfo will continue to ensure access to key information for efficient and equitable decision-making.





Annexes

Annex I: Authors Vitae.....	210
Annex II: Abbreviations.....	217

Annex I: Authors Vitae

ADRIEN, Marie-Hélène is the current President of the International Development Evaluation Association (IDEAS). She is also the President of Universal Management Group a Canadian consulting firm. Dr. Adrien is a senior consultant in international development. She has 20 years of experience in the evaluation of programmes and organizations, and in building/developing evaluation capacity. She has applied her expertise in Canada and in 35 countries, particularly in Africa, the Caribbean and Central America. Dr. Adrien has contributed to the field of evaluation through her publications, including *Enhancing Organizational Performance: A Toolbox for Self-assessment* (IDRC, 1999); *Organizational Assessment: A Framework for Improving Performance* (IDRC, IDB, 2002); and, a *Guide to Conducting Reviews of Organizations Supplying M&E Training* (World Bank, 2002). Dr. Adrien is a lecturer at the World Bank and Carleton University IPDET programme.

BAMBERGER, Michael has worked on programme evaluations and gender impacts 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 in Brazil, Colombia, Costa Rica, El Salvador, Peru, and Venezuela. During his 22 years with the World Bank, he worked as an 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 assignments for the World Bank; U.S. Agency for International Development (USAID); Asian Development Bank; International Food Policy Research Institute (IFPRI); the United Nations Development Programme (UNDP); U.N. Department of Economic and Social Affairs; and, U.N. Evaluation Office. Mr. Bamberger has published widely on development evaluation. In 2006 he co-authored (with Jim Rugh and Linda Mabry) *Real World Evaluation: Working Under Budget, Time, Data and Political Constraints*.

CONNER, Ross is professor emeritus in the School of Social Ecology, Department of Planning, Policy and Design, at the University of California Irvine, USA. Prior to his recent retirement, he was the founder and director of the Center for Community Health

Research at the university. He is currently President of the Board of Trustees of the International Organization for Cooperation in Evaluation (IOCE), involving national and regional evaluation organizations throughout the world. He continues to work on and pursue his interests in community health and evaluation, in the US and abroad. His most recent work has focused on cancer control working with Chinese and Korean communities. In addition, he recently completed an assessment and strategic review of The California Endowment's 'Communities First' grant program, which has supported over 1,000 diverse California communities to select and act on health improvements. He is the author or coauthor of nine books and numerous articles in various areas, including health, education, criminal justice and leadership development. His writings also include papers on evaluation utilization in program improvement and policy formation, experimental and quasi-experimental designs in evaluation, international and cross-cultural evaluation issues, and evaluation training. He is an active international speaker and workshop leader on evaluation topics.

DJOKOVIC-PAPIC, Dragana graduated in Sociology from the Faculty of Philosophy at the University of Belgrade. She has worked for the Statistical Office of the Republic of Serbia since 1993 in various departments, including the Department of Population Statistics; the Department of Registers and Cadastre; the Department of Living Standards; and, the Department of Socio-Economic Indicators and Workforce Monitoring. Currently, she is the Head of the Social Standards and Indicators Department. She is the author and co-author of numerous articles. At present, she is working as Project Manager for the LSMS 2007 Serbia. Ms Djokovic-Papic is a member of the National MDG Working Group and Gender Equality Council, both of which are Government bodies.

HANCIOGLU, Attila is currently the Global Coordinator of the UNICEF-supported Multiple Indicator Cluster Surveys. Mr. Hancioglu holds a B.A. degree in Sociology, and M.A. and Ph.D. degrees in Technical Demography. Prior to joining UNICEF in 2004, he was Head of Department and Associate Professor at the Hacettepe University Institute of Population Studies in Ankara, Turkey. During his 23-year residency at this Institute, Mr. Hancioglu worked on and managed a large number of household surveys, including five Turkish Demographic and Health Surveys between 1983 and 2003, and the Multiple Indicator Cluster Surveys of Turkey and Azerbaijan, and worked on national and international projects supported by UNICEF, UNFPA, UNDP and the World Bank, and national organiza-

tions in Turkey. Mr. Hancioglu has written or co-written more than 35 articles, books, and survey reports, and has delivered papers at more than 25 scientific conferences. His main areas of interest include survey methodology, measurement and analysis of mortality, measurement of poverty, indirect demographic techniques, and child rights. Mr. Hancioglu is a Turkish Cypriot by birth.

JANKOVIC, Vladica graduated in Communications at the Faculty of Electrical Engineering. He has worked for the Federal Statistical Office, the World Food Programme and the UNHCR in the area of database development. Since 2004, he has been working at the UNICEF Office in Belgrade on the development and support of a monitoring system based on DevInfo. Mr Jankovic has not only participated in the development of the DevInfo Serbia National Database from the very start, but also in the Settlement level databases, MICS 2005 and the Roma database. Among his other duties, he is supporting the development of a DevInfo-based monitoring system at local/municipal level; liaison with the Republic level; and, enhancing the monitoring system and networking, at both levels.

JOBIN, Denis is the Vice President of the International Development Evaluation Association and is currently a Senior Policy Analyst for Programme and Policy Integration, in the Pollution Prevention Directorate at Environment Canada. 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), where he currently works. 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.

KARLSSON VESTMAN, Ove is professor of education at Mälardalen University, and Associate Dean of both the Faculty Board and the Department of Social Sciences. He was one of the founders and the first Vice-President of the Swedish Evaluation Society (SVUF). Since 2002, he has been the Director of the Mälardalen Evaluation Academy (MEA), a multidisciplinary evaluation and research centre that includes researchers from several disciplines and Swedish universities and that focuses on the evaluation of policy, social programs and organizations. Professor Karlsson Vestman's work concentrates on one of MEA's main aims, building evaluation capacity to promote

social change and poverty reduction for the increase of social justice. In this work, he typically is using participatory and mixed-method approaches. An example of his recent work in this area is a joint Sweden-Russia formative evaluation of social programs to address drugs and women-battering and to improve elderly care in the Leningrad County area. In addition to papers on his projects, he has published articles on dialogue methods and on critical theory, and he has written about the role of values and social justice in evaluation.

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) that 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, helping to design and implement the Government Performance and Results Act which is the hallmark of the US's strategic and programme planning model. She is the co-author of *Ten Steps to Results Based Monitoring and Evaluation*, now in its 4th printing and available in five languages. She is also the author of numerous papers on government management; results based management and poverty monitoring system development.

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 & 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.

MCWHINNEY, Deborah is the Deputy Representative in UNICEF's office in Bosnia and Herzegovina (BiH). She has an M.A. in Political Sociology and has over 10 years experience of development and human rights in Africa and Eastern Europe, the last 5 of which have

been with UNICEF in Serbia and Bosnia and Herzegovina. She is a UNICEF trainer on human rights based approaches to programming, and in results-based management. Ms McWhinney recently oversaw the implementation of the Joint Country-Led Evaluation in collaboration with governmental counterparts in BiH.

PARKER, David is Deputy Director of UNICEF's Innocenti Research Centre based in Florence, Italy. In this capacity he oversees and contributes to the Centre's studies on the impact of socio-economic policies and international standards on children. He previously served as coordinator of UNICEF's programme of cooperation in China, and in technical posts with UNICEF in New York and for the South Asia region based in Kathmandu. Mr Parker is a health economist and policy analyst by training, he holds graduate degrees from Princeton and Harvard Universities.

PETROVIC, Oliver studied Medicine at Belgrade University's School of Medicine. Currently, he is undertaking a PhD in Public Health. He is a UNICEF Project Manager and the Head of the Early Childhood Department at the Belgrade Office. His major achievements are work on the design and implementation of the new national mother & child health care policy; several innovative projects focused on excluded children and their integration into social systems and life; and, the coordinated establishment of a child rights monitoring system in the country. Mr Petrovic is the co-author of the MICS I-III publications, *Early Childhood Development (Breastfeeding, Nutrition, ARI)* and *Facts For Life*.

PRON, Nicolas Charles has been working for the United Nations for 15 years, of which 12 years were spent in the field in Africa and Asia, where he implemented UN Country Programmes. Mr. Pron is currently posted in UNICEF New York where he is in charge of Data Dissemination. He is the Global Administrator of the DevInfo 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 Master Degree in International Administration from the Sorbonne University and a Master Degree in International Development Law from the René Descartes University in Paris.

RIST, Ray C. was a former Senior Evaluation Officer in the Operations Evaluation Department of the World Bank. His previous position in the Bank was as Evaluation Advisor and Head of the Evaluation and Scholarship Unit of the World Bank Institute. Prior to coming to the World Bank in 1996, his career included 15 years in the United States Government with appointments in both the Executive and

Legislative Branches. He served as a university professor with positions at Johns Hopkins University, Cornell University, and George Washington University. Dr. Rist was the Senior Fulbright Fellow at the Max Planck Institute in Berlin, Germany, in 1976 and 1977. He has authored or edited 24 books, written more than 125 articles, and lectured in more than 60 countries. Dr. Rist serves on the editorial boards of nine professional journals and also serves as chair of an international working group which collaborates on research related to evaluation and governance.

SALAH, Mohamed Azzedine is the Deputy Representative of the UNICEF Moldova Office. Before joining UNICEF he trained and worked as an epidemiologist in Algeria. He joined UNICEF in 1993 as project officer in health and nutrition. In 2005 and 2006, he was in charge of the Monitoring and Evaluation tasks and the Social Policy Project. During this period, Mr. Salah managed with the Ministry of Economy and Trade and UNDP, the UNICEF component of the Joint UNICEF-UNDP project on Economy and Growth Poverty Reduction Strategy.

SEGONE, Marco is the Regional Chief, monitoring & evaluation advisor in the UNICEF Regional Office for Central and Eastern Europe and the Commonwealth of Independent States (CEE/CIS). Mr. Segone represents UNICEF on the Board of Trustees of the International Programme Evaluation Network (IPEN). He was Vice-President of IOCE (International Organisation for Cooperation in Evaluation), one of the founders of the Latin America and the Caribbean Network for Monitoring, Evaluation and Systematization (ReLAC), the Brazilian Evaluation Network (Rebrama) and the Niger M&E Network (ReNSE). He worked in Bangladesh, Pakistan, Thailand, Uganda and Albania in integrated development projects. In 1996 he joined UNICEF and worked for the Regional UNICEF Office for Latin America and the Caribbean, for UNICEF Niger and UNICEF Brazil. He has written and edited about 20 books and articles, including "Creating and developing Evaluation professional organizations", "New trends in development evaluation" and "Democratic evaluation". He has presented about 50 papers in international Conferences worldwide.

VADNAIS, Daniel joined UNICEF/New York 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 repre-

sentatives. 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 Program. 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 interfaith conference to focus solely on children's issues that took place at Princeton University. Mr. Vadnais, a native from Québec, holds a Masters Degree in Demography from the University of Montreal.

VASIC, Vladan has been the Mayor of the Municipality of Pirot since December 2003. He is one of the youngest people in the history of the Pirot municipality to have been elected to that position. Mr Vasic graduated from the Faculty of Electronics at Nis University in 1996. After graduation, he worked for six and a half years as a computer systems engineer. He completed his postgraduate M.A. studies at the Faculty of Economics in 2003. His work focused on the concept of integral computer information systems and their application in the company, "Prvi Maj".

VUKOVIC, Azemina has an MSc. in Natural Sciences from the University of Sarajevo and was trained in the design of electoral systems at the University of Essex. Before the war, she was the Director of the BiH Public Fund for Higher and Secondary Education. Since 1997, she has been working as a leader of different projects related to poverty reduction, socio-economic policy, country development, monitoring and evaluation and education. She worked for the OSCE in BiH for a number of years and for EU-based consultancies. Since 2004, she has been the Head of Office for the BiH Mid-term Development Strategy in the Economic Policy and Planning Unit (EPPU). As of March 2007, the EPPU has been transformed into the Directorate for Economic Planning, where Ms. Vukovic has been acting Director and Head of the Sector for the Preparation of BiH Development Documents, Analysis of Social Inclusion and M&E.

Annex II: Abbreviations

ACE	Asociacion CentroAmericana de Evaluacion
AEA	American Evaluation Association
AES	Australasian Evaluation Society
AfrEA	African Evaluation Association
BiH	Bosnia and Herzegovina
CCA	Common Country Assessment
CEE/CIS	Central and Eastern Europe/Commonwealth of Independent States
CES	Canadian Evaluation Society
CLE	Country-Led Evaluation
CLES	Country-Led Evaluations and Systems
CSOs	Civil Society Organizations
DAC-OECD	Development Assistance Committee of the Organization for Economic Cooperation and Development
DEP	The Directorate for Economic Planning, DEP (previously the Economic Policy and Planning Unit (EPPU), Government of Bosnia & Herzegovina
ECG	Evaluation Cooperation Group
EES	European Evaluation Society
EGPRSP	Economic Growth and Poverty Reduction Strategy Paper, Republic of Moldova
EPPU	The former Economic Policy and Planning Unit, now known as The Directorate for Economic Planning, DEP

GAO	General Accounting Office, USA
IDEAS	International Development Evaluation Association
IEG	Independent Evaluation Group, World Bank
IMES	Integrated Monitoring & Evaluation Strategy
ILO	International Labour Organization
IOCE	International Organization for Cooperation in Evaluation
IPEN	International Programme Evaluation Network
IPDET	International Programme for Development Evaluation Training
MDGs	Millennium Development Goals
MoET	Ministry of Economy and Trade, Republic of Moldova
MTDS	Medium-Term Development Strategy
MTEF	Mid-Term Expenditures Framework
M&E	Monitoring and evaluation
MTR	Mid-term Review
MTSP	Medium Term Strategic Plan, UNICEF
NDP	National Development Plan
NGOs	Non-Government Organizations
NPA	National Plan of Action for Children
NSSD	National Strategy for Sustainable Development
ODA	Official Development Assistance
OECD	Organization for Economic Cooperation and Development

OECD-DAC	Development Assistance Committee of the Organization for Economic Cooperation and Development
OED	Operations Evaluation Department, World Bank
PD	Paris Declaration
PIUs	Project Implementation Units
PRSP	Poverty Reduction Strategy Papers
QED	Quasi-experimental design
RCTs	Randomized control trials
ReLAC	The Latin America and Caribbean monitoring and evaluation Network
SAWC	Situational Analysis of Women and Children
SIDA	The Swedish International Development Agency
SORS	Statistical Office of the Republic of Serbia
UNDAF	UN Development Assistance Framework
UNEG	United Nations Evaluation Group
UNFPA	United Nations Population Fund
UNHCR	United Nations High Commission for Refugees
UNICEF	United Nations Childrens' Fund

UNICEF Regional Office for CEE/CIS
Palais des Nations
CH 1211 Geneva 10
Switzerland
www.unicef.org/ceecis

January 2008