

SOLUTIONS RELATED TO CHALLENGES OF **INDEPENDENCE,** **CREDIBILITY AND** **USE OF EVALUATION**

PROCEEDINGS FROM THE THIRD INTERNATIONAL
CONFERENCE ON NATIONAL EVALUATION CAPACITIES
30 SEPTEMBER – 2 OCTOBER 2013, SÃO PAULO, BRAZIL



*Empowered lives.
Resilient nations.*

SOLUTIONS RELATED TO CHALLENGES OF INDEPENDENCE, CREDIBILITY AND USE OF EVALUATION

PROCEEDINGS FROM THE THIRD INTERNATIONAL
CONFERENCE ON NATIONAL EVALUATION CAPACITIES
30 SEPTEMBER – 2 OCTOBER 2013, SÃO PAULO, BRAZIL

SOLUTIONS RELATED TO CHALLENGES OF
INDEPENDENCE, CREDIBILITY AND USE OF EVALUATION
PROCEEDINGS FROM THE THIRD INTERNATIONAL
CONFERENCE ON NATIONAL EVALUATION CAPACITIES,
30 SEPTEMBER – 2 OCTOBER 2013, SÃO PAULO, BRAZIL

© UNDP June 2014

eISBN: 78-92-1-056761-9

Manufactured in the United States of America

Printed on recycled paper

Permission to reproduce any part of this publication will be freely granted to educational and non-profit organizations.

Suggested citation: Independent Evaluation Office of UNDP,
*Proceedings from the Third International Conference on National
Evaluation Capacities*, UNDP, New York, June 2014.

Editor: Jeffrey Stern, Suazion

Design: Suazion

Printing: Consolidated Graphics

Publishing and dissemination: Sasha Jahic, UNDP

SOLUTIONS RELATED TO CHALLENGES OF

INDEPENDENCE, CREDIBILITY AND USE OF EVALUATION

PROCEEDINGS FROM THE THIRD INTERNATIONAL
CONFERENCE ON NATIONAL EVALUATION CAPACITIES,
30 SEPTEMBER – 2 OCTOBER 2013, SÃO PAULO, BRAZIL

KEY MESSAGES

- Evaluations improve the effectiveness of public policy, programmes and service delivery. Building a culture of evaluation requires developing national evaluation capacities, which in turn fosters accountability, good governance and transparency. These factors are the cornerstones of effective democracies.
- The independence, credibility and use of evaluations are intrinsically linked. Together, they support the establishment of national monitoring and evaluation systems.
- **Independence of evaluation:** Independence in evaluation is vital and relevant, as it helps support evaluation's ultimate function: to improve development results. When the consultative process of evaluation is considered inclusive whilst retaining independence of leadership, it enhances credibility to evaluation and furthers the potential for the use of its results. National evaluation systems must have financial and organizational independence, with visible independence between the organization conducting the evaluation and those being evaluated.
- **Credibility of evaluation:** Credibility of evaluations is enhanced when the evaluation process is consultative and inclusive, when the evaluators are competent, and when the organizations that sponsor and implement evaluations are committed to clear evaluation standards and processes. This requires evaluation systems to institutionalize a consultative process, promote the professionalization of evaluation, invest in building evaluation capacity, and establish clear and visible policies, standards and guidelines for evaluation.
- **Use of evaluation:** National evaluation capacities are central to promoting the use of evaluative evidence in policymaking. Dialogues among governments, parliaments and civil society, along with advocacy for evaluation, are key strategies for promoting the use of evaluation. The use of evaluation in policymaking is further dependent on the legitimacy and quality of evaluation.

KEY OUTCOMES

- Deeper understanding of challenges and innovations in an effort to build national evaluation capacities in different country contexts.
- Relationships between government and civil society representatives towards a mutual commitment to creating demand for evaluation and building national evaluation capacities.
- Articulation of 18 commitments by conference participants centred around four main strategies to build national evaluation capacities: (1) promote evaluation use through in-country and global advocacy; (2) define and strengthen evaluation processes and methods; (3) engage existing and new stakeholders in exchanges and collaboration; and (4) explore options for different institutional structures for managing evaluation.
- Recognition of areas for South-South cooperation, South-based innovations in building national evaluation capacities and South-based experts who can lead the way.
- An agreed landmark (International Year of Evaluation – EvalYear 2015) to monitor progress and promote mutual accountability to commitments for strengthening national evaluation capacities.



TABLE OF CONTENTS

Acknowledgements	10
Preface	12
Foreword.	15
Executive Summary	16
Conceptual Framework	26

COMMISSIONED PAPERS 31

1. Encouraging the Effective Use of Evaluations to Improve Programme Management and Design.	32
PAULO MARTINO JANNUZZI	
2. Innovative Solutions to Challenges Linked to Independence, Credibility and Use of Evaluations.	42
MARÍA BUSTELO	
3. Reflections on Independence in Evaluation	53
HANS-MARTIN BOEHMER, XIMENA FERNANDEZ ORDONEZ AND NEHA SHARMA	
4. Innovations in Monitoring and Evaluating Results.	68
THOMAS WINDERL	

COUNTRY PAPERS 91

1. Afghanistan – Use of Evaluation: Local Governance M&E System in Afghanistan	93
MOHAMMAD HAMED SARWARY	
2. Afghanistan – Political Participation for Development Evaluation in South Asia: An Afghanistan Perspective.	99
RANGINA KARGAR	

- 3. Cambodia – Evaluation of Development Activities at the Macro Level: Challenges and Experiences in Cambodia 105**
THENG PAGNATHUN
- 4. Indonesia – Monitoring and Evaluation System in Indonesia 114**
ARIF HARYANA
- 5. South Asia – Why National Evaluation Policies Matter in South Asia. 119**
HON. KABIR HASHIM AND ASELA KALUGAMPITIYA
- 6. Sri Lanka – Country-led National Evaluation System: Independence, Credibility and Use of Evaluations; Challenges and Solutions 124**
VELAYUTHAN SIVAGNANASOTHY AND VAIDEHI ANUSHYANTHAN
- 7. Malaysia – An Evaluation of Public Perception towards Government Projects: A case study of Constitution XYZ 132**
MOHD KHIDIR BIN MAJID AND SHAHRAZAT BINTI HAJI AHMAD
- 8. Nepal – Institutionalization and Use of Evaluations in the Public Sector in Nepal 137**
TEERTHA DHAKAL
- 9. Argentina – Policy Evaluations of National Public Administration: Overview 143**
NORBERTO PEROTTI
- 10. Brazil – A Model to Evaluate the Maturity of the Brazilian Public Administration’s Evaluation Systems 147**
SELMA MARIA HAYAKAWA CUNHA SERPA AND
GLÓRIA MARIA MEROLA DA COSTA BASTOS
- 11. Brazil – Evaluating Social Development in Brazil: Analysis of the Cycle of Production and Dissemination of Evaluation Studies by the Ministry of Social Development and Fight against Hunger 155**
MARCO ANTONIO CARVALHO NATALINO,
ALEXANDRO RODRIGUES PINTO AND
MARTA BATTAGLIA CUSTODIO

12. Dominican Republic – Analysis of Lessons Learned by the Dominican Republic in Local and Institutional/ Government-Level Monitoring, Evaluation and Participatory Follow-Up	160
HERNAN RODRIGUEZ MINIER	
13. Dominican Republic – Experience of the Dominican Republic in the Evaluation of the Solidaridad (Solidarity) Programme	163
IVETT SUBERO ACOSTA, DORIS MELO AND SÓCRATES BARINAS	
14. Mexico – Overcoming Mistrust in Poverty and Evaluation Figures: The case of CONEVAL in Mexico	169
GONZALO HERNÁNDEZ LICONA AND AGUSTÍN ESCOBAR LATAPÍ	
15. Mexico – Evaluation Use and its Institutionalization in the Federal Evaluation System in Mexico	175
GABRIELA PÉREZ-YARAHUÁN	
16. Mexico – Incentives in the Use of Evaluations as a Tool for Improving Public Policy	180
THANIA DE LA GARZA AND RASEC NIEMBRO	
17. Peru – Evidence-based Public Policies: The Experience of the Ministry of Development and Social Inclusion	185
ANIBAL VELÁSQUEZ VALDIVIA, URSULA MARTÍNEZ ANGULO AND FABIOLA CÁCERES PAURINOTTO	
18. Trinidad and Tobago – Building a National Performance Framework for Evaluation.	192
DINESH D. RAGOO	
19. Benin – The Process of Institutionalizing Evaluation in Benin: Progress on Questions of Usage and Independence	196
ARISTIDE N. DJIDJOHO	
20. Ghana – Building M&E Capacities to Enhance the National M&E System in Ghana: The WayForward	203
BRUNO B. DERY	
21. Kenya – Credibility of Evaluation: The Kenyan Case	211
SAMSON MASESE MACHUKA AND JENNIFER MUTUA	

- 22. Malawi – Monitoring and Evaluation of the National Budget in Malawi 216**
WINSTON NYASULU
- 23. Malawi – Promoting Use of Evaluation Results by Senior Government Officials in Malawi 222**
TED SITIMA-WINA
- 24. South Africa – Reflections on the South African Experience with Evaluation and the Use of Evaluative Evidence to Orient Public Policy Formulation 227**
IAN GOLDMAN, STANLEY NTAKUMBA AND CHRISTEL JACOB
- 25. Egypt – Institutionalizing and Streamlining Development Monitoring and Evaluation in Pre- and Post-Revolutionary Egypt 232**
DOHA ABDELHAMID
- 26. Morocco – Gender Evaluation of Public Policies from a Human Rights Perspective. 238**
CHAFIKI MOHAMED
- 27. Morocco – Evaluation of the National Initiative for Human Development: Independence and Use 243**
MOHAMED MOUIME
- 28. Canada – Alphabet Soup: Linking the Letters to Promote Evaluation Use 248**
MARTHA MCGUIRE AND JAMES MELANSON
- 29. Albania – Challenges of a New Evaluation National Society 252**
FATION LULI AND MARIE GERVAIS
- 30. United States of America – The Diversity of the Evaluation Enterprise Influences its Use in the United States 256**
STEPHANIE SHIPMAN AND JODY FITZPATRICK
- 31. World Bank – Science of Delivery: Implications for Monitoring and Evaluation 262**
ARIANNE WESSAL, CLAY G. WESCOTT
AND MATTHEW COURTNEY TREUTH, JR

ANNEXES	269
Annex 1: Agenda	271
Annex 2: Participants.	274
Annex 3: Conference Assessment.	281
ATTACHMENT: THE EVENT IN PHOTOS	283

ACKNOWLEDGEMENTS

Several organizations and people are responsible for making the Third International Conference on National Evaluation Capacities a success. As co-hosts, the Independent Evaluation Office of the United Nations Development Programme (UNDP) and the Brazilian Ministry of Social Development and Fight against Hunger, are grateful for the opportunities this conference yielded for the respective institutions and look forward to further collaborations.

The support of UNDP management, particularly Ms Rebeca Grynspan (UNDP Associate Administrator) and Mr Jorge Chediek (UN Resident Coordinator, UNDP Country Office in Brazil) are greatly appreciated.

Our gratitude also goes to Mr Marcelo Cabral, Ms Cecília Ishikawa Lariú and Mr Alexandro Rodrigues Pinto from the Brazilian Ministry of Social Development and Fight against Hunger, as well as to Ms Marianne Pinotti (Municipal Secretary for People with Disabilities and Reduced Mobility, São Paulo) and Mr Rogério Haman (São Paulo State Government Secretary for Social Development).

The conference was made possible thanks to the generous financial support provided by the Governments of Finland, Norway and the United Kingdom. The organizers also appreciate these governments' substantive contributions as participants at the conference.

We gratefully acknowledge the role of the Advisory Group in guiding the organizers in conceptualizing, organizing and implementing the conference. The Advisory Group consisted of:

- Ms Riitta Oksanen (Senior Advisor, Development Evaluation, Office of the Under-Secretary of State, Ministry for Foreign Affairs of Finland);
- Ms Paula Montagner (Deputy Secretary of Evaluation and Information Management, Ministry of Social Development and Fight against Hunger in Brazil);
- Ms Deborah Rugg (Director, Inspection and Evaluation Division, UN Office of Internal Oversight Services, United Nations Evaluation Group [UNEG] Chair);
- Ms Taiana Fortunato Araujo (Brazilian Monitoring and Evaluation Network Management Committee, Economist of the National Institute of Metrology, Quality and Technology [INMETRO], Ministry of Development, Industry and Foreign Trade in Brazil);
- Mr Marco Segone (United Nations Children’s Fund [UNICEF] and EvalPartners Co-Chair);
- Ms Natalia Kosheleva (President, International Organization for Cooperation in Evaluation, and EvalPartners Co-Chair);
- Ms Thania de la Garza Navarrete (Adjunct General Director of Evaluation, Consejo Nacional de Evaluación de la Política Social [CONEVAL]);
- Mr Mohamed Mouime (Director of Information Systems Division, National Observatory for Human Development, Morocco);
- Ms Christel Jacob (Director, Evaluation and Research Unit, Department of Performance Monitoring and Evaluation in the Presidency, Republic of South Africa); and
- Mr Kobus van der Merwe (Chief-Director, Governance Monitoring, Office of the Public Service Commission, Republic of South Africa).

The support and participation in the conference by the Brazilian Monitoring and Evaluation Network (especially Ms Marcia Paterno Joppert), EvalPartners, Regional Centers for Learning on Evaluation and Results (CLEAR) and evaluation offices in the UN and multilateral systems are also recognized.

Staff of the UNDP Bureau for Development Policy, including Mr Romolo Tassone, Ms Xiaojun Grace Wang and Ms Jennifer Colville, and Ms Inka Mattila of the UNDP Regional Bureau for Latin America and the Caribbean also participated actively.

The following staff of the Independent Evaluation Office of UNDP contributed to the conference and/or to developing materials for this report: Mr Juha Uitto, Ms Ana Rosa Soares, Ms Concepcion Cole, Mr Oscar Garcia, Ms Joan Gordon, Ms Ximena Rios, Ms Michelle Sy, Ms Vijayalakshmi Vadivelu, Ms Sasha Jahic, Mr Anish Pradhan and Mr Roberto La Rovere.

The following people also contributed while temporarily associated with the Independent Evaluation Office of UNDP for this conference and/or this report: Ms Neha Karkara, Ms Tessie Tzavaras Catsambas and Mr Nicholas J. Cuozzo.

PREFACE



Our world is more complex, interconnected and unpredictable than it has ever been, and so are the challenges that we face. The pressure on national governments, non-state actors, development agencies and donors to achieve and demonstrate sustainable results is huge, and will only increase as the post-2015 development agenda emerges. Evaluation can provide critical information to ensure that development interventions take longer-term and interconnected challenges into account despite the short-term nature of the planned actions and projects.

National evaluation capacities are increasingly critical to countries' overall ability to capture and demonstrate results to a variety of constituencies and to promoting learning and enabling decision makers to make informed decisions and plan strategically. We at the United Nations Development Programme (UNDP) are convinced that the success of development interventions depends in part on the ability to carry out credible evaluations and use them to make evidence-based decisions.

National evaluation capacities conferences, such as this one, provide forums for discussion of evaluation issues confronting countries and enable participants to draw on other countries' recent and innovative experiences. The previous two national evaluation capacities conferences in Morocco and South Africa emphasized the need to build better institutional capacities to manage development programmes. The focus was on improving organizational systems and processes, and on developing incentives for better performance and results. National systems that facilitate independent and credible evaluations play an important role in achieving these priorities by generating evidence and objective information on how improvements can be made. This is why institutionalizing evaluation systems and practices is so important.

National evaluation capacities are increasingly critical to countries' overall ability to capture and demonstrate results to a variety of constituencies and to promoting learning and enabling decision makers to make informed decisions and plan strategically.

Let me reaffirm that UNDP is fully committed to supporting efforts to strengthen national evaluation capacities. UNDP has invested considerable effort in the past to strengthen results-based management, programme performance, learning from evaluation and results reporting. We are seeing a stronger culture of results taking root in the organization. In line with the UNDP Evaluation Policy, UNDP programme units promote and coordinate South-South and triangular cooperation in support of strengthening national capacities, particularly for country-level evaluation. South-South cooperation is a powerful tool for addressing common development challenges. Similarly, South-South solutions can play a strong role in promoting and strengthening national evaluation capacities and addressing the urgent concerns confronting development evaluation today.

As we move towards designating 2015 as the International Year for Evaluation, I call upon all partners to further redouble their efforts to strengthen national evaluation capacities and to harness the wealth of knowledge and expertise in development thinking in the global South. In this increasingly complex development environment, it is incumbent upon us to keep evaluation relevant by recognizing and responding to the challenges of its use, credibility and independence.



Ms Rebeca Grynspan

UN Under-Secretary-General and UNDP Associate Administrator

Key messages from the Opening Keynote Address



More than 160 participants attended the conference from 63 countries, making it the largest evaluation event in the United Nations system.

FOREWORD



The Independent Evaluation Office of the United Nations Development Programme (UNDP) and the Brazilian Ministry of Social Development and Fight against Hunger came together to co-host the Third International Conference on National Evaluation Capacities.

The purpose of the conference was to share experiences, challenges and solutions from programme countries with national monitoring and evaluation (M&E) systems, including countries that may be considering creating one. The conference provided a forum to identify lessons and challenges in implementing national M&E systems and the use of evaluation. Overall, the conference helped to identify supply and demand for technical assistance in strengthening institutional capacity for national M&E systems under the umbrella of South-South and triangular cooperation. The conference was organized around three interrelated topics: independence, credibility and use of evaluation.

More than 160 participants attended the conference from 63 countries, making it the largest evaluation event in the United Nations system. The event was also webcast to 140 countries across the globe. The conference was specifically designed to provide opportunities for an open exchange and direct interactions among national institutions commissioning evaluations, institutions and professionals conducting evaluations and decision makers using evaluations.

This report captures key messages and outcomes from the conference, contributing to knowledge sharing and South-South cooperation among countries to strengthen evaluation-related efforts. To increase relevance for policymakers, this report presents a set of four commissioned papers that provide a conceptual framework for the conference's theme. The report also includes 31 country papers (the majority of which were authored by conference participants) that provide national perspectives on the issues of independence, credibility and use of evaluation.

The Independent Evaluation Office of UNDP and the Brazilian Ministry of Social Development and Fight against Hunger are committed to continuing to support the dialogues initiated at the conference. We hope that this report serves as an advocacy tool to promote the continuity of networks, cooperation and agreements reached at the conference.

Mr Indran Naidoo

Director, Independent
Evaluation Office of UNDP

Mr Paulo Jannuzzi

Secretary for Evaluation and Information Management,
Ministry of Social Development and Fight against Hunger

EXECUTIVE SUMMARY

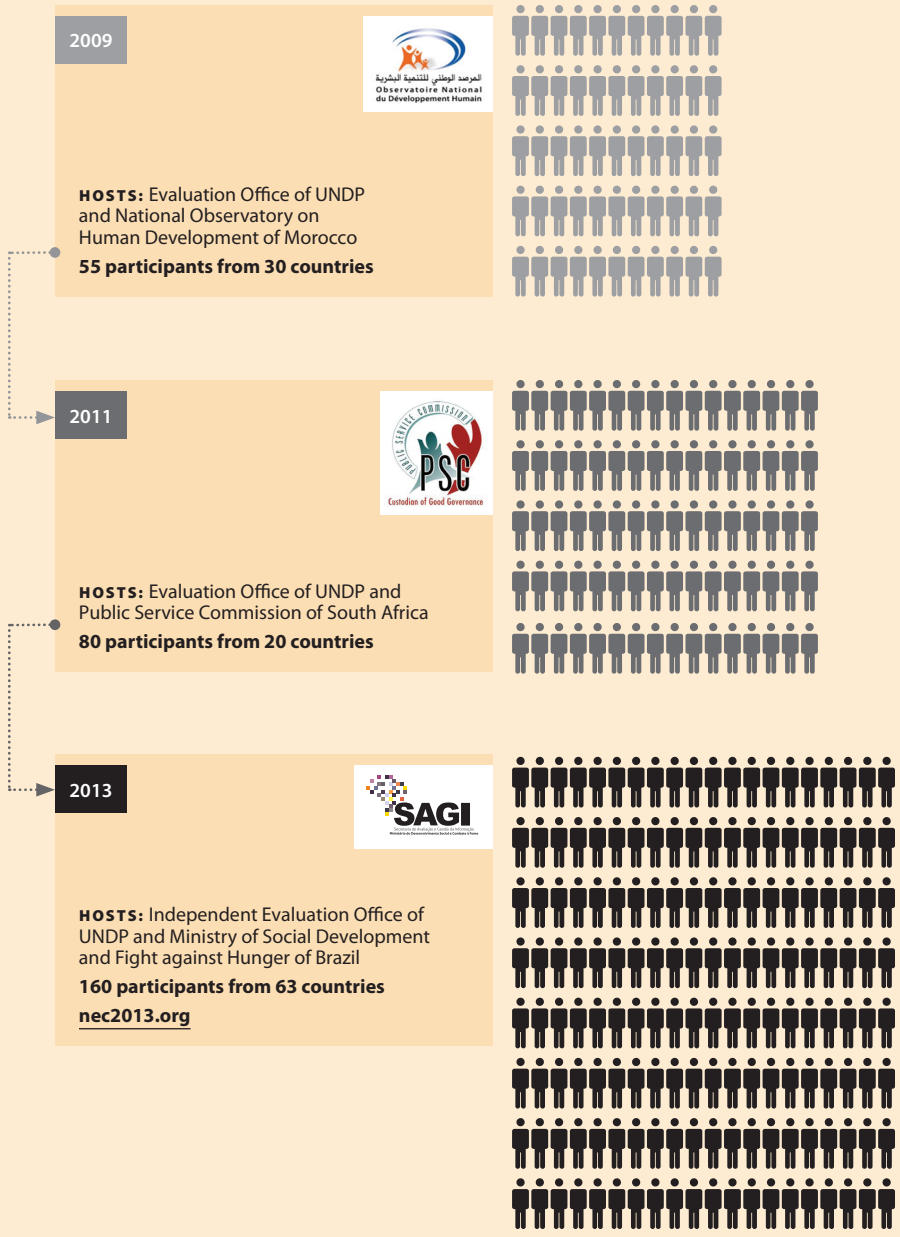
The Third International Conference on National Evaluation Capacities was held from 29 September to 2 October 2013 in São Paulo (Brazil). The conference was co-hosted by the Independent Evaluation Office of the United Nations Development Programme (UNDP), the Brazilian Ministry of Social Development and Fight against Hunger (SAGI). The overarching theme of the conference was *Solutions to Challenges Related to Independence, Credibility and Use of Evaluation*.

This conference built upon the deliberations of the First International Conference on National Evaluation Capacities¹ (Morocco, 2009) and the Second International Conference on National Evaluation Capacities² (South Africa, 2011). National evaluation capacities conferences provide forums for discussing evaluation issues that confront countries, enable participants to draw on innovative experiences of other countries and work towards South-South solutions.

More than 160 participants from 63 countries attended the 2013 conference, including representatives of national institutions responsible for commissioning, conducting and using evaluations of public policies, programmes and projects. In addition, policy-makers, parliamentarians, evaluation experts, practitioners, academia, civil society, Voluntary Organizations of Professional Evaluation (VOPEs) and officials from UNDP, other UN system and development agencies took active part in the conference. With live webcasts of key sessions, a large number of global viewers within and outside the UN were also able to join the proceedings.³ This was the first conference on national evaluation capacities of such scale, actively engaging a range of stakeholders from across the globe.

-
- 1 The First International Conference on National Evaluation Capacities was held in Casablanca (Morocco) from 15 to 17 December 2009. The conference was co-hosted by the Evaluation Office of UNDP and the National Observatory on Human Development of Morocco. Approximately 55 participants from 30 countries participated in his conference. A report is available at nec2013.org/downloads/NEC-2009-Proceedings.pdf.
 - 2 The Second International Conference on National Evaluation Capacities was held in Johannesburg (South Africa) from 12 to 14 September 2011. The Evaluation Office of UNDP and the Public Service Commission of South Africa co-hosted. Approximately 80 participants from 20 countries participated in this conference. A report is available at nec2013.org/downloads/NEC-2011-Proceedings.pdf.
 - 3 More than 400 people visited the live streaming website. At any point there were 64 simultaneous live connections.

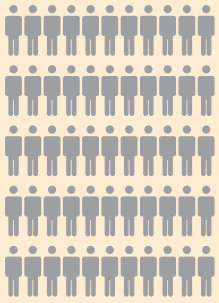
FIGURE 1: THE EVOLUTION OF THE INTERNATIONAL CONFERENCE ON NATIONAL EVALUATION CAPACITIES



2009



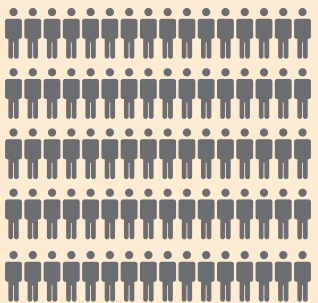
HOSTS: Evaluation Office of UNDP and National Observatory on Human Development of Morocco
55 participants from 30 countries



2011



HOSTS: Evaluation Office of UNDP and Public Service Commission of South Africa
80 participants from 20 countries



2013



HOSTS: Independent Evaluation Office of UNDP and Ministry of Social Development and Fight against Hunger of Brazil
160 participants from 63 countries
nec2013.org



In order to capture the wealth of knowledge and experience of the participants, each day of the conference was dedicated to a specific sub-theme (use, credibility and independence of evaluations). On each sub-theme, parallel solution forums were organized, where participants engaged in discussions within smaller groups. The solutions were presented in a plenary session so as to encourage commitments on South-South cooperation to promote evaluations. In the run-up to the conference, online discussions⁴ on the theme of the conference also provided participants with opportunities to share experiences in promoting evaluation practices and culture.

In her opening address, Ms Rebeca Grynspan hailed the importance of evaluation to democracy, as it is critical for good governance (voted as the third highest priority for the UN). Grynspan noted the role of evaluations in bridging the gap between experience and knowledge. She stressed that evaluations must be more flexible, collaborative and responsive, without compromising the rigour and independence needed for their credibility. She challenged evaluators to measure what is important and not just that which is easy to measure, to incorporate cultural considerations and to focus on South-South exchanges and learning. Finally, she urged governments to institutionalize evaluation to help in scaling up programmes.⁵

The conference **declared 2015 as the International Year of Evaluation (EvalYear)**. EvalYear provides a key opportunity to position evaluation in the policy arena by raising awareness of the importance of embedding evaluation systems in the development and implementation of international, national and subnational development goals and strategies. In the context of EvalYear, the discussions highlighted several innovative initiatives that promote critical partnerships among governments, parliamentarians and VOPEs to strengthen the demand for and use of evaluation. A panel of representatives from Sri Lanka's government, its parliament, and the Sri Lanka Evaluation Association embodied an example of whole-system collaboration for building national evaluation capacity. More information on EvalYear is available at mymande.org/evalyear.

INNOVATIVE SOLUTIONS TO CHALLENGES LINKED TO EVALUATION USE

In his keynote address, Mr Paulo Jannuzzi⁶ noted that, in order to promote the use of evaluation, evaluation objectives and questions must be aligned with evaluation methods. This will ensure that evaluation is relevant to the public policy issues at hand. Jannuzzi further attested that enhancing the credibility of an evaluation enhances its use. Recognizing the political nature of evaluation, an inclusive and participatory process of evaluation builds its credibility. When evaluators are willing to listen to every party, and the evaluation incorporates and represents different perspectives, the evaluation gains legitimacy. Jannuzzi

4 A summary of the Community of Practice Discussions on conference's themes is available at nec2013.org/downloads/NEC-2013-COP-summary.pdf.

5 Ms Grynspan is the UN Under-Secretary General and UNDP Associate Administrator. Her full speech can be accessed at nec2013.org/downloads/AA_Statement_NEC2013.pdf.

6 Secretary for Evaluation and Information Management, SAGI.

also emphasized the importance of effective and timely communication targeting different audiences to ensure the use of evaluation results. *The paper on the use of evaluations is available in the Commissioned Papers section of this report.*

Following three parallel solution forums on the use of evaluation,⁷ the plenary session noted that the presence of a national monitoring and evaluation policy has proven to be helpful in promoting the use of evaluations in some countries. However, one of the most difficult challenges for countries has been where to place—and how to structure—the evaluation mandate and function within government. From the participants' discussion, here is a perspective on the challenges impeding evaluation use from the field:

“There are various factors that limit the demand and use of evaluations by the Government of Afghanistan. Particularly, lack of understanding of the senior management and leadership on the importance of M&E information... There are various factors that affect the [use of evaluations]... such as political willingness, not completely relying on [evaluation] reports, lack of technical understanding and awareness of the issues raised, and interventions and favouritism.”

AFGHANISTAN

Another critical bottleneck for promoting the use of evaluation has been the lack of appropriate channels that connect evaluations to programmes.

“M&E data must be fed back into the ministries to guide in the implementation of on-going activities and the planning of future ones. However, there is little feedback that goes to the implementing agencies. To improve on the use of evaluations, there is need for proper feedback to the implementing agencies.”

MALAWI

An additional challenge has been broadening the use of evaluation beyond a limited number of stakeholders.

“Today, evaluation appears to be highly ranked as an instrument for government change among public officials in the federal government, particularly so for the change of rules and the acquisition of information about programme performance. This use is perceived by government actors within the Executive. The perception on the use by actors outside the government, such as [members of] Congress, beneficiaries or citizens, remains to be achieved.”

MEXICO

Participants also learned from examples of countries such as India, Malaysia and the United States. India is currently implementing the Performance Monitoring and Evaluation System, which covers 80 departments and 800 responsibility centres. The Government of Malaysia requires line ministries to conduct evaluations and provides evaluation guidelines. In the United States, the Accountability Office⁸ is an independent, nonpartisan government agency

7 The topics for the three parallel solution forums were: (1) Evaluation capacities to ensure use of evaluation; (2) Innovative approaches or solutions to ensure the use of evaluation findings and implementation of recommendations; and (3) National evaluation systems and incentives to promote the use of evaluation.

8 See [gao.gov](https://www.gao.gov).

that works for Congress. The discussions noted that developing structures that are mandated by law to use evaluation results can create accountability.

In Sri Lanka, the strategy of promoting collaboration among government, parliament and civil society actors to increase the visibility of evaluation has had positive results:

“The Government of Sri Lanka has given high priority to ensure value-for-money in public management. As a result, concepts such as Managing for Development Results Development Evaluation and Performance Audit are getting very high focus in public management.”

SRI LANKA

The discussions further highlighted that, in addition to policy and structure, the profile of evaluation must be raised among the policy-makers and the public. This means that evaluation reports must be published and disseminated openly and, if possible, in local languages. Reports should be user-friendly (consider videos, radio or television reporting as evaluation outputs), and they need to present helpful insights that illuminate different public policy options and issues. Countries' innovations for dissemination included an intuitively coloured dashboard from BAPPENAS (the planning ministry of Indonesia) that makes it extremely easy to review and understand evaluation results. This is further highlighted in the following experience:

“The information provided by the Mid-term Evaluation (MTE) was very beneficial for policy-makers, because the MTE was conducted at the right time. Performance information from MTE was available just in time, when the process of planning and budgeting for the fiscal year ... started. The MTE results were presented by the Minister and discussed in the cabinet meeting chaired by the President.”

INDONESIA

Building a culture of evaluation further requires creating an appropriate incentive structure that promotes the use of evaluation in policymaking. In Malaysia, for example, funding for follow-on projects is tied to the completion of evaluations in existing projects.

INNOVATIVE SOLUTIONS TO CHALLENGES LINKED TO EVALUATION CREDIBILITY

In her keynote address, Ms María Bustelo⁹ noted that the credibility of evaluation is inextricably linked to inclusion or exclusion in evaluation. In other words, who participates in the evaluation, who pays for it, what institution hosts or leads the evaluation, who leads the evaluation, who is interviewed and who provides feedback for the evaluation jointly determine its credibility. Bustelo suggested that credibility not only depends on the quality and independence of individual evaluators, but also on the perceptions of independence and competence of the institutions that commission evaluations.

In some countries, a close link with the government gives credibility to the evaluation; the same may not hold true in others. Governments should ensure that the framework of their M&E system fits within their broader system of government. Credibility is also linked

9 Associate Professor, Department of Political Science and Public Administration, Complutense University of Madrid.

to evaluation capacity. Strong evaluation capacities often lead to good quality evaluations that adhere to common evaluation standards and follow legitimate, inclusive processes. Investing in training in evaluation is critical for the credibility of evaluations and of the field of evaluation overall. Finally, regular and transparent communication is essential to enhancing the credibility of the evaluation process and using its findings. *The paper on the credibility of evaluations is available in the Commissioned Papers section of this report.*

Following two parallel solution forums on the credibility of evaluations,¹⁰ the plenary session highlighted that transparent sharing of findings and results enhances credibility. Further, the credibility of evaluations also depends on the skills, competencies and reputations of evaluators. Peer reviews of evaluations can keep evaluators sharp and increase evaluation quality. In addition, putting together evaluation teams with diverse membership from national and international levels—including government, academia and other stakeholders—can improve an evaluation’s credibility. Substantiating this discussion, an experience from the United States highlighted the following:

“Credibility is influenced by the choice of study design, the competence of its execution and by the author’s institutional affiliation and reputation for quality and objectivity.”

UNITED STATES

South Africa provided a further example of placing the evaluation structure within its country context in a way that promotes authority, credibility and a commitment to evaluation use:

“A significant step change occurred in 2010 with the creation of a Department of Performance M&E (DPME), which has moved very quickly to establish a range of M&E systems, including a national evaluation system. DPME has explicitly focused on trying to create a utilization-focused and demand-driven evaluation system.”

SOUTH AFRICA

A clear commitment to credibility and independence was shared by participants from Brazil:

“SAGI aims to accomplish the difficult task of guaranteeing both the independence and credibility of the research itself, and that the evaluation will be actually used to the qualification of public policies.”

BRAZIL

INNOVATIVE SOLUTIONS TO CHALLENGES LINKED TO THE INDEPENDENCE OF EVALUATION

In his keynote address, Mr Hans Martin Boehmer¹¹ highlighted the different dimensions and definitions of independence. Independence can be understood at various levels, including at the structural, organizational, individual and financial levels. Independence is important

10 The topics for the two parallel solution forums were: (1) Tools and methodologies to enhance evaluation credibility; and (2) Functional evaluation systems to promote credibility of evaluations.

11 Senior Manager, Independent Evaluation Group, Strategy, Communication, and Learning Department, World Bank.

when it prevents bias and ensures that there are no economic gains (or losses) from certain conclusions in evaluation.

Boehmer further suggested that independence is not an end in itself, but has to be seen in the context of an evaluation's ability to achieve evaluation results. For independence to have an effect on results, evaluation functions need to have an important learning role. He argued that some of the ways in which independence is interpreted or operationalized could have neutral or negative effects on the credibility and use of evaluations. Boehmer noted that as national monitoring and evaluation systems evolve, establishing better links between independence and results presents an opportunity for innovation and advancement in the monitoring and evaluation field. *The paper on the independence of evaluation is available in the Commissioned Papers section of this report.*

Following the two parallel solution forums on the independence of evaluations,¹² the plenary session noted that independence of evaluation must be considered in a country context. Champions in the government, civil society and media can promote credibility and independence of evaluations by creating demand for evaluation. It was emphasized that educating the media and the public on the value of evaluation and how evaluation results have to be interpreted can instil a culture of evaluation in the country. Participants further shared practices that promote independence of evaluations in their country context:

"The main decisions on the evaluation of social programmes and the measurement of poverty are taken by the majority of researchers who are elected, not appointed by the President or by the Minister of Social Development. This characteristic has given CONEVAL effective independence from the Executive and even from the Congress."

MEXICO

"For the purpose of routine evaluations, an independent expert evaluation team is appointed with representation from the Department of Project Management and Monitoring, independent sector specialists, academia and research institutions. Similarly, in the case of evaluations of a special nature, where capacity constraints are there, the study is outsourced to an independent academic or research institution."

SRI LANKA

"Evaluation systems ... support decision-making processes and organizational learning, ... transparency of programmes, and policies' performance and results. This kind of information is essential to increasing trust and confidence of all stakeholders in the policy process and contributes to [greater] coordination, accountability [and] more effective governance."

BRAZIL

The importance of innovations in evaluations was strongly emphasized in the conference deliberations. It was noted that countries are increasingly using innovative approaches to manage the performance of public policies, programmes and service delivery. This is because evaluations must respond to the higher demands placed on it and to the fast-changing

12 The topics for the two parallel solution forums were: (1) Enabling environment and institutional arrangements for independence of evaluations; and (2) Approaches/models for enhancing public policy accountability and learning.

environment. On the other hand, technological advances provide opportunities for innovation in evaluations. Six innovative approaches in evaluations were shared with participants: outcome harvesting, crowd sourcing, data exhaust, intelligent infrastructure, remote sensing and data visualization.¹³

In addition, innovative projects for strengthening the demand for and use of evaluation to inform policy-making, especially in the context and spirit of EvalYear, were presented at the conference. These projects are part of EvalPartners (the global movement to strengthen national evaluation capacities) Innovation Challenge.¹⁴

The conference also provided space for detailed experience sharing around the use of evaluations by the Brazilian Government in relation to its 'Brazil Without Poverty' Plan, Family Grant Programme, Cistern Programme and the Food Acquisition Programme. The Brazilian Government also shared information on its monitoring and evaluation surveys, tools and training courses. National and regional VOPEs from Africa also provided specific ideas on how VOPEs can best support governments in promoting national evaluation capacities.

At the close of the conference, it was suggested that the Fourth International Conference on National Evaluation Capacities in 2015 could be hosted in Asia. The conference concluded with a presentation of 18 agreed-to commitments proposed by the participants (*more can be read ahead in the section 'Commitments emerging from the conference'*). These commitments are particularly relevant in the context of EvalYear, as they encourage south-south cooperation and dialogue to build national evaluation capacities at international, regional and national levels.

USEFUL LINKS



Documents and papers

nec2013.org



Twitter feed (#NECbrazil was used in social media)

storify.com/UNDP_Evaluation/following-the-third-international-conference-on-na-1



Video

2013 NEC conference: youtube.com/user/evaluationoffice

Interviews with conference participants:

youtube.com/playlist?list=PLduAEjS6wFdKffErxoBwsADrSiagoYoZN



2013 Community of Practice discussion summary

nec2013.org/downloads/NEC-2013-COP-summary.pdf

13 For more information on innovation in evaluations, see nec2013.org/documents/papers/Innovations-in-mande.pdf.

14 For more information on these projects, see mymande.org/evalpartners/evalpartners_announces_the_winners_of_the_innovation_challenge.

COMMITMENTS EMERGING FROM THE CONFERENCE

Seeded by thoughtful plenary presentations, panels of government and VOPE representatives, and engaging discussions in small groups, participants came up with 18 commitments to further national evaluation capacities. These 18 commitments¹⁵ can be broadly clustered around four key strategies that can help build national evaluation systems. Some commitments may support multiple strategies.

The strategies include:

1. Promoting evaluation use through in-country and global advocacy;
2. Defining and strengthening evaluation process and methods;
3. Engaging existing and new stakeholders in exchanges and collaboration; and
4. Exploring options for different institutional structures for managing evaluation.

It is expected that UNDP partners, participants from the conference and other stakeholders will engage in opportunities to exchange knowledge and discuss how national governments and partners can cooperate under the aegis of South-South cooperation to implement these 18 commitments. UNDP is committed to developing and implementing a strategy to follow up on and monitor efforts, cooperation agreements and results of national governments and partners linked to these commitments and the International Year of Evaluation (EvalYear), 2015.¹⁶

UNDP will also help to identify and follow up on potential cooperation among conference participants and other interested parties. UNDP will link interested parties to potential partners and its own internal programmatic units able to support South-South cooperation initiatives to implement the commitments.

15 See nec2013.org/documents/nec-2013-proposed-commitment.ppt.

16 See mymande.org/evalyear.

STRATEGY	COMMITMENTS
Promote evaluation use through in-country and global advocacy	Collaborate to build and strengthen credible national data systems to improve the integrity of such systems, in order to better link performance of policies and programmes.
	Develop systems to promote the transparent follow-up of evaluations, such as management response tracking systems and citizens' commission that allow for effective monitoring of the implementation of evaluation recommendations .
	Create/strengthen Parliamentarians' Forums for development evaluation in different regions to advocate for use and conduct of evaluations.
	Develop/connect national registries/national statistical systems to monitoring and evaluation systems with increased frequency of data collection to support evaluations and decision-making.
	Assign resources (a percentage of the initiatives' costs) for the conduct of evaluations when designing/approving projects/programmes/policies.
	Use external evaluators to facilitate/moderate self-assessments and reviews.
Define and strengthen evaluation process and methods	Develop approaches, based on lessons learned, on how to incorporate cultural dimensions into evaluation in different regional and national contexts.
	Develop standards , based on lessons learned, to ensure proper triangulation of evidence, checks and balances, and qualitative data use, not just perception-based.
	Develop standards , based on lessons learned, to ensure stakeholders' involvement in evaluations while still guaranteeing the independence of assessments.
	Incorporate gender capacities/perspectives in national monitoring and evaluation systems.
Engage existing and new stakeholders in exchanges and collaboration	Develop and implement a transparent results-based monitoring and evaluation framework to track the efforts and results of the implemented commitments proposed in this conference.
	Develop/strengthen/support/expand joint peer-to-peer systems and mentoring programmes among professional associations of evaluators and government evaluation units.
	Facilitate partnership/cooperation among governments, VOPEs, parliaments and private-sector initiatives to strengthen the understanding of what evaluation is and how it can be useful for different actions.
	Support joint regional/national events to take stock of developments in these commitments (in 2014), including the sharing/learning of good practices to validate data from multiple sources, manage sensitive data and disseminate evaluation results.
Explore options for different institutional structures for managing evaluation	Study the alternatives, assessing the pros and cons of different options of institutional set-ups for evaluation , such as national evaluation legislation/ policy, where appropriate, taking the country context into account, and establishing a set of minimum requirements based on lessons learned.
	Have an online platform (NEC Community of Practice) to present/exchange experiences, keep participants connected and follow up on the commitments.
	Translate material on evaluation into different languages.
	Map and analyse effectiveness of coordination mechanisms and practices between central evaluation units and sector ministry units and local government evaluations.

CONCEPTUAL FRAMEWORK

INTRODUCTION

Evaluating the performance of public policy is considered a fundamental ingredient that fosters accountability, promotes good governance and improves programme effectiveness. Evaluating the performance of public policy can help deepen democracy by creating the conditions for holding governments accountable for their performance and increasing transparency. Evaluation of public action is embedded in a country's political, economic and social context and needs to take history and national context into account. Countries have different institutional configurations for assessing the performance of public actions, and evaluation should be one of them.

Given the complexity of institutionalizing monitoring and evaluation (M&E) systems, it is not surprising that there are a number of challenges at each stage of the process. Efforts to build and sustain effective evaluation systems face challenges of institutional design, political dynamics, technical capacity and resistance to change. Successive UN General Assembly resolutions, including the Quadrennial Comprehensive Policy Review of operational activities for development of the United Nations system¹⁷ and UNDP Executive Board decisions, have encouraged the UN development system and UNDP in particular to support national evaluation capacities.¹⁸

OBJECTIVES

To enhance the understanding and appreciation of evaluation as a powerful tool of public accountability and learning, the conference had the following objectives:

- a) To share experiences, challenges and solutions from countries with national M&E systems, including those that may be considering creating one or have important experiences with other types of evaluation efforts;

17 UN General Assembly resolution A/RES/67/226, Quadrennial Comprehensive Policy Review. New York, 22 January 2013. See un.org/en/ga/search/view_doc.asp?symbol=A/RES/67/226.

18 UNDP Executive Board, DP/2011/3, *Evaluation Policy of UNDP*. New York, January 2011.

- b) To identify lessons and constraints in implementing national M&E systems and the use of evaluation; and
- c) To identify opportunities for technical cooperation in strengthening institutional capacity for national M&E systems under the umbrella of South-South and triangular cooperation.

Following two earlier international conferences on national evaluation capacities, the Third International Conference on National Evaluation Capacities provided a platform to share experiences and expand on solutions to address common challenges related to establishing national M&E systems of public policies and initiatives. Recognizing that monitoring and evaluation are closely related, the conference concentrated more on evaluation than on monitoring.

The conference placed emphasis on three interconnected aspects of establishing national M&E systems: how to ensure the independence, credibility and use of evaluations. As the third exercise of this nature, the conference learned from the experience gained by the Independent Evaluation Office of UNDP in organizing the 2009 and 2011 conferences. Among other benefits, this resulted in better follow-up mechanisms, dialogue continuity, and the continuity of networks and agreements reached. Due to the participatory nature of the conference, the results will help improve international standards of evaluation in its application to public programmes, policies and initiatives.

INDEPENDENCE OF EVALUATIONS

Meeting the requirement of independence generally implies freedom from political influence and/or organizational pressure. According to the norms and standards set out by the United Nations Evaluation Group (UNEG), an evaluation is said to be independent when it is “objective, free from undue influence, and has full authority to submit reports directly to appropriate levels of decision-making.” This definition also implies that management must not impose restrictions on the scope, content, comments and recommendations of evaluation reports. In order to avoid conflicts of interest, evaluators must not be directly involved in policy setting, design, implementation or management of the subject of evaluation. Evaluators should, however, have full access to all relevant information required for the evaluation.¹⁹

The conference addressed the challenges faced by governments when establishing monitoring and evaluation systems that are considered independent. It answered questions such as: Who is responsible for evaluating public programmes and policies in a country? Who commissions evaluations? Is the evaluation function located independently from other management functions? How are conflicts of interests or undue pressures over evaluators avoided? Can evaluations be submitted to the appropriate level of decision-making?

¹⁹ See “UNEG Norms and Standards for Evaluation in the UN System”, New York, 2005. This definition of evaluation independence is in line with the one provided by the Development Assistance Committee of the OECD in the 2002 “Glossary of Key Terms in Evaluation and Results-Based Management”, available at oecd.org/fr/cad/evaluation/glossaryofkeytermsinevaluationandresultsbasedmanagement.htm.

The independence of the evaluation function and its location in the public administration triggered interesting discussions conference.

The conference also addressed the complementary role that could be played by self-evaluation and independent evaluation. How are they interrelated? Are mechanisms in place for checks and balances between spheres of the state, for example between the executive and legislative branches? How do citizens participate in the process?

CREDIBILITY OF EVALUATIONS

The credibility of evaluation generally depends on the expertise and independence of the evaluators, the degree of transparency of the evaluation process and the quality of evaluation outputs. Credibility requires that evaluations should report successes as well as failures. Beneficiaries of programmes or policies should, ideally, participate in evaluations in order to promote credibility and commitment.

The conference addressed the challenges faced by governments when ensuring the credibility of evaluation systems. It answered questions such as: Are evaluations designed, planned and implemented following international quality standards in evaluation? Are the evaluations explicit in the use of methodologies for data collection, analysis and interpretation? Do evaluators have access to all the required information to conduct evaluations? How do the systems ensure the quality of evaluations? Are evaluations conducted in an ethical way? Is data integrity respected, or is there manipulation of quantitative and qualitative data to influence findings and conclusions? Do evaluations consider issues linked to human rights, gender equality, ethnicity and cultural behaviour? How is credibility in evaluation processes enhanced and ensured?

The conference also addressed the complementary roles of monitoring and evaluation. There is a need for a strong link between monitoring and evaluation; however, it is important to mark the distinction, because only evaluation can question the assumptions behind public action. In that sense, evaluation can question if the state is doing things right—and also if it is doing the right thing—something that a monitoring exercise is unable to assess.

UTILITY OF EVALUATIONS

For evaluation to have an impact on decision-making, evaluation findings must be perceived as relevant and useful and be clearly and concisely presented. They should fully reflect the interests and needs of different stakeholders. Importantly, ensuring the utility of evaluations is only partly under the control of the evaluation function. Commissioning, receiving and using evaluations is also a responsibility of decision makers within the state and societal structures.

The conference addressed the challenges faced by government institutions for better use of evaluation systems, but went one step ahead of the 2011 International Conference on National Evaluation Capacities that focused on the use of evaluation by, on one hand, assessing the country cases in which evaluation use has moved significantly ahead and, on the other hand, portraying documented cases of linkages among evaluation use, independence and credibility in conducting evaluations.

The conference addressed questions such as: Are there incentives to motivate potential users to demand and/or to use the evaluations as input for their own decision-making processes? Which stakeholders seem to be most interested in demanding and using evaluations and why? What happens after an evaluation is conducted? Do evaluations consider approaches that would call attention to use of evaluation findings in an inclusive and comprehensive way for decision-making? Do evaluations appropriately consider gender equality and human rights?

The conference also addressed linkages between evaluation and other processes, such as planning and budgeting. Is the availability of evaluation synchronized to government cycles and decision-making processes (for example, the budgetary and planning cycles at the country and agency levels). Are evaluations linked to planning processes? What is the timing for evaluations? Are evaluations linked to budgetary allocations? How is accountability ensured? What can be learned from evaluations?

CONFERENCE FORMAT

The conference adopted the form of a dialogue that sought to collectively arrive at answers to questions, differing from a debate format where participants try to prove the validity of their positions. A dialogue is a common search for meaning and a way of addressing common challenges. All the participants discussed each topic separately and in sequence. Four discussion papers were commissioned to evaluation experts as input for dialogue, including papers on independence of the evaluation function, credibility of evaluations, use of evaluation and innovations in evaluation. Each paper posed questions to conference participants to start a dialogue. The papers identified sub-themes that were discussed in smaller groups. The authors of the papers were invited as keynote speakers. Participants were invited to share their national experience in addressing the challenges under analysis.

The conference provided space for an open exchange of experiences. Participants described the roles played by their institutions in addressing one of the conference topics. The conference issued a call for short country papers; papers were shared in advance and now constitute a critical input to this report.

PARTICIPANTS

The challenges faced in starting national M&E systems of public policies are not exclusive to developing countries. For this reason, the conference invited representatives of national institutions responsible for commissioning, conducting and using evaluations of public policies of programme and non-programme countries. This approach facilitated the exchange of experiences and learning from each other.

- The representatives of national institutions commissioning evaluations dealt mainly with the topic of evaluation independence.
- The representatives of institutions conducting evaluations contributed to the dialogue on evaluation credibility.
- The representatives responsible for using evaluations in national decision-making processes contributed to the dialogue on evaluation use.

COMMISSIONED PAPERS

THIS SECTION INCLUDES FOUR COMMISSIONED PAPERS IN THE AREA OF USE, CREDIBILITY, INDEPENDENCE AND INNOVATIONS IN EVALUATION.

1. ENCOURAGING THE EFFECTIVE USE OF EVALUATIONS TO IMPROVE PROGRAMME MANAGEMENT AND DESIGN
2. INNOVATIVE SOLUTIONS TO CHALLENGES LINKED TO INDEPENDENCE, CREDIBILITY AND USE OF EVALUATIONS
3. REFLECTIONS ON INDEPENDENCE IN EVALUATION
4. INNOVATIONS IN MONITORING AND EVALUATING RESULTS

1. ENCOURAGING THE EFFECTIVE USE OF EVALUATIONS TO IMPROVE PROGRAMME MANAGEMENT AND DESIGN

PAULO MARTINO JANNUZZI

Secretary for Evaluation and Information Management
Ministry of Social Development and Fight against Hunger in Brazil

Professor, National School of Statistical Sciences
Brazilian Institute of Geography and Statistics

Researcher, Statistical Information and Monitoring and Evaluation
of Policies and Social Programmes in Brazil and Latin America project
National Research Council

INTRODUCTION

Around the world, universities, research centres, multilateral agencies, public administration officials and others are rapidly increasing the production of information and knowledge about public policies and programmes. In many countries, recurring subjects of applied research into public policies include education and public health programmes, income transfer programmes and actions to fight hunger and poverty. These research efforts mobilize sociologists, economists, statisticians and other monitoring and evaluation (M&E) professionals. The number of academic journals that are dedicated to this theme and the conferences that deal with M&E matters (e.g. two international conferences on national evaluation capacities held in Morocco and South Africa) highlight the relevance of public policy information and knowledge.

However, although M&E systems have produced vast amounts of empirical evidence and comprehensive and rigorous evaluation studies on such policies and programmes, it seems additional efforts are necessary to ensure that the information and knowledge produced is effectively used to formulate public policy and to improve routine programme activities.

Public programmes are complex systems that involve various processes and activities. So, in general, specific and rigorous public policy information and knowledge are dense and difficult to understand—even by the staff involved in formulating and coordinating programmes. Programmes involve many different agents in their daily operations,

with many different skills and learning capabilities. Depending on the country, resources and programme coverage and complexity, thousands—or even hundreds of thousands—of agents contribute to management, preparation and delivery of products, services and benefits. Agents can make a difference to programme improvement if they can understand the information and knowledge produced by M&E systems.

Although policymakers, managers and programme staff do not need exhaustive data or studies about their programmes, they do need information that is clear, consistent and relevant to decision-making. Data should be organized geographically and by operational issue, answer questions relating to the programme's implementation stage, and include information on costs, deliveries, outputs and outcomes. Good and relevant information and knowledge is customized to the different needs at the formulation, monitoring or summative evaluation stage of public policy and programmes.

The main idea this paper discusses is that, given the operational complexities of programme management and its need for innovation, the effective use of evidence depends largely on the relevance of information and knowledge to the formulation, decision and managing processes. This dependency holds whether the data was gathered by monitoring panels, evaluation surveys or studies. Effectiveness also depends on the strategies used to disseminate information and knowledge to personnel involved in the programme, from the field or street-level bureaucrats to the strategic decision-makers. Dissemination strategies of customized information and knowledge to all technical staff involved in public policy can make the difference to the challenge of incorporating changes into programme design and operations.

This paper presents the argument that innovation in public programmes (through effective use of information and knowledge), seems to depend less on the technical sophistication and independence of evaluation study and more on the clarity and objectivity that the information and knowledge responds to the specific needs of technical staff and managers. This does not deny the importance of robust evaluation studies. However, if studies do not answer the most crucial demands for information from the perspective of those involved, there is a risk that the resulting data and information will be of minimal use.

This paper is organized into two sections. It begins with a more conceptual discussion of M&E systems and their integration into the policy and public programme cycle. The extent to which managers and staff are interested in—and actually use—the information and knowledge produced in an M&E system depends on the adequacy of its design and purpose to answer those questions that the programme team and officials consider relevant to improving the programme. Therefore, the nature of the questions the M&E system answers and the choice of instruments (e.g. monitoring indicators, implementation studies, impact and results surveys) determine managers' and technical staff's involvement and interest in using the system's products. Besides credibility and independence, evaluation studies must be relevant to their potential users.

The second section of this paper is dedicated to the different strategies used to disseminate monitoring tools and evaluation studies for public users of the M&E system. Information and knowledge produced in this environment can be complex and not easily assimilated by

managers and technical staff. It is insufficient to merely produce an extensive research report or to have an informative online application with multiple functionalities. Lectures, publications and courses need to be tailored to their audiences, seeking to present them with an evaluation study's most relevant and interesting aspects.

PRODUCING INFORMATION AND KNOWLEDGE RELEVANT AND USEFUL TO MANAGEMENT AND PROGRAMME IMPROVEMENT

M&E systems have many conceptual definitions in specialized literature, which, according to MacDavid and Hawthorne (2006), Mackay (2007), Owen (2007), and Cunill and Ospina (2008), can be broader or more operational. Using these concepts, M&E systems can be defined as a set of articulated processes for raising, organizing, analysing and disseminating information and knowledge about public policy and programmes. These processes vary according to the different needs of decision makers and operational managers, extend over the policy and programme life cycle, and include diagnoses of the social problem, the formulation and design of programmatic solutions, actual implementations in the field and overall evaluation stages. Processes aim to support improvements to a programme's design and management, to ensure greater transparency of government action, and to provide evidence on the merits and effectiveness of policies and programmes.

This definition makes it clear that M&E systems provide knowledge and information for analysing product and service delivery, correcting any failures of government action, identifying the impacts of policies and programmes, and determining the costs of production of programme delivery. By definition, M&E systems are important mechanisms for ensuring greater transparency in the use of public resources. They also contribute to decisions relating to the merits and relevance of policies and programmes.

However complementary, the three basic purposes of a M&E systems—information to help improve programmes, public transparency and budget merit evaluation—compete with each other over evaluative efforts and available human resources, and largely define the methods and techniques chosen for the work. Of course, the primary purpose, the evaluation focus and, consequently, the effective use of M&E products depends on where such a system is based. For example, if an M&E system is based in a sectoral ministry or programme agency, the creation of monitoring instruments and evaluation research will aim to provide the means to continuously improve the programme's implementation and results. If the M&E system is based in a public control body or parliament, the evaluative focus will be on producing and organizing information on the results and impacts of public policies and programmes on society. If a system is based in a body responsible for budget management and/or medium-term planning, it is natural that the processes and activities will be oriented towards producing studies on the cost-effectiveness and impacts of public programmes, and guiding public resource allocation.

Clarity over the evaluative focus of an M&E system is a key factor in ensuring that technical staff involved in policy and programme implementation, civil society, parliament and budget managers successfully and effectively use evaluative information and knowledge. The evaluative focus also helps to orient the main efforts and activities of M&E staff. This, therefore,

determines the choice of instruments and methods used to generate relevant and useful information and knowledge.

M&E systems that are oriented towards the needs of management and programme improvement (about which this paper is particularly concerned) are characterized by research designs, surveys, information systems or monitoring indicators that focus on specific aspects of programme implementation. Such systems generally depend on qualitative methodological strategies, such as discussion groups and in-depth interviews, taking field staff and programme beneficiaries as main sources of information. The aim is to generate rapid empirical evidence concerning the programme's management and any difficulties experienced during planned implementation.

As part of field evaluation, surveys may not need to be performed if the set of monitoring indicators, created from the programme management systems' database, is able to answer basic evaluative questions. Appropriate choices of key indicators with a detailed geographic and demographic focus may provide valuable, accessible information and can be used effectively by technical staff and managers. As 'thermometers', these indicators may diagnose 'fever' (or healthy status) at critical points in a programme's intervention model. This can help technical staff and managers make informed decisions on how to address problems or enable such staff to commission specific research (or 'clinical investigation', to continue the metaphor) to investigate the causes of implementation problems (or the fever's causes and the reasons for its persistence). (Jannuzzi 2011a)

Nationally representative sample surveys and research with a quasi-experimental design are certainly important tools and products of M&E programme improvement for sectorial ministries. However, the time and effort they require makes them more useful as ways for transparency and budget practitioners to appraise merit, legitimacy and impact; instead, M&E programme improvement staff should be involved in a broader technical agenda. Staff time and concerns cannot be captured by impact or national evaluation surveys.

It should be recognized that in order to legitimize the political priorities given to certain social issues, and in the interests of public accountability and efficient use of scarce public finances, quantitative research using probability samples (such as those conducted by national statistical agencies) and impact assessments with control groups and beneficiaries offer important measures of the adequacy of public programme design, coverage, beneficiaries, results and impacts, and distinctiveness. However, in deciding whether to maintain, modify or discontinue a policy or programme, evaluation studies are not the only inputs. Such decisions are not merely technical; rather, they are primarily political, because they have implications for the lives of beneficiaries and for the programme's institutional arrangements.

Large surveys, or those that are methodologically or operationally complex, can be justified at the outset of a policy or programme in order to define the situation the policy or programme seeks to address. Further surveys of a similar scale, however, should wait until after any problems in programme implementation have been identified and resolved (Rossi et al. 2004). Otherwise, implementation problems may cause evaluation studies to conclude that a programme's outcomes and impacts are minimal or absent. In turn, this may create a

hasty mistrust of public opinion regarding the public policy or programme's merits and may negatively impact officials' perceptions of the utility of M&E products and research. The fact is that, despite the prestige conferred by certain academic communities, quasi-experimental evaluation research is not the most legitimate scientific approach, nor is it the gold standard for programme evaluation (Worthern et al. 2004).

There are several ethical conflicts and operational constraints on their realization that have been widely noted in international literature (Jannuzzi 2011b). Moreover, daily programme management requires packets of information and knowledge far beyond those produced by such research designs. Rather than producing evidence for a 'revolutionary innovation' as intended by impact assessments, it is necessary to have information that can contribute to the continuous improvement and incremental innovation of public action. Without denying the importance of empirical evidence that is nationally representative, collected with technical rigour and analytically deep, managers and technical staff involved in programme implementation need a more eclectic methodological approach to gathering information, as well as knowledge of the complexity of social problems and programme operations.

Misconceptions about the relevance of a mixed-method approach (rather than a quantitative and impact evaluation focus) have contributed to scepticism about the value of M&E tools among managers and technical staff in the public sector. Thus, in the context of scarce human and financial resources, M&E systems should focus on responding to management and programme needs, using a structure of centralized databases extracted from computerized management systems or from the countless spreadsheets and paper controls executed by programme managers. These M&E tools may include key programme indicator dashboards, logical framework processes or research recommendations on dealing with implementation issues. In situations of limited resources and time, it can be more useful and productive for programme management teams to use evaluation studies of secondary data, study reviews, international publications and meta-evaluations of similar programmes in other countries.

In addition to clarity of evaluative focus and appropriate choices of methodology, if an M&E system is to offer information and useful knowledge to improve programmes, it is essential that the system secures the participation of technical staff and managers in drafting instruments. If it is true that external teams can ensure technical credibility for an evaluation study (assuming that they are competent, reputable and committed to a multidisciplinary view and to a mixed-method evaluation approach), then the relevance and ownership of results depends on the extent of programme managers' and technical staff's involvement in evaluation processes. Though internal teams typically know more about a programme's most pressing problems and difficulties, they nonetheless often need technical support from specialized consultants (as well as information from field surveys) in order to properly understand the causes of such issues and to recognize appropriate resolution strategies.

Running field evaluation surveys or conducting evaluation studies based on secondary data requires skilled people and teams. However, it is worth noting that the professional market for evaluation consultants is far from perfect in many developing—and even some

developed—countries. It must be recognized that increased demand for the evaluation of social programmes has outstripped the availability of properly qualified private consulting firms and academic research groups. The pool of evaluators is dominated by companies specializing in public opinion and market research, areas where problems require less complex design and effort to understand compared to social programmes. Although these evaluators may have a good academic pedigree, the firms' or consultants' knowledge of the reality of public programmes is often limited. Social surveys, particularly those related to programme evaluation, can be more complex than market or academic research and therefore require a more robust, specific and responsible approach, particularly because the results will guide critical decisions about the design, results and merits of government actions.

In this context, it is necessary for the technical teams of the M&E system to intensely monitor the contracted evaluation study. Technical teams should not just leave the hired firm or consultant to design the survey sample, the questionnaires or the training and supervision of field staff. If programme managers want the answers to specific problems, the same managers must participate in all evaluation processes. Not all contracted companies appreciate the experience of having their technical procedures questioned or altered by internal teams. However, mixed teams of contracted and internal personnel can help achieve a synergy of internal knowledge management and external technical expertise. This can help create knowledge products by combining their respective areas of expertise, and may increase the overall legitimacy and relevance of the evaluation effort.

The credibility of results and the legitimacy of evaluation processes are two values that must be pursued jointly; making technical and political choices based on studies and knowledge that are limited in their operational survey and analytical scope is worse than not having information for decision-making. In some situations, it may be preferable to have no evaluation than to rely on misconceived, mishandled or rashly contracted research.

DISSEMINATING RELEVANT AND USEFUL INFORMATION AND KNOWLEDGE FOR MANAGEMENT AND THE IMPROVEMENT OF PROGRAMMES

Formulating, evaluating and managing policies and programmes requires, like any other activity in human organizations, training of technical staff and managers that are involved in decision-making and those involved in providing services (i.e. field workers). Leadership, communication, ethics and response to the public are among the training topics for technical staff and managers, and are as important as project management and evaluation methodology development. Experience suggests that a significant reason that public programmes fail or lack impact lies in the difficulties in maintaining continuing education programmes for teams involved in policy development.

The impact of public programmes would certainly be higher if those involved understood more about programmes' objectives, logical design and related activities, the role of each institution and staff member, and the characteristics of target beneficiaries. Although courses and operational training may have been planned for in the framework of many social programmes, they may not be fully adapted with materials, regulatory documents, classrooms and teaching staff to train multiple technical persons involved. In some situations,

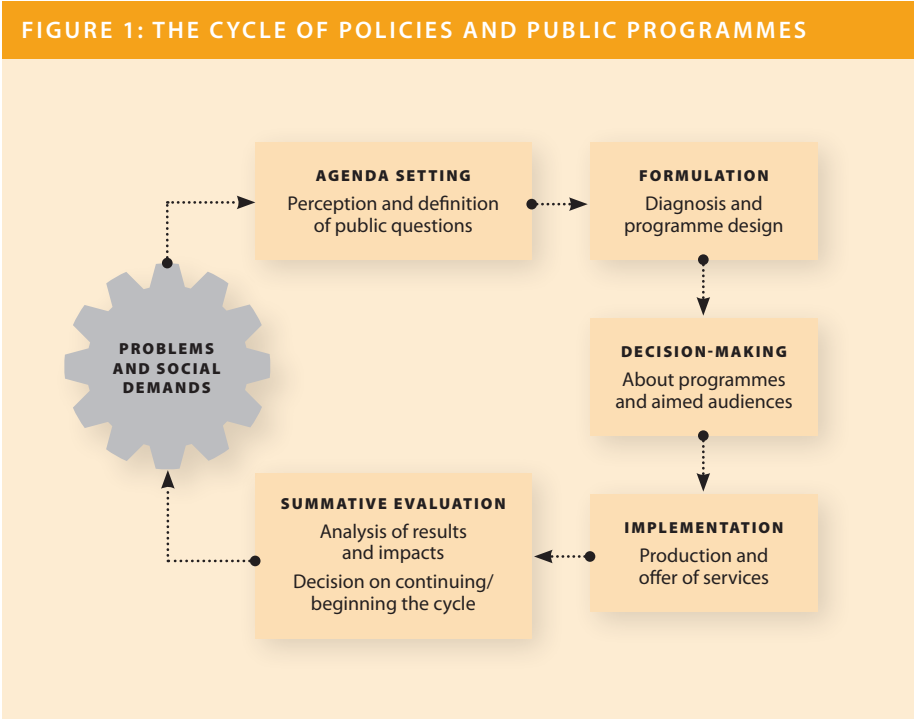
the trainees may not even be engaged or informed about the training. There is much to be done in terms of training human resources involved in public service delivery and management of public programmes (an issue beyond the scope of this paper). However, it must be a matter of concern for multilateral organizations, with the same emphasis attached to the dissemination of methods and techniques as to planning and evaluating social programmes and projects.

This section addresses two central issues: dissemination strategies in M&E and training on M&E tools. Both are important for enhancing the informed use of M&E products and studies, particularly those designed to improve public programmes. As with information in science, technology and innovation, information and knowledge in public policy are complex and require training programmes. Although indicators such as the infant mortality rate or monetary extreme poverty are part of the technical vocabulary of evaluators and the academic research community working on evaluation of social programmes, they are not necessarily part of the vocabulary of programme managers and technical staff. Similarly, evaluation reports and their results may be differently understood by evaluators and a programme's technical staff.

If the knowledge produced by M&E is to reach broader audiences, it is necessary to make its products (e.g. reports, indicators, evaluation studies) understandable and attractive to a range of public users. It is not enough to simply post all data sheets, indicators and evaluation reports on the Internet. Data production does not generate demand for knowledge. It is necessary to develop tailored products for targeted audiences of technical staff and managers by appropriately adapting format, content and complexity. Results from evaluations should also be disseminated through lectures or multimedia recordings, and they should be readily accessible to Internet users.

There are a number of Internet-based virtual applications that provide many interactive and visual resources and links to other documents. Executive summaries of evaluation reports, small datasheets (one-page papers) with the essential results with graphs, maps and descriptive reports may have a utility and aesthetic appeal greater than that of tables, dashboards or massive publications with content that is inscrutable for those without specialized training. Results of econometric models developed with evaluation data are frequently presented, but have limited capacity for diffusion to the uninitiated public. It is surely possible to make such outcomes more tangible and concrete for technical staff and managers who want to learn more about programmes.

Efforts to electronically disseminate M&E content to technical staff and managers may be more effective when combined with continuing education, in either classroom or distance-learning settings. There are always technical staff and managers interested in deepening their knowledge of M&E but unable to find an appropriate, relevant course in a university or research centre. M&E training programmes should be organized for technical staff and programme managers interested in developing their skills and improving their understanding of monitoring tools, evaluation and the application of information and knowledge. The training programmes should be organized using the basic cycle of policy and programmatic processes (see Figure 1).



In classical political science textbooks, the public policy formulation process has been repeatedly presented as the cycle of successive steps, with a number of stages (Jann and Wegrich 2007). Despite long-standing criticism of the simplified way in which this diagram shows the political process as an empirical truth, the separation of steps demonstrates that the process gives different emphasis to programme planning, implementation and evaluation. This model lends itself well to teaching, particularly for the way it contextualizes the issue for technical staff and programme managers.

In this model, the first step, agenda setting, defines the political agenda and corresponds to the multiple paths and processes that culminate in recognizing a social issue as a public problem and the need for government action to solve it. In other words, it legitimizes the introduction of the issue on the policy agenda. The next step, formulation, refers to the processes and activities involved in developing possible solutions, legislation and programmes to deal with the defined social issue. In the next step, decision-making, crucial choices are made on the interventional model, institutional arrangements and the target audience—narrower or broader—considering the feasibility of alternative solutions and their budgetary implications. The fourth step, implementation, corresponds to launching the actions, allocating resources and developing processes to guarantee the delivery of public programmes. Finally, the summative evaluation of policies and programmes reviews the extent to which the work is

TABLE 1: BASIC TRAINING PROGRAMME IN M&E

COURSE	OBJECTIVE	TIMETABLE
1. Diagnostics for programme formulation	Develop capacity to use information sources from programmes and official statistics to diagnose a situation and propose a public programme.	20 to 40 hours
2. Tools and indicators for programmes monitoring	Prepare participants to use information systems and to develop methodologies to build indicators for public programme monitoring.	20 to 40 hours
3. Introduction to research methods and evaluation studies	Develop skills essential for the understanding evaluation results and methodologies, their potential and limitations.	40 hours

solving the defined problem. This step assesses any requirement to change the programme in order to ensure its effectiveness, to discontinue the programme if the problem is no longer part of the agenda or to adapt to a new reality, restarting the cycle.

It is appropriate to note that, as part of the cycle, evaluation takes place after implementation. It is a more reflective process that helps inform the decision to continue or stop the programme. It is distinct from M&E activities, which are characterized by indicators (among other things). However, such investigative tools can be used at any time during the cycle. Re-naming this decisive stage of the cycle to ‘summative evaluation’ could help to avoid confusion between these two distinct activities.

Although there are different ways to implement a training programme based on this cycle, it would seem appropriate to organize it into three modules, each of 40 hours. Its content should include diagnosis/formulation of programmes, research tools and M&E studies, with complementary objectives. The course will become gradually more complex, as envisaged in Table 1, dealing with construction of indicators in the beginning and the methodologies of social research used on evaluation studies in the final stage. In addition to presenting M&E concepts and methodologies, it will be important for programme managers and technical staff to submit case studies from their own experience that show the effective use of course content.

FINAL CONSIDERATIONS

The effective use of M&E products and surveys depends very much on factors related to the supply of information and knowledge produced and to the demand from potential users. Evaluation can focus on the production of information to improve programme management, to produce data for public transparency and/or to inform budget decisions. Those evaluations are targeted at a range of users with different demands for information and knowledge.

Once the focus is established, the methods used to develop knowledge products must be appropriate in terms of content, costs and schedule so that they meet their users' needs. Information and knowledge produced by M&E systems are complex, and efforts should be made to disseminate them with the most proper strategies—as customized publications and training courses. As the publishing market has demonstrated, especially with the advent of the Internet, there are many different and creative ways to communicate simple or complex messages to a range of audiences.

There is certainly much to be done to maximize the effective use of public policy assessments. National experiences presented at the Third International Conference on National Evaluation Capacities in 2013, as those presented during previous conferences, have been showing how different countries are dealing with it. Let's share our experiences and challenges!

REFERENCES

- Garcia, Ronaldo Coutinho. 2001. "Subsídios para organizar avaliações da Ação Governamental", in *Textos para discussão*, Brasília, n. 776: IPEA.
- Grau, Nuria Cunill, and Sonia Ospina. 2008. *Fortalecimiento de los sistemas de monitoreo y evaluación (M&E) en América Latina: informe comparativo de 12 países*. Washington: Banco Mundial.
- Jann, Werner and Kai Wegrich. 2007. "Theories of the Policy Cycle." In *Handbook of Public Policy Analysis: Theory, Politics and Methods*, edited by Frank Fischer, Gerald J. Miller and Mara S. Sidney, 43-62. Boca Raton: CRC Press.
- Jannuzzi, P.M. 2011a. "Avaliação de programemas sociais no Brasil: repensando práticas e metodologias das pesquisas avaliativas". In *Planejamento e políticas públicas*. Brasília, 36: 251-275.
- Jannuzzi, P.M. 2011b. "Monitoramento analítico como ferramenta para aprimoramento da gestão de programemas sociais". In *Revista Brasileira de Monitoramento e Avaliação*, Brasília, 1: 38-66.
- Mackay, Keith Robin. 2007. *How to Build Monitoring and Evaluation Systems To Support Better Government*. Washington D.C.: World Bank.
- McDavid, James C. and Laura Hawthorn, eds. 2006. *Programme Evaluation and Performance Measurement: An Introduction to Practice*. Thousand Oaks: Sage.
- Owen, John M. 2004. *Programme Evaluation: Forms and Approaches*. New York: Guilford Press.
- Rossi, Peter H., Mark W. Lipsey and Howard E. Freeman. 2004. *Evaluation: A Systematic Approach*. Thousand Oaks: Sage.
- Worthen, B.R. et al. 2004. *Avaliação de programemas: Concepções e práticas*. São Paulo: EdUsp/Gente.

2. INNOVATIVE SOLUTIONS TO CHALLENGES LINKED TO INDEPENDENCE, CREDIBILITY AND USE OF EVALUATIONS

MARÍA BUSTELO

Associate Professor, Department of Political Science and Public Administration
Complutense University of Madrid

This paper opens a discussion on the credibility of evaluation and on how to enhance and ensure credibility. Many issues affect the credibility of an evaluation—the expertise and independence of the evaluators, the degree of transparency in the evaluation process and the quality of outputs, to name but some. The cultural context is also important—the values on which an evaluation rests, the way that evidence is deemed credible, the institutions that support evaluation systems and structures, the people that contribute to an evaluation, and how the evaluation is shared, communicated and reported.

My standing point in writing this paper comes from my experience training evaluators and conducting the Masters programme on Evaluation of Programmes and Public Policies at the Universidad Complutense de Madrid (Spain), which started its 12th course in 2013-2014. It also comes from my service at the Board of the European Evaluation Society since 2009, the last two years of which I have been honoured to lead the presidency of this regional society. For context, I will be using Spain, which is my own country where I have been working on evaluation for the last 25 years, but also the more ample and diverse European region. And I will be referring more than once to the inspiring European Evaluation Society Public Hearing at the European Parliament on ‘Evaluation in Democracy’, held in Brussels in April 2013.²⁰

THE POLITICAL NATURE OF EVALUATION AND THE NEED FOR CREDIBILITY

The political nature of evaluation has been amply recognized by the evaluation community since the late 1980s. Eleanor Chelimsky, in her excellent keynote address at the American Evaluation Association's annual meeting in San Diego in 1997 (which was later published as an article), discussed "the role of experience in formulating theories of evaluation practice," noting that evaluation and politics are "viscerally connected" (Chelimsky 1998). For her, "evaluations have been deeply affected by the way politics works in a democratic, pluralistic society, and especially by the continually changing nature of the political environment." Also, because "in a world of highly sophisticated and continuous jockeying for political advantage, advocacy abounds. Not only do policymakers have their own political agendas, they are also besieged by pressure groups, vested interests and lobbyists, all with their war stories about 'success' or 'failure,' and all trying, with money, power, and data, to move policies and programs in specific directions" (Chelimsky 1998, 37-38). So we need to be aware of and include that political nature in the very heart of the evaluation concept and theories.

The need for credibility is particularly important, considering that politics are central to evaluation practices. For Chelimsky, credibility implies impartiality, and that should be something that is preserved in the evaluation process, where evaluations need to be perceived as impartial. So not only should they be impartial, they should also be seen as impartial by stakeholders. In Chelimsky's substantial experience in the United States General Accounting Office (now the Government Accountability Office), this impartiality was often preserved by correcting possible perceptions of an evaluator's bias by hiring the opposite bias in the same evaluation team or by correcting for potential bias during the evaluation process (especially during the literature review, the methodological design and the report phases). These procedures were useful in a context such as the General Accountability Office, which is part of the legislative branch and directly serves the Congress. In other contexts, evaluation credibility might be more derived from the transparency of the process than from the absence of biases. Irrespective of the strategy, it is crucial that an evaluation is conducted so that it is perceived by the general public and the stakeholders as credible.

Indeed, even in contexts less politicized than the General Accountability Office, the political nature of evaluation poses challenges for the independence, credibility and use of evaluations. And that is why evaluation is about much more than dealing with methodologies and techniques that allow us to acquire good enough evidence of what has happened in a concrete project, programme or policy. Or, as I like to say, the methodology of evaluation needs to go far beyond what is understood as social science research, including most of what is called 'applied social research'. We need specific ways to think about and conduct evaluations, ways that are unique to evaluation and are embedded in its political nature and purpose—improvement, accountability and enlightenment.

The political nature of evaluation means that, apart from providing systematic and rigorous evidence, it should rise, be tailored and respond to and from a particular context to be credible and usable. This political nature should also be acknowledged in independent evaluations. Even if an evaluation appears to be apolitical, it is invariably and inevitably political because it is formed from the interactions of actors with different perspectives, and

power positions and interests (some of them in apparent contradiction, but all of them legitimate). There is interdependency among the actors, and they operate in a concrete territory and period of time. Any project, programme or policy I can think of that can be subject to evaluation can be described in such a context.

This important presence of different actors and stakeholders brings us to the next point.

WHO PARTICIPATES AND HAS SAY IN AN EVALUATION? A KEY ISSUE FOR CREDIBILITY

Credibility is also inevitably linked to the participation of stakeholders in evaluation processes. A participatory approach to evaluation, apart from the benefits of inclusiveness, promotes stakeholder ownership of evaluation processes and results. Consequently, it also raises perceptions of the credibility of those processes and results by those who have participated and feel they have a say in the evaluation.²¹ Of course, this is easier said than done, and real stakeholder participation is a process that requires time and political sensitivity by commissioners and evaluation teams.

Some of the ways to involve stakeholders include getting to know and understand the context and the programme to be evaluated, identifying key stakeholders and their information needs (including those needs in the evaluation questions and contrasting or even negotiating them with stakeholders), gathering systematic information and evidence from stakeholders and other sources, and contrasting and/or elaborating collaborative conclusions and recommendations. This will help the evaluation to be perceived as more credible, because the process has been transparent and stakeholders have had opportunities to contribute to several parts of the process.

Often, the context in which an evaluation takes place, or constraints of time and resources and of terms of reference, do not allow real stakeholder participation in all parts of the process. But we should be aware that to enhance credibility, it is necessary to methodologically design an evaluation process in which stakeholder participation in some parts of the evaluation process is possible.

So far, we have discussed stakeholder participation as a methodological requirement for enhancing ownership and hence evaluation credibility. But other issues regarding the rationale for stakeholder participation go beyond this rather utilitarian and pragmatic perspective, and such issues are also crucial for evaluation credibility. Many evaluation theorists remind evaluators about the centrality of stakeholder participation. For example, Monnier talks about the importance of the 'social utility' of evaluation, noting that social utility can only be attained with the participation of all stakeholders involved in the programme or public policy to be evaluated. The evaluation, he says, should not only have institutional, technical or scientific legitimization, it should also have, above all, political legitimization,

21 The norm 10.1 of the "Norms for Evaluation in the UN System", published by UNEG in 2005, reads: "Transparency and consultation with the major stakeholders are essential features in all stages of the evaluation process. This improves the credibility and quality of the evaluation. It can facilitate consensus building and ownership of the findings, conclusions and recommendations". See uneval.org/documentdownload?doc_id=21&file_id=562.

which is given by stakeholder participation.

However, going through the motions of seeking participation without a genuine intent to maximize participation could lead to favouring those that hold the levers of power. The deliberative democratic evaluation model (House and Howe 2000), for example, proposes procedures that ensure that the disadvantaged are respectfully included in the evaluation process. For House, the deliberative democratic evaluation “aspires to arrive at unbiased conclusions by considering all relevant interests, values and perspectives; by engaging in extended dialogue with major stakeholders; and by promoting extensive deliberation about the study’s conclusions—in addition to employing traditional evaluation methodologies” (House 2005). This recognition of structural and systemic inequalities, which produce, by default, biases or preferences for the favoured or powerful, is also a key reason for the need for an equity-based, gender-sensitive and human rights perspective in evaluation. All these approaches remind us that there is a need to make deliberative efforts to level the playing field by allowing the weakest to have a voice in the evaluation process.

Stakeholder participation is neither simple nor straightforward. New forms of policy-making and new instruments are necessary to deal with new and complex public problems and policy challenges—locally, nationally, regionally and globally. Such complexities increase the scope for democratic participation and the number of stakeholder interactions. Governments have moved from a centralized and hierarchical, top-down form of government to a form of governance that involves multiple decentralized and contested types of public and private actors. To be credible in this changing environment, evaluation practice and theory cannot ignore new realities and offer simple evaluative designs for complex situations and interventions.

Another important perspective regarding credibility and stakeholder participation is the citizenship’s perspective. In democracy, as Tarja Cronberg, European Parliamentarian and host of the European Evaluation Society ‘Evaluation in Democracy at the European Parliament’ event, says, evaluation can be a tool for empowering citizens. It can be also a way of promoting social learning, identifying priority policy interventions and reducing the democratic deficit (in this case, in the European Union). At the same time, citizens must be able to trust institutions for evaluation to play a role. But if an evaluation is planned, implemented and used properly from a citizen’s perspective, it should also help to recover and build institutional trust.

CREDIBILITY NOT ONLY DEPENDS ON THE QUALITY AND INDEPENDENCE OF THE EVALUATORS, BUT ALSO ON THE INSTITUTIONS AND THE SYSTEMS WHERE THE EVALUATIONS ARE CONCEIVED, PLANNED AND MANAGED

Evaluators do not play a lone role in evaluation. They are hired by commissioners, who establish terms of reference, and clients, who normally have a say on how the evaluation is conceived and performed. Commissioners and clients do not operate alone either. They play a role in organizations, which have a purpose, a way of working and, sometimes, strategic aims in their evaluation work. Evaluations may be conducted using established evaluation systems or procedures, for example, for ways of developing terms of reference or finding and hiring evaluators. Thus, credibility not only depends on the quality and independence of the

evaluators, but also on the institutions and systems where these evaluations are conceived, planned and managed.

However, the evaluation community has a tendency to think—sometimes exclusively—from an evaluator’s point of view. A typical example is the set of standards, principles and guidelines for ethical conduct, which some evaluation societies have adopted in the last two decades. Most, including the Joint Committee on Standards for Educational Evaluation, which established the first evaluation code, seem only to think from the evaluator’s point of view. They recommend what evaluators should do to conduct a good, useful, correct and ethical evaluation. Even the UN standards—the Norms for Evaluation in the United Nations System, which were established by the United Nations Evaluation Group (UNEG) in 2005 and have a UN-system perspective in most sections (definition, responsibility of evaluation, policy, intentionality, impartiality, independence, evaluability, quality, competencies of the evaluation, transparency and consultation, evaluation ethics, follow-up and contribution to knowledge building)—entrust the two norms in the section on quality of evaluation and the five in evaluation ethics to the evaluator. The UK Evaluation Society’s Guidelines for Good Practice in Evaluation²² are exceptional in this regard, as in addition to being for evaluators, they are also for commissioners, self-evaluation and evaluation participants.

In sum, we have thought more about what evaluators can do to enhance evaluation credibility than about how institutions and commissioners should promote credibility. What should institutions or governments do? I would suggest that they have a clear policy on evaluation, its purpose and who should benefit. For example, an enabling environment for stakeholder participation is more frequently facilitated—or obstructed—by organizations, clients and commissioners than by evaluators.

The instruments and mechanisms for implementing a clear evaluation policy, which ultimately enhances credibility, may vary greatly and will depend on the characteristics of the political system and the general context and culture of each country, among other things. For some countries in Europe, the creation of a specific institution for evaluating public policies for all sectors might work (Spain). Elsewhere, the goal can be better pursued through advances in each policy sector (France). Some might find that passing concrete legislation that requires evaluation is particularly useful, while others realize that this mechanism might have unintended effects, such as excessive bureaucratization or simple lack of implementation (Italy). Some evaluation systems reside in the legislative branch (Switzerland), while the executive is responsible in other countries (Sweden). Each political and administrative system, and each political situation, may require a different solution for promoting credibility.

Spain offers an illustrative example. The *Agencia Estatal de Evaluación de Políticas Públicas y Calidad de los Servicios (AEVAL)* was established in 2006, after a very well-considered project in which different expert commissions participated. In the early part of the project in 2004, experts called for an agency that was dependent on the legislative branch to have better conditions for independence, and hence, credibility. That call was made with an understanding of the difficulties that its implementation would have in Spain’s non-presidential political system, which is

22 See evaluation.org.uk/assets/UKES%20Guidelines%20for%20Good%20Practice%20January%202013.pdf.

formed of closed electoral lists and demands for strong party discipline in the parliamentary system. Finally, because of the formal difficulties that call would require, the recommendation was to start the agency at the executive level and eventually change it to depend on the senate when a foreseen constitutional reform of the senate would come.

What we did not know at the time is that Spain was about to enter a period of major economic crisis, which was preceded by a political and institutional crisis that started with a tremendous—even hysterical—confrontation between the two main parties, reflected especially in parliament. Having AEVAL depend on the legislative branch in this context would have had unintended consequences on credibility from administrators and the citizenship. From my perspective, AEVAL had a greater chance of enhancing its credibility and independence from the executive than if it had been in the legislative.

Unfortunately, AEVAL was born at the onset of a deep economic crisis. Dramatic budget cuts and its gradual reduction as a political priority meant that AEVAL did not develop as planned. Regarding its credibility, AEVAL started to serve the public administration at the beginning of a major political crisis, which led to a substantial and ever deeper political disaffection by the Spanish public, accompanied by wide political and institutional distrust (the general dissatisfaction at the functioning of democracy has continued to increase, from 42 percent in 2007 to 62 percent in 2011).²³ In my opinion, having reached this point, the only way that institutional and political trust will be recovered in Spain will be by giving real voice to the citizenship. And evaluation could play an important role in that recovery.

EVALUATION CAPACITY BUILDING FOR BOTH EVALUATORS AND COMMISSIONERS

It is clear that evaluation credibility depends to a large extent on the expertise of the evaluators—and on that of the commissioners. That is why specific education on evaluation, good training courses and internships are so important for sound evaluation capacity building.

However, there are few specific graduate programmes in evaluation, whether in Europe²⁴, in North America or in other regions of the world. After many years dedicated to training evaluators, I am convinced that taking one course in evaluation in a general research methods, management, public policy or development Masters or PhD is frequently not enough to consider someone prepared to be good evaluator or commissioner. Neither are 20- or 40-hour professional training courses. A very important part of any good evaluation training should be in 'learning by doing', which is why it is crucial to provide internship and practicum experiences to new evaluators. But previous to this 'learning by doing', there is

23 Public trust in institutions dropped enormously between 2007 and 2011. Although we do not yet have the updated data, it has probably continued to drop in 2013, given the latest corruption cases and, in my opinion, the bad management of those cases by the Spanish Government. As examples, here are the average means in trust for 2007 and 2011 for different institutions: The King, from 7.20 to 5.79; The Constitutional Court from 6.08 to 5.00; the Congress of Deputies from 5.54 to 4.87; the Senate from 5.45 to 4.07; the Spanish Government from 5.46 to 3.93; the EU from 6.67 to 5.79; political parties from 4.22 to 3.38; and the unions from 4.58 to 3.26. Data from CIS (*Centro de Investigaciones Sociológicas*) and from prof. Francisco Llera's presentation at the 6th Harvard Summer Seminar on Sociological and Political Research, August 2013.

24 See europeanevaluation.org/images/file/Study_programmes/16_profiles_November2012.pdf.

also a vast body of knowledge and evaluation theory, practices and dispositions that are part of an evaluation curriculum, and that can and should be taught.

In relation to capacity building, it is necessary to mention the efforts the evaluation community has made towards its professionalization. Those efforts include the elaboration of codes of practice. These are sets of norms, guides and standards adopted by evaluation societies mainly during the 1990s and the early 2000s (see Bustelo 2006). These codes started to establish some boundaries of what an evaluator should do as part of a good, useful and ethically correct evaluation (with the exception of the UK Evaluation Society guidelines mentioned above, which added guidelines for commissioners and others).

Recent years have seen those codes develop into a focus on competencies or capabilities, that is, the capabilities evaluators should have and develop. The European Evaluation Society developed a capabilities framework and validated it through a survey to its membership.²⁵ This same stream has recently turned into a discussion about possible designation or accreditation of evaluators. Canada has already implemented an accreditation scheme and the International Development Evaluation Association (IDEAS) has initiated its own review. While the issues involved are controversial and sensitive, the European Evaluation Society believes that the time is right to engage in a public debate about the pros and cons of a potential peer review system geared to professional development and designation. Thanks to EvalPartners, which has provided funding for the European Evaluation Society and the UK Evaluation Society to sponsor a joint workshop designed to debate the potential for a Voluntary Evaluator Peer Review system, a discussion has been launched in the European Evaluation Society membership. The Voluntary Evaluator Peer Review proposal has been designed under the aegis of the European Evaluation Society's Professionalization Thematic Working Group, and the society's board unanimously agreed that the working group's proposal deserves exposure to a broad evaluation audience as the first step in a proactive consultative process.

Why am I linking these professionalization efforts to the necessary evaluation capacity building as a way of enhancing credibility? Because these efforts have been, and will continue to be, an invaluable aid for building evaluation capacity by identifying what evaluators should know, the capabilities they should be trained in and, eventually, be tested against for designation or accreditation. Evaluation credibility should also be promoted with a system of 'accredited' evaluators and also commissioners. An evaluation's quality and credibility do not depend only on evaluators, so these capabilities should also be considered for commissioners and evaluation units elsewhere. Credibility also depends on political systems and cultures, institutions and contexts. Any framework or system that helps to define what a credible evaluation should be like must never be imposed, should be the object of periodic review and negotiation in the community, and should generally be flexible enough to be useful in different contexts.

From my perspective, an indicator of a degree of consolidation in the evaluation function is the fact that evaluation commissioners and managers recognize that they also need to learn evaluation and look for some sort of evaluation training. One of the best features of

25 See europeanevaluation.org/files/EES_Capabilities_Survey_Evaluation.pdf.

an evaluation course is to be seated together, exchanging and learning at least these two different roles and perspectives—evaluators and evaluation managers and commissioners—from each other. However, more thought has to be given to the possibility of training some different capabilities and skills for different roles in evaluation. That is one of the reasons why I think we should look at the evaluation capacity building picture from different angles and perspectives. As an illustrative example, I have seen commissioners so enthusiastic in their evaluation work that they have misinterpreted their role and have told the evaluation team exactly how they should conduct the evaluation. But the job of a good commissioner is not about elaborating the terms of reference (as I have unfortunately seen more than once), where too much time is spent on explaining the concrete methodology and methods to be used. We all know evaluation teams need a good explicit framework and context, but they also require enough freedom to be able to do a good job.

CREDIBILITY AND METHODOLOGY: WHAT COUNTS AS CREDIBLE EVIDENCE AND FOR WHOM?

Evaluation credibility is related to the data that should be gathered empirically to answer evaluation questions. This has a first level, which is related to the scope of information to be gathered and the reliability of the information acquired by an evaluation. The quantity and reliability of information, along with lack of bias, is what we get through third persons or secondary sources. Who have we got information from? Have we gathered information from the whole spectrum of stakeholders, or just from some of them? There are also the issues of honesty and impartiality, because credibility requires that evaluations report both successes and failures.

Credibility is inevitably related to the methodological perspective or how empirical evidence is gathered and analysed. It is related to methods (for example, questionnaires and interviews) and methodology (for example, case studies, surveys and experiments). But it is also related to the more philosophical question of social inquiry, about the nature of reality, about what constitutes knowledge and how it is created; that is, about epistemology and ontology. Moreover, what is considered credible evidence is clearly mediated by key philosophy of science notions, such as the concept of paradigm. The debate on what it is considered credible evidence comes from an old and recurring discussion on how best to study social phenomena: the quantitative-qualitative debate.

In the evaluation field, the question about what constitutes credible evidence used to support claims relating to the impact of a practice, programme or policy, have fiercely reappeared when some international organizations, networks and federal departments in the USA have identified the randomized controlled trial as the 'gold standard' design for generating 'scientific' evidence of evaluated programmes or policies. This created much discomfit in the evaluation community during the last decade and generated responses from several evaluation societies, including from American Evaluation Association in 2003 and the European Evaluation Society in 2007.²⁶ In 2006, Claremont University organized a symposium

26 See the European Evaluation Society statement on methodological diversity at europeanevaluation.org/library.htm.

on ‘What Counts as Credible Evidence in Applied Research and Evaluation Practice?’, in which known evaluation academics, both in the experimental-quantitative and the non-positivistic-constructivist and qualitative approaches were invited to participate. That symposium led to the production of an edited volume with the same title as the symposium, which was organized around the social inquiry paradigms as a frame for the debate on credible evidence (Donaldson, Christie and Mark 2009).

Although one could argue that the very differentiation between experimental and non-experimental approaches is somehow normative and tends towards the condition defined—defining ‘the other’ by the absence of that condition—this was the first time that the credibility and the concepts of evidence and impact were debated openly and in depth from a methodologically plural perspective, and not exclusively related to a concrete epistemological stance, such as experimentalism.

From my perspective, there can be no other way, because the discipline of evaluation has evolved from joint perspectives and multiple methods and approaches, allowing debate among evaluators who come from very different traditions of study. Due to its practical and applied nature, and the need for credible evidence from different perspectives to answer varied questions in different contexts, evaluation has been one of the first fields in which quantitative and qualitative researchers and evaluators have exchanged views, networks and talked to each other. Moreover, the evaluation community was quickly ready to embrace the mixed-methods approach. In this “era of paradigm pluralism” (Greene 2013, 111) and necessary understanding among different perspectives about social inquiry, this mixed-methods approach is broadly accepted in the evaluation community. As Donna Mertens and Sharlene Hesse-Biber say in the editors’ note in their recent volume, *Mixed Methods and Credibility of Evidence in Evaluation*, in *New Directions for Evaluation*:

*An old Italian proverb reads, “What’s old is new, what’s new is old” (Melfi, 2011). This quote characterizes the story of mixed methods in the evaluation community in that mixed methods have been used by evaluators for many years. Many evaluators intuitively came to the conclusion that evaluations on complex social programs could be enhanced by the use of multiple methods; hence the combination of both quantitative and qualitative data in the same study is nothing new. Attention to mixed methods in evaluation was apparent in the *New Directions for Evaluation* edited by Jennifer Greene and Valerie Caracelli in 1997 (Greene & Caracelli, 1997). Since that time, attention to mixed methods has increased exponentially, as evidenced by the launch of the *Journal of Mixed Methods Research* in 2007, which had an initial impact factor of 2.219 and ranked fifth out of 83 journals in the social sciences, interdisciplinary category, according to the 2010 *Journal Citation Reports* by Thomson Reuters (2011). The American Evaluation Association Topical Interest Group: *Mixed Methods in Evaluation* was founded in 2010 and quickly became one of the largest of [The American Evaluation Association’s Topical Interest Group]. And, the *Sage Handbook of Mixed Methods in Social and Behavioral Research* (Tashakkori & Teddlie, 2010) is in its second edition.*

MERTENS & HESSE-BIBER, 2013: 1

This important piece of work aims to ask “about the connection between the call for evidence-based programmes and the potential contribution of mixed methods to the creation of credible evidence. The purpose of this issue is to examine the contributions of mixed methods evaluation and its emerging philosophies, theories, and practices (...) as well as opening up the possibility of enhancing credibility with evaluations that start from several

paradigmatic stances, such as post-positivism, pragmatism, constructivism and transformativism” (Mertens and Hesse-Biber 2013, 3).

The mixed-methods approach claims that this mixed usage should not only function at the methods level, but also at the methodology and epistemology level. As Mertens and Hesse-Biber say, “it is important to understand that mixed methods is not just about (mixing and combining) methods. The use of any given method or set of methods in an evaluation is also tightly linked to specific epistemologies, methodologies (theoretical perspectives), and axiological assumptions, as well as being connected to particular stakeholder perspectives” (Mertens and Hesse-Biber 2013, 5-6).

This volume puts forward the need to be conscious and explicit about such theoretical perspectives and assumptions. Jennifer Greene argues that mixed-methods evaluators should be “explicit about the paradigmatic assumptions that frame and guide their work,” and that “careful explication of just what is being mixed in a mixed-methods study contributes to the subsequent warrant for and thus credibility of results.” For her, “it is a critical responsibility of the inquirer” to make explicit assumptions, such as the “nature of the social world, what counts as warranted knowledge, defensible methodology, and the role of social inquiry in society.” Inquirers should also “justify the values they invoke—values of distance, engagement, inclusion, objectivity, generalizability, contextuality, social action” and so forth. This is particularly important in evaluation contexts, because they are saturated with value (Greene 2013, 111-112).

In sum, credibility, evidence and impact are not concepts exclusively valid for positivism stances, so they should be explored and defined by other paradigmatic perspectives. Positivism has been the dominant paradigm for many years, but is not necessarily the case anymore—as the methodological and paradigmatic pluralism in the evaluation community has demonstrated. Mixed-methods evaluators propose advancing the debate of credible evidence by making explicit values as well as ontological, epistemological and methodological choices, so that paradigmatic and methodological transparency is needed for credibility. In my opinion, we should insist that this transparency is exercised not only by mixed-methods evaluators, but by all evaluators. For the sake of credibility, I believe that ‘classical’ and dominant understandings, such as experimentalism, should not be taken for granted, and the paradigmatic and methodological choices that drive those perspectives should be explained and made explicit. This would be a real acknowledgement that there are other modes of inquiry that are not hierarchically inferior. In this way, methodological pluralism would become real—no longer would one perspective be the ‘norm’, while other ‘alternative’ choices have to be justified.

COMMUNICATION AND REPORTING: ANOTHER KEY ISSUE FOR EVALUATION CREDIBILITY

As a final important point in this discussion on evaluation credibility, I would like to mention communication and reporting. Although the lack of time and space in this guiding paper does not allow me to develop this point further, I would like to mention at least three aspects or questions for discussion in the online community of practice. These are the issue of transparency and what should be made explicit in the report; a need for a fluid communication

with stakeholders during the process; and the accessibility of evaluation reports.

If an evaluation is to be credible, it must be transparent and state the political and institutional contexts in which it took place. It should also consider—and include in the evaluation report—paradigmatic, methodological and values-related issues.

Fluid communication with commissioners and stakeholders during the evaluation process is known to promote better use of evaluation results (see, for example, Torres, Preskill and Piontek 1996). But fluid communication is also important in enhancing the credibility of the evaluation process and the evidence resulting from it. If commissioners and stakeholders have regular and transparent feedback during the evaluation process, this results in a higher probability of perceived credibility by them.

A key issue for transparency and credibility is related to the accessibility of evaluation reports. The evaluation process may have been well-designed, the evaluation may have been conducted honestly and transparently, but if the final report is not accessible to the public, it will be almost impossible to be perceived as credible by those without access. The ease of access to information through the Internet has vastly improved openness, but it also challenges the cases in which there is no immediate access to evaluation reports.

REFERENCES

- Bustelo, Maria. 2006. "The Potential Role of Standards and Guidelines in the Development of an Evaluation Culture in Spain". *Evaluation*, 12(4): 436-452.
- Chelimsky, Eleanor. 1998. "The Role of Experience in Formulating Theories of Evaluation Practices". *American Journal of Evaluation*, 19(1): 35-55.
- Donaldson, Stewart I., Christina A. Christie, and Melvin M. Mark. 2009. *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Thousand Oaks. Sage Publications.
- Greene, Jennifer. 2013. "Reflections and Ruminations". In D. M. Mertens and S. Hesse-Biber (Eds.), *Mixed Methods and Credibility of Evidence in Evaluation. New Directions for Evaluation*, 138: 109-119.
- House, Ernest R. 2005. "Deliberative Democratic Evaluation". In *Encyclopedia of Evaluation*, edited by Sandra Mathison, 105-9. Thousand Oaks, CA: Sage Publications.
- House, Ernest, and Kenneth R. Howe. 2000. "Deliberative Democratic Evaluation in Practice." In George F. Madaus, Michael S. Scriven, Daniel L. Stufflebeam (eds.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer-Nijhoff.
- Mertens, Donna M., and Sharlene Hesse-Biber. 2013. "Mixed Methods and Credibility of Evidence in Evaluation". In D. M. Mertens & S. Hesse-Biber (Eds.), *Mixed methods and Credibility of Evidence in Evaluation. New Directions for Evaluation*, 138: 5-13.
- Monnier, E. 1995. *Evaluación de la acción de los poderes públicos*. Madrid, Instituto de Estudios Fiscales. Ministerio de Economía y Hacienda.
- Torres, R., H. Preskill and M. Piontek. 1996. *Evaluation Strategies for Communicating and Reporting*. Thousand Oaks: Sage.

3. REFLECTIONS ON INDEPENDENCE IN EVALUATION

HANS-MARTIN BOEHMER, Senior Manager
XIMENA FERNANDEZ ORDONEZ, Evaluation Officer
NEHA SHARMA, Consultant

Independent Evaluation Group
The World Bank Group

This paper integrates a synthesis and analysis of the discussions held on the online platform leading up to and during the Third International Conference on National Evaluation Capacities.²⁷ It is also based on an initial paper that was produced to serve as background for the presentation on independence and related discussions during. The background document and the discussions focused on the relationship between independence and evaluation, seeking to contribute to linking how credibility and use interact with independence. The background paper was based on an analysis of different examples of how organizations and countries deal with the issue of independence in their M&E systems.²⁸

INTRODUCTION

Countries, development organizations and scholars often struggle to determine what independence in development evaluation really implies, what the optimal level of independence for an evaluation function should be and how to achieve that level. In these conversations, the independence of an evaluation function is often discussed in isolation from the ultimate role of evaluation. Also, independence is often portrayed in absolute terms. That is, evaluation functions are either independent, or they are not.

27 The paper does not reflect the opinions of the Independent Evaluation Group of the World Bank. It also does not necessarily reflect the views of its authors, since it attempts to summarize and analyse the conversations held during the Third International Conference on National Evaluation Capacities.

28 Analysis of Country M&E System cases was based on IEG Evaluation Capacity Development Working Paper Series; see ieg.worldbankgroup.org/evaluations/evaluation-capacity-development-eed.

The conversations before and during the conference highlighted the view that in the development realm, independence in evaluation is only relevant for its help in supporting evaluation's ultimate function: improving development results. Independence is not an end in itself. Therefore, different levels and types of independence might be appropriate for different situations. There is, unfortunately, no recipe or blueprint for achieving 'optimal independence'.

This paper also posits that independence should be seen as a characteristic that helps reduce the biases that an evaluation function might have (or be perceived as having). Reducing such biases should increase an evaluation's credibility, which in turn should increase the use of its evidence to feed decisions. The assumption is that 'evidence-based decision-making' will ultimately improve development results.

Independence is relevant to evaluation because assessing the results of development projects, programmes and policies is complex, and many biases can emerge in the evaluation process. However, the relevance of each of these biases—and therefore the need to deal with it in a particular situation—depends on whether or not it affects relevant actors' perceptions regarding the credibility of the evaluation function and the evidence it produces. Evaluation functions face many biases—some real and some perceived. Independence cannot preclude some of them; they just need to be managed. And even though independence is an important component that influences credibility, merely establishing independence may not be enough. Similarly, credibility is not the sole determinant that leads to the use of evidence. In fact, independence in evaluation is often seen, not necessarily correctly, as endangering the adoption of the evidence produced by evaluations.

The experiences of many countries' and development organizations' efforts to influence development results through evaluation shows that a battery of responses, often articulated through institutional M&E systems, is helpful in aligning independence with other important characteristics that support evidence-based decisions. The Independent Evaluation Group of the World Bank regards independence as one of the keys to ensuring evaluation credibility and the use of the evidence it produces. Other organizations and countries that have different realities have chosen different paths; M&E systems can help find an 'optimal' or 'useful' type and level of independence for a particular situation.

The first section of this paper explores why independence is important in evaluation. The second section examines the definition of independence and its relevant dimensions. The third section focuses on the current debates about independence, including the trade-offs, overlaps and supportive functions between independence and commonly implemented solutions to uphold it.

WHY IS INDEPENDENCE IMPORTANT IN EVALUATION? THE LINK BETWEEN INDEPENDENCE AND RESULTS

“Independence is not an end in itself (but it is meant to ensure) “objective, impartial, uncompromised and unbiased findings that will make the evaluations credible and valuable.”

ROBERT PICCIOTTO (2013)

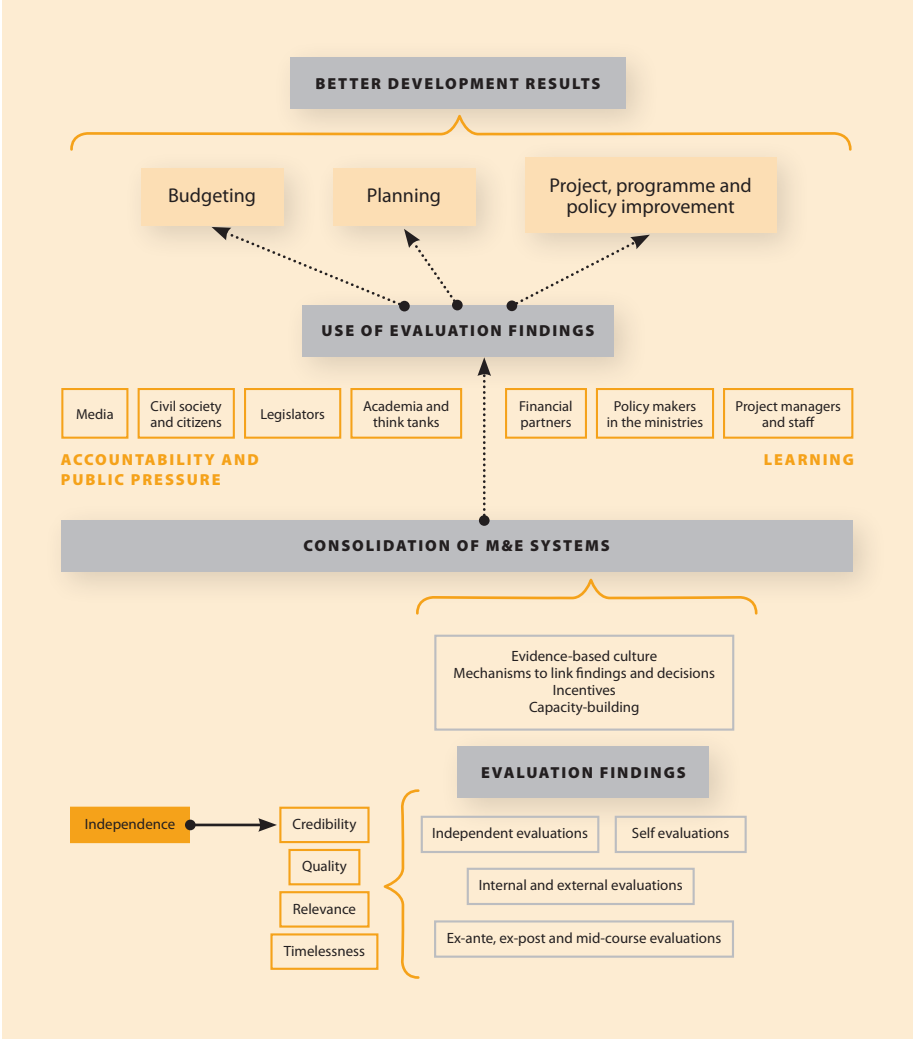
“The idea of independence has a long association with evaluation (and in many cases is seen as) an essential element of good evaluation practice” (INTEVAL 2011, 1). Independence becomes particularly relevant in development evaluation because of the complex nature of development policies, programmes and projects; the poor availability of data; the multiplicity of actors; and the perceptions that particular interests might have in confirming positive results rather than truly assessing what happened and why. This evaluative context can differ among areas; for example, in medicine, data is more widely available and there is greater opportunity to conduct well-controlled experiments than elsewhere.

Discussions of independence in evaluation often leave out the ultimate role that evaluation itself is meant to play. Evaluation findings can help countries and agencies know what progress has been made (if any), why it did or did not happen, the mechanisms that aided it and, in some cases, what portion of those results can be attributed to specific project, programme or policy feature (Thomas and Tominaga 2010, 376). Evaluations can supply information for feedback loops and nudge real-time improvements by incorporating past learning and lessons into later planning. Evaluation can help improve the design and implementation of ongoing projects, programmes and policies; promote the scale-up and replication of successful practices; avoid mistakes; and incorporate learning into higher-level planning, budgeting (or prioritization) and design of programmes. Ideally, evaluations should enable stakeholders to ask tough questions, challenge assumptions, allow for changes in courses of action, learn progressively, make space for reforms and inform policy debates (Boehmer, in Rist 2011).

What purpose does independence serve in development evaluation and how does it operate? If we were to map a theoretical ‘results chain’ for independent evaluations, we would see that independence is meant to prevent bias (promote the objectivity) of an evaluation function, which in turn strengthens its credibility (see Figure 1). This increased credibility should result in a rise in the utilization of the evidence. Better-informed decisions or evidence-based decision-making should then improve development results. In some cases, the evidence influences decision makers directly. In others, it mobilizes accountability mechanisms that influence them indirectly.

Depending on the actors that the evaluation function is trying to influence, not all biases will hold the same importance. The biases that will matter will be those that affect the credibility of the evidence in the eyes of the actor who needs to act on it. Independence, however, is not the only factor that affects credibility, and credibility might not be enough to ensure the actor incorporates an evaluation’s findings into their decision-making process. Lastly, though evidence-based decision-making could improve development results, it is likely not the only determinant.

FIGURE 1: A THEORETICAL RESULTS CHAIN FOR INDEPENDENT EVALUATION



In his 2013 study of evaluation independence, Picciotto found that “without independence, evaluation is perceived by the public to be subservient to vested interests.” When evaluation results appear to be tied to personal gain, profits or corporate interests, the ability to achieve desired results becomes seriously compromised. Conversely, establishing independence can validate results, which may significantly improve the ability to achieve desired results (Picciotto 2013, 22).

Because of this connection, the discussion on independence in evaluation cannot be separated from the discussion on avoiding bias. Independence can decrease conflicts of interest, where the evaluator is “in bed with’ the program being evaluated [which is] typical of much program monitoring by agencies and foundations where the monitor is usually the godfather of the program, sometimes its inventor, and nearly always its advocate at the agency” (Scriven 1991, 192-193). Independence—or true freedom from political, financial or personal objectives—is supposed to render unbiased findings. Both real and perceived biases are relevant, and so both perceived and real independence is, too.

The discussion on independence also needs to be linked to the actors that the evaluation is trying to influence and their perceptions regarding the biases that affect credibility. By increasing the credibility of evidence, key stakeholders (e.g. parliaments, opposition parties, civil society organizations, citizens, universities, think tanks and the media) have better tools available and might be more likely to hold governments and agencies accountable and exert public pressure for course corrections (Mackay 2007). Further, managers and planners can see more clearly their mistakes and missed opportunities and pre-emptively incorporate such information into their decisions.

Biases and other constraints can emerge at every stage of an evaluation, including when the subject is chosen, the questions determined, the methodology designed, the engagement and consultation strategy with stakeholders set, the information analysed, the evaluator chosen, the evaluation terms of reference set, the evaluation implemented, the recommendations constructed or the results are reported and presented. Factors such as who is paying for the evaluation, how the evaluation is managed and who the evaluators are, are also critical regarding the emergence of biases.

Independence can abolish or reduce biases and increase objectivity. However, independence—or at least the ways in which it is often understood—also has the potential to diminish the expected effects of evaluation on results.

Independence is not the only way to improve credibility. For example, making the data used for evaluation available (i.e. giving individuals, academia, think tanks and others the chance to analyse it) is one way to increase the legitimacy of the evidence, or at least diminish questions regarding its objectivity. When external third parties have the opportunity to use their own methods and draw their own conclusions and recommendations, the actual evaluation findings can be validated, refuted or understood as one interpretation.

Moreover, independence is a necessary but not a sufficient condition for evidence to affect results. Evaluation findings can only play a predominant role in promoting effective and efficient projects, programmes and policies if they are incorporated into the relevant decisions around them. Contrary to the assumption that credible evidence will influence

decisions, many aspects beyond credibility affect use. Key factors in determining the role that evaluations play in achieving better results include the relevance of the topics evaluated, the evaluation's guiding questions, the timeliness and quality of the evaluation, the dissemination of the findings, the existing structures link findings and decision-making, the incentives in place, the level of capacities to engage with evidence and the culture of an organization.

In conclusion, though independence could help reduce many biases that affect credibility, other biases that are not possible to avoid through independence will need to be managed; independence is only one of the ways to increase the chances of the evidence from evaluations influencing results. A country or an organization first needs to understand the perceived or real biases evaluations face and then choose the type and level of independence that appropriately responds to a given situation.

WHAT IS INDEPENDENCE? FORMAL DEFINITIONS AND DIMENSIONS

Independence: Freedom from external control or influence.

NEW OXFORD AMERICAN DICTIONARY

Though there is a lot of discussion on independence in evaluation, there is rarely agreement on what independence really is. The definition of the United Nations Evaluation Group (UNEG) regards an evaluation function as independent if it is "objective, free from undue influence, and has full authority to submit reports directly to appropriate levels of decision-making" (UNEG 2005). Implied in this definition is the directive that "management must not impose restrictions on the scope, content, comments and/or recommendations of evaluation reports." To prevent conflicts of interest, UNEG also asserts that evaluators "must not be directly involved in policy-setting, design, implementation or management of the subject of the evaluation" and adds that "evaluators should have full access to all relevant information required for the evaluation". Somewhat less straightforward, but more extensive, is the definition of the Organisation for Economic Co-operation and Development's Development Assistance Committee's (OECD-DAC), which considers an evaluation independent if it is "carried out by entities and persons free [from] control of those responsible for the design and implementation of the development intervention" (OECD-DAC 2002). Similar to UNEG, the OECD-DAC definition includes 'full access to information' and 'full autonomy' as necessary characteristics of an evaluation's design, implementation and reporting for it to be considered 'independent'.

Both structural and individual independence (also called functional or intellectual) are relevant in evaluation (Heider 2014 and Picciotto 2013). Even though this paper focuses on the institutional and organizational dimensions of the relationship between independence and evaluation, it should be noted that for an evaluation to be considered truly independent, these requirements must apply to the evaluation function or unit and to the actual individual or team conducting an evaluation.

Individual evaluators also suffer from numerous natural biases and risks that can corrupt findings, such as anchoring, framing and attention bias. Anchoring refers to focusing on the one aspect of a programme that confirms preconceived notions about the project; framing refers to drawing different conclusions from the same information that others would have available, depending on how or by whom that information is presented; and attention bias refers to paying attention to emotional stimuli instead of facts. Ignorance, ideology and inertia can also be important biases affecting both organizations and individuals (Boehmer, in Rist 2011). An evaluator needs ‘independence of mind’, the state of mind that permits one to provide an opinion without being affected by influences that compromise professional judgment, thereby allowing an individual to act with integrity and exercise objectivity and professional scepticism (International Federation of Accountants 2003).

The question that arises is how countries and organizations should use these definitions and dimensions when structuring evaluation functions. Independence is still often seen in absolute terms: evaluation functions are seen as either independent or not. In reality, complete independence, or the avoidance of all biases, is hard to achieve. So should countries and organizations aspire to ‘tick’ all these boxes? Current discussions around this subject have taken quite a nuanced approach, increasingly linking independence to organizational learning theories and specific contexts (e.g. relevant decision makers, perceived and real biases) (Picciotto 2013). Countries and development organizations have also taken very different routes to ensuring the independence, credibility and use of evidence.

Despite the availability of good definitions, independence is still often oversimplified and equated or opposed to other terms and constructs. The next section of this paper explores some of the common views regarding evaluation independence, how to reduce biases and the likely effects these can have on credibility, use and results. Some of these ‘solutions’ might actually go against the use of evidence.

THE CASE OF THE INDEPENDENT EVALUATION GROUP: INDEPENDENCE WITHIN

In the World Bank Group, structural (organizational) independence (actual and perceived) is seen as one of its two pillars for the credibility of its evaluation function and its influence on the World Bank Group’s results (the other pillar is the ‘quality and relevance’ of its work). Independence contributes to the overall governance of the World Bank Group by allowing stakeholders to use unbiased and reliable findings. According to its 2011 self-evaluation, the Independent Evaluation Group’s (IEG) organizational systems, reporting structures and procedures are consistent with those established by the Evaluation Cooperation Group Good Practice Standards for independence.

But IEG is not truly separated from the World Bank; it enjoys a kind of ‘independence within’. It is an independent evaluation function that remains inside the World Bank architecture. This independent evaluation function strongly complements self-evaluation at the World Bank. In this position, IEG—much like other evaluation departments—faces internal pressures to function as an internal self-evaluation group instead of an independent group. In this situation, issues such as engagement become critical. However, in 2004, an external review noted that engagement with management did not undermine independence; “on the contrary, such interaction should be increased to ensure the usefulness of evaluation products.” (Continued on p. 60)

The Case of the Independent Evaluation Group: Independence Within (continued from p. 59)

IEG INDEPENDENCE CRITERIA	
CRITERION	FACTORS THAT HELP MEET CRITERION
Organizational Independence	IEG reports directly to the World Bank Board through the Committee on Development Effectiveness (CODE) and is thus organizationally independent from management and operational staff whose activities are being evaluated. IEG's scope of responsibility extends, without restriction, to all the determinants of the World Bank Group's operational results.
Behavioural Independence	IEG's Work Program and Budget are endorsed by CODE and approved by the Board. The IEG budget is separate from management budgets, and in the end, management does not have authority over IEG's budget or its use. IEG's reports are transmitted to the Board through the Director General for Evaluation, without any clearance from management. Although management is given the opportunity to review and comment on draft reports, IEG decides how and whether or not to address such comments.
Avoidance of Conflict of Interest	IEG staff does not evaluate activities that they were previously responsible for or were involved in. The head of IEG and its Director Generals are not eligible for employment in other positions in the World Bank Group or for consulting assignments.
Protection from Outside Influence	IEG's three-year rolling consolidated work programme and budget are prepared independently of management for endorsement by CODE and approval by the Board.

The IEG has also attempted to increase the influence of its evidence on World Bank decisions—and ultimately its effect on development results—through the Management Action Record (MAR), a system that mandates and facilitates incorporating emerging recommendations into the Bank's policies and programmes, and in a sense, mediates the relationship between IEG and management. MAR allows IEG to track the progress made by World Bank Group management in adopting its recommendations from sector, thematic and corporate evaluations.

MAR has helped both IEG and management by: improving the quality of IEG recommendations (by providing clearer links to findings, prioritizing, improving clarity and specificity, and integrating considerations of feasibility and cost-effectiveness); strengthening engagement and building understanding and agreement with World Bank Group management while drafting recommendations; increasing the number of specific actions to implement IEG recommendations and clarifying timelines and monitoring arrangements; enhancing assessment of progress on implementation; and reducing inconsistencies between IEG and management ratings.

IEG and World Bank Group management jointly began to reform MAR in 2011. After intensive coordination and testing with the three World Bank Group institutions, the new standardized MAR system for tracking recommendations across institutions was rolled out in April 2013. Currently, MAR provides stakeholders with a central repository of findings, recommendations, management responses, detailed action plans and implementations. IEG will continue to provide access and training to all interested members of the Board and World Bank Group management and will also develop a series of briefings on the available tools. As part of IEG's commitment to enhance transparency and access for external stakeholders, IEG has made MAR data available on its external website. The MAR system currently houses 193 individual IEG recommendations across the World Bank Group (World Bank, International Finance Corporation and Multilateral Investment Guarantee Agency). In 2013, IEG followed up on 98 recommendations active in the system.

These same criteria are shared by the Evaluation Cooperation Group, which was founded in 1995 by the heads of evaluation departments of multilateral development banks. The members of the Evaluation Cooperation Group include the evaluation departments of: African Development Bank, Asian Development Bank, European Bank for Reconstruction and Development, European Investment Bank, Inter-American Development Bank Group, International Monetary Fund and World Bank Group.

WHO ARE THE FRIENDS AND FOES OF INDEPENDENCE? PERCEPTIONS, REALITIES AND MYTHS

"Independence combined with disengagement increases information asymmetry, ruptures contacts with decision makers and restricts access to relevant sources of information ... thus, the basic challenge of evaluation governance design consists in sustaining full independence without incurring isolation."

ROBERT PICCIOTTO (2008)

There are numerous preconceptions regarding the relationship between independence and other principles, concepts or constructs. 'Engagement' is often seen as opposed to independence, and 'separation' and 'distance' are often portrayed as its ally. Independence is also often seen as a control function, linked solely to accountability. And to preserve independence, *ex post*, externally conducted evaluations are often the chosen tool. Independent evaluations are readily used as an indication of accuracy and of good quality.

These constructs are linked to attempts to reduce biases and increase organizational and behavioural independence. But sometimes, actions taken in the name of independence can actually have a detrimental effect on the objectives that independence is pursuing. Also, some of these concepts and constructs contain some truth, some do not, and the complexities of reality often preclude such binary determinations. Some of these concepts overlap, some contradict, and others reinforce the notion of 'independence'. This paper does not attempt to resolve such tensions, but rather endeavours to shed some light on how they affect the credibility and use of evidence.

Independence is often understood as 'separation' or lack of engagement

Independence is often interpreted as drawing and maintaining a clear line between the evaluation function or evaluator, the topic of evaluation and those responsible for the design and implementation of the policy, programme or intervention under evaluation. It is often assumed that maintaining separation or 'an arm's length' from decision-making can help prevent bias and provide objectivity to the evaluator. The fear is that intentional or unintentional biases might develop if a relationship develops between the evaluation unit and the managers of the programmes or interventions being evaluated. However, the line between the evaluator and the programme or project is often blurry.

The Evaluation Cooperation Group recognizes that independence does not mean isolation: cross-fertilization of knowledge and experience enriches operational and evaluation activities. Experts reinforce the idea that "evaluation cannot fulfil its potential unless it connects effectively to its management, the supreme authorities that govern the organization and the broader society" (Picciotto 2013).²⁹ Additionally, "selecting the right topics and providing timely inputs require deep understanding of the institution that can be gained through close engagement." Open and transparent conversations among evaluators,

29 "But in doing so it should maintain its objectivity, exercise full freedom of inquiry and resist capture. Evaluation needs to remain functionally and structurally independent" (p. 22).

programme managers, implementers, beneficiaries and civil society regarding all aspects—from programme design through to discussion of programme outcomes and findings—may improve an evaluators' credibility and enhance the evaluation's learning and problem-solving functions.

Establishing an effective balance between independence and engagement could indeed help link evaluations to results. Several cases have shown that although it might take several iterations, it is possible to structure engagements in such a way that they do not violate independence. "Engagement ... provides internal credibility and can lead to constructive change" (World Bank 2013).

Evaluations are also often considered independent when the evaluation function or evaluator sits outside of the institution or is not part of its organizational structure. Being 'far' could offer objectivity, which is why many organizations prefer outside/external evaluators. However, though outsiders can provide a fresh perspective, distance alone does not ensure independence. Simply existing external to an organization does not ensure that there are no political, personal or financial incentives that would invalidate independence. Further, although external evaluators can bring third-party insights, they still face biases. For example, fee dependency of external evaluators can constrain independence as much as proximity; the incentive set-up may not always be conducive to independence (Picciotto 2013). External consulting firms and evaluators with a strong client orientation and business sense often have powerful incentives to earn contract renewals and obtain new contracts, which can corrupt independence by prioritizing customer satisfaction over analytical rigour.

Another concern with external evaluators is that their understanding of the programme or its context may not be sufficient enough to provide useful findings. Additionally, external evaluators may have fewer incentives than those conducting a self-evaluation to go the extra mile to make the evaluation findings useful for learning. Yet another constraint on the use of evidence is the fact that external evaluators do not always have a firm grasp of implementation realities and may establish questions, frameworks or methodologies that are not realistic or relevant.

In the past, independent evaluation was often associated only with accountability and interpreted as an external control

Independent evaluation functions are often considered to be a policing or 'watchdog' function, and are therefore only linked to accountability. This view is reinforced by the fact that evaluations often use language that can hinder communication of the main messages. It is important that independent evaluators present balanced reports, highlighting not only areas that need improvement, but also the constraints and other implementation roadblocks that organizations face. Balanced, realistic reports will increase organizations' receptivity to findings and to evaluations in general.

'Learning' is increasingly being recognized as an integral objective of evaluation. "In evaluation we do not aim to 'learn in the abstract' or test theoretical concepts, but learn about what happened during the process of implementing a policy, strategy, program [sic], or project... [Learning and accountability] are ... two sides of the same coin: accounting for

what happened during implementation when working towards achieving the plans set out in a project design—an agreement and commitment between at least two parties—is the source for learning about the replication of success and avoidance of unnecessary mistakes on that journey to realize expected outcomes” (Heider 2014). Independence does not mean that the work is done when findings are published. Influencing results means that evaluation functions include the responsibility to actually try to foster the adoption of their findings and recommendations.

Independence is also often conflated with an evaluation’s type and timing

Although less discussed, there may be presumptions that additional biases emerge if the evaluation function comes in early in a project or programme’s life cycle. This perception is related to the question of whether it is possible to evaluate work that was influenced by the evaluation function itself. However, as evaluation is an integral part of the policy cycle and tries to influence decision-making (even if only through learning from other evaluations), this problem is present every time findings affect a decision, even indirectly.

Although commissioning evaluations after programme implementation may help reduce some bias and conflicts of interest, it does not guarantee their avoidance. And an evaluation is not necessarily independent simply because it is conducted *ex post*. Similarly, the opposite might not mean complete lack of independence. An evaluation function can conduct independent formative and mid-course evaluations. This would allow for incorporating corrections based on the findings. In other words, it would feed into the feedback loops for programme learning.

If evaluation functions become the history department of a programme, they may not have a significant impact on results: inputs will come too late to allow the organization to change course, and the context may change too fast for lessons to still be relevant. If evaluations do not link appropriately to feedback loops, there is a serious risk that such evaluations may compromise development results and lead to “unjustified replication of popular fads based on anecdotal evidence” or prevent “structured learning ... [or] replication of innovations that may be performing better than is appreciated” (World Bank 2013). Other types of engagements, such as assessments of evaluability practiced by the Inter-American Development Bank, could help develop more accurate programme or policy results frameworks, indicators³⁰ or M&E plans, leading to programme design improvements. Though there are risks and biases related to an earlier engagement that need to be managed carefully, in certain situations and contexts the potential benefits to results may outweigh these risks.

Self-evaluations are often portrayed as the enemy of independence

Self-evaluations and independent evaluations often have complementary functions and can support each other. Conducted by the implementing organization to evaluate its own work, self-evaluations benefit from deep knowledge of the organization, its mission and its

30 Poor quality at entry is a strong predictor of poor eventual outcome, although there might be other factors correlated with that relationship.

programmes. Self-evaluations are more likely to overcome information asymmetries than independent ones (DFID 2008). They are also more likely to be well-received by management and programme implementers. Therefore, self-evaluations' findings have a higher probability of being used for mid-course corrections. Self-evaluations are also important elements of an organization's culture of seeking and using evidence (Mayne 2008 and INTEVAL 2011).

However, there are valid concerns regarding an organization's ability to conduct self-evaluations that are free of biases, given managerial relationships, ownership of or accountability towards projects being evaluated, and the link between results and funding requirements. Moreover, there are moral hazards of judging one's own work that are related to a reluctance to accept failure. It is crucial to complement and balance self-evaluation with checks and balances or an independent evaluation function tasked with attesting to the validity of self-evaluation findings (Picciotto, NEC 2013 community of practice).

There is no direct link between independence of an evaluation and its quality or accuracy

Without independence, the quality of an evaluation will usually be questioned. However, "independence on its own does not guarantee evaluation quality: relevant skills, sound methods, adequate resources and transparency are also required. Independent but incompetent evaluations can be misleading, disruptive and costly" (Picciotto 2013, 20). On the other hand, self-evaluations can be of very high quality. In any evaluation, it is critical to ensure that the evaluators have the appropriate skills and the technical, contextual and procedural knowledge to conduct the evaluation. The evaluator selection process and the quality control mechanisms for evaluations of any type (e.g. steering or technical groups) are key in this regard. Independent evaluations are often also perceived as being more accurate than self-evaluations.³¹ However, other types of biases not related to proximity to the programme may be equally present and skew the evidence they produce.

The proposed responses to decreasing biases or preserving independence can go against the core of what independence tries to achieve. In some cases, the risks of emerging biases can be managed. Under certain circumstances, sacrificing independence for better results may be a reasonable option. At other times, even if the biases are merely perceived, the danger is just too big. Real or perceived, these tensions and trade-offs are usually better managed within an M&E system rather than individually.³²

CONCLUSIONS AND QUESTIONS GOING FORWARD

Independence has traditionally been a very important aspect of development evaluation. Integrally linked to the credibility and use of evidence, the relevance of independence cannot be detached from the ultimate goal of evaluation: to improve development results.

31 As Ken Chomitz, Senior Adviser at IEG explains, just like in econometrics, the "unbiased estimator can be the one with the biggest error."

32 The analysis of the Country M&E System cases was based on the Independent Evaluation Group ECD Working Paper Series, available at ieg.worldbankgroup.org/evaluations/evaluation-capacity-development-ecd.

However, some of the ways in which independence is interpreted or operationalized can have neutral or negative effects on credibility and use, and therefore do not assist with improving results. There has traditionally been a false dichotomy between independence on the one hand and learning and incorporating evaluative evidence for decision-making on the other (Picciotto, NEC 2013 community of practice). But if independence is to have an effect on results, evaluation functions need to have an important learning role and not operate just as a control function.

The discussions and cases analysed as the basis for this article highlighted that, when it comes to independence, one size does not fit all. Case studies and examples can provide inspiration and illustrate different approaches organizations and countries have taken with regards to independence. However, there is no single model or blueprint to attaining independence, nor is there a definitively appropriate response to particular biases or threats to independence. However, some lessons emerge:

Engagement with programmes and managers does not necessarily endanger independence; it can actually help with the credibility and use of evidence. Conversely, separation and distance between management and evaluation can endanger the adoption of evidence. Irrespective of an evaluation's independence, timing and type, its quality should always be a concern because it is directly related to credibility. Often, the most successful evaluation systems combine self-evaluations and independent evaluation functions. Risks and biases need to be managed carefully, but this is true for all aspects of evaluation, not just when dealing with increased engagement or self-evaluation.

M&E systems can play a significant part in better linking and balancing independence, the use of evidence in feedback loops and learning from lessons for future projects, programmes and policies. Independence is a very important tool that countries and development organizations have in fostering credibility and use, but it cannot be the only one. Designing M&E systems often entails making tough decisions, including the type and level of independence. In making those choices, countries and organizations need to ensure that they are not endangering adoption and use in the name of independence. M&E systems tend to respond to the reasons they emerged (decision-making needs, relevant actors, levels of capacity and readiness for evaluation), and they can play a critical role in fostering evaluations that are credible, relevant, timely and of good quality. In order to design an M&E system that is a good fit for its context, it is critical to incorporate diagnostics that facilitate a better understanding of the reality on the ground.

For some systems (e.g. the Independent Evaluation Group, the United States, and the United Kingdom's Department for International Development), independence is a pillar. For others (e.g. Canada), independence is sought after³³ but is not the main feature. Even among those systems that have chosen to prioritize independence, many have started transitioning towards designing more straightforward links to foster the incorporation of evidence

33 The analysis of the Country M&E System cases was based on the Independent Evaluation Group ECD Working Paper Series. See the cases at ieg.worldbankgroup.org/evaluations/evaluation-capacity-development-ecd.

into decision-making. The links established among the evaluation findings, project and programme improvement, and functions such as planning and budgeting, are critical. Many independent evaluation functions, such as those of the National Council for the Evaluation of the Social Development Policy in Mexico and the Independent Evaluation Group, have started to set up robust mechanisms to follow up on the recommendations of evaluations.

Different institutional arrangements, processes, incentive structures, M&E tools and methodologies have helped address the many risks and biases related to credibility that were particularly relevant for their contexts. Variations in cultures, time periods and topics need different degrees of independence. As M&E systems evolve, establishing better links between independence and results still presents an opportunity for innovation and advancement in the M&E field.

One of the most interesting questions going forward is how other countries and organizations have achieved an optimal level of independence that allows for improving development results. Which arrangements have been successful in linking independence, credibility, and the use of evidence? What competencies and incentives have been successful in fostering these links? We look forward to the further debates on this topic.

REFERENCES

- Heider, Caroline, "Independence for Accountability and Learning", Forthcoming.
- International Federation of Accountants. 2003. *Handbook of International Auditing, Assurance and Ethics Pronouncements*. International Federation of Accountants.
- Mackay, Keith Robin. 2007. *How to Build M&E Systems To Support Better Government*. Independent Evaluation Group. Washington: World Bank.
- Mark, Katharine, and John R. Pfeiffer. 2011. "Monitoring and Evaluation in the United States Government: An Overview". ECD Working Paper 26. Washington: World Bank.
- Mayne, John. 2008. *Building an Evaluative Culture for Effective Evaluation and Results Management*. Institutional Learning and Change Initiative (ILAC).
- Organisation for Economic Co-operation and Development, Development Assistance Committee (OECD-DAC), Working Group on Aid Evaluation. 2002. "Glossary of Key Terms in Evaluation and Results Based Management". Available at oecd.org/development/peer-reviews/2754804.pdf. Last accessed 5 February 2014.
- Picciotto, Robert, and Eduardo Wiesner Durán (Editors). 1998. *Evaluation and Development: The Institutional Dimension*. Vol. 1. Transaction Publishers.
- Picciotto, Robert. 2008. "Evaluation Independence at DFID". Independent Advisory Committee for Development Impact.
- Picciotto, Robert. 2013. "Evaluation Independence in Organizations". *Journal of MultiDisciplinary Evaluation* 9 (20):18-32.
- Rist, Ray (Editor). 2011. *A Vision for Development: Dialogues on the Work of Vinod Thomas*. Independent Evaluation Group. Washington: World Bank.

Scriven, Michael. 1991. *Evaluation Thesaurus*. Sage Publications.

Thomas, Vinod, and Jiro Tominaga. 2010. "Evaluation for Better Development Results." *Journal of Development Effectiveness* 2(3):371-386.

United Nations Evaluation Group (UNEG). 2005. "Standards for Evaluation in the UN System". Available at uneval.org/papersandpubs/documentdetail.jsp?doc_id=22. Last accessed 5 February 2014.

World Bank Independent Evaluation Group. 2011. "Self-Evaluation of the Independent Evaluation Group". Washington: World Bank.

World Bank Independent Evaluation Group. 2013. "Management Action Record Database." Washington: World Bank.

4. INNOVATIONS IN MONITORING AND EVALUATING RESULTS

THOMAS WINDERL

Consultant for Planning, Monitoring and Evaluation

Paper commissioned by the Knowledge,
Innovation and Capacity Group, UNDP

THE DEVELOPMENT CONTEXT

The rapid emergence and adoption of innovations in monitoring and evaluating results³⁴

Countries are increasingly using **innovative approaches to manage the performance of public policies, programmes and service delivery**. These approaches are fostering more inclusive, collaborative and responsive processes across the development cycle: from planning to implementation, monitoring and evaluation. Two critical commonalities among the innovations explored in this paper are: 1) the increased frequency of input and feedback; and 2) the expanded definition of and outreach to stakeholders, including those not traditionally part of the development process. Increased participation is, in some cases, direct (e.g. citizens providing input through SMS reporting or storytelling), and indirect in other cases (e.g. information being collected and analysed remotely and in the aggregate). Many of the innovations are also characterized by their relatively low cost and lower degree of formality and rigidity.

These innovations bring a significant benefit to the development process by enabling more frequent testing of theories of change and facilitating timely course corrections based

34 Partially based on "Innovations in Monitoring and Evaluation 'as if Politics Mattered'", Concept Note, Roche/Kelly, 04-08-2011, mande.co.uk/2011/coming-events/conferences/innovations-in-monitoring-and-evaluation-as-if-politics-mattered/; (2) "Answering Calls for Integrative Capacity Development for Sustainable Development, Group 3: Monitoring and Evaluating Sustainable Development Results: Flexible, Nimble and Dynamic Approaches", 15/17-10-2012, Bratislava, Slovakia, draft; (3) "Using Technology for Development Project Monitoring & Evaluation: An Interview with Linda Raftree", 02-07-2013, bestict4d.wordpress.com/2013/07/02/using-technology-for-development-project-monitoring-evaluation-an-interview-with-linda-raftree/.

on evidence. By gathering frequent input on the building blocks of policies, programmes and service delivery from those most affected, hurdles and bottlenecks are more easily identifiable. When organizations are capable of absorbing this information and have systems flexible enough to respond to it, they achieve better results: more relevant policies, more effective programmes and improved service delivery.

Innovations in monitoring and evaluating results are emerging and being adopted at such a rapid pace for a number of reasons.³⁵ On the one hand, M&E has to respond to the higher demands placed on it and to the fast-changing environment. On the other hand, as technology moves forward, opportunities for innovation in M&E are opening up.

A number of factors are driving innovation in M&E:

The need for flexible and faster M&E

Increased unpredictability, rapidly changing circumstances and a dynamic environment for public action require more flexible, dynamic and nimble approaches to M&E that capture and adapt to rapidly and continuously changing circumstances and cultural dynamics. Traditional approaches of diligently checking if a public policy, programme or service is 'on track' in achieving a predefined milestone is often not sufficient anymore. Further, feedback loops of traditional monitoring (with quarterly and annual monitoring, mid-term reviews, final evaluations, annual reporting, etc.) have often proven to be too slow to influence decision-making in time. More timely real-time updates are required for better use of monitoring information and evaluation findings.

Theories of change need intermediate outcomes that can be measured quickly and easily

There is an increased emphasis on measuring outcomes (changes in behaviour and performance) as a result of public policy, programmes and service delivery. Due to their nature, however, outcomes are typically more difficult to monitor and evaluate, since data is often not readily available and primary data collection is typically required. A theory of change that includes a more proximate series of outcomes that can be measured and reported on more quickly and easily ('fast-cycle measurables') can be used as a meaningful tool to manage and assure the quality of policies, programmes and service delivery.

Civil society demands accountability

In countries with an increasingly energetic civil society, there is growing public demand for greater transparency and public accountability. This, in turn, requires more rigorous monitoring and evaluation of public policies, programmes and service delivery. A lack of objective evidence on the performance of policies, programmes and service delivery may contribute to a lack of accountability and even misappropriation of resources.

35 For a detailed critique of current M&E practices from a women's rights perspective, see e.g. "Capturing Change in Women's Realities: A Critical Overview of Current Monitoring and Evaluation Frameworks and Approaches", Batliwala/Pittman, Association for Women's Rights in Development, December 2010; available at awid.org/About-AWID/AWID-News/Capturing-Change-in-Women-s-Realities.

Avoiding cognitive bias

More traditional M&E methods, such as focus groups or surveys, require interpretation by experts who may build in their biases or reinterpret, rather than aggregate, citizens' inputs in order to uncover patterns. With increasing application of behavioural economics to policy making, this potentially detrimental impact of cognitive biases on decision-making is becoming more obvious.

Single method is not sufficient anymore

Public policies, programmes and service delivery operate in increasingly complex and ever-changing social, economic, ecological and political contexts. No single M&E methodology can adequately describe and analyse the interactions among all of these different factors. Mixed methods allow for triangulation—or comparative analysis—which is better suited to capture complex realities and to provide different perspectives on the effect of public policies, programmes or service delivery.

In addition, a number of factors are currently enabling innovation to take place in M&E:

More mature civil societies

In many countries, a more matured civil society is increasingly willing and able to participate in the planning, monitoring and evaluation of public policies, programmes and service delivery. This is partially also due to new information and communication technology tools.

Boom of information and communication technology

Advances in and the spread of information and communication technology open up a wide range of new opportunities for innovations in M&E. This includes the spread of access to the Internet and mobile phone networks, the proliferation of mobile phones and other hand-held devices, better and cheaper satellite and aerial remote sensing, the production of inexpensive sensors (such as pyro-electric heat sensors and pressure slab sensors), as well as sophisticated software for data analysis and mining.

Rise of 'big data'

The explosion in the quantity and diversity of high-frequency digital data holds the potential—as yet largely untapped—to allow decision makers to track the performance and effects of social policies, programmes and service delivery to better understand where adjustments are required. Big data is an umbrella term for call logs, online user-generated content (e.g. blog posts and Tweets), online searches, satellite images, and mobile-banking transactions. Big data usually requires computational techniques to unveil trends and patterns and turn them into actionable information.

ISSUE ANALYSIS

Features of innovations in monitoring and evaluating results

An innovation is the introduction of something new, whether an idea, method, or device (Merriam-Webster dictionary). Typical categories of innovations for monitoring and evaluating the performance of public policies, programmes or service delivery are: a) technological

innovations; b) innovative products; c) innovative services; d) innovative processes; or e) innovative interactions and partnerships. In this paper, we consider a product, process, service or a technology to be an innovation in M&E if at least two of the following criteria are met:³⁶

Significant process improvement

Innovations in M&E are technologies, products, services, processes or interactions that have shown a significant impact on how M&E is done (not just innovation for innovation's sake) or have a clear potential to change M&E in order to improve the value or usefulness of monitoring information and evaluation findings. Typically, innovations with a great potential impact also address a core need or core challenge in M&E.

Catalytic change

Innovations in M&E have to go beyond incremental change and reframe, re-imagine, or recombine different existing elements to yield a new pathway for M&E. In other words, an innovation in M&E is not simply a better, faster, cheaper way of doing the same thing. Innovation requires going beyond current models of thinking in M&E. That is why it often takes outsiders or unconventional partnerships to break old paradigms in M&E.

Concrete

Innovations in M&E must be sufficiently concrete. Ideas and theoretical approaches are not innovations (although they can lead to innovations). Innovations are concrete if they are already being implemented (at least as pilots), can be replicated and are potentially scalable across different contexts and regions.

Innovations focus on monitoring, less so on evaluations— but distinctions get blurred

Most of the innovations examined here can be directly used for monitoring public policies, programmes and service delivery, while only a few innovations focus exclusively on evaluation (e.g. multilevel mixed evaluation methods, outcome harvesting). This could imply that in the current environment, the push for innovations is mostly driven by the need and the possibility for better, more frequent and real-time monitoring. The clear distinction between monitoring and evaluation³⁷ in traditional M&E, however, appears to get more and more blurred:

- Many of the innovative tools can be used for monitoring as well as for evaluations (e.g. crowdsourcing, micro-narrative, mobile data collection, data exhaust, data visualization).

36 Partially adopted from: EvalPartners Evaluation Challenge (mymande.org/evalpartners/innovation_challenge); "Six Fundamental Truths About Innovation" (blog), Stikeleather 2013, Management Innovation eXchange, 28-02-2013 (managementexchange.com/blog/six-fundamental-truths-about-innovation); "Principles – Stories of Innovation", UNICEF (unicefstories.org/principles); "Innovation for development: what is really different?", Quaggiotto 2013, Voices from Eurasia (blog), UNDP, 18-03-2013 (europeandcis.undp.org/blog/2013/03/18/innovation-for-development-what-is-really-different).

37 Monitoring is continuous, often internal and tracks delivery and the achievement of results; evaluation is one-off, typically external and goes beyond results by questioning their value.

- With better data collection tools for monitoring, information that was traditionally only collected occasionally through evaluations (e.g. through a baseline, mid-term and final survey) now becomes available on a continuous basis.
- The increasing demand for real-time information increases the need for solid monitoring information over much less frequent evaluations.

INVENTORY OF INNOVATIONS

Eleven innovations have been identified based on extensive research and analysis. Increased frequency of input and broader citizen participation are key features in most of the innovations presented in this paper. In addition, many present cost-conscious and flexible approaches to managing and assuring quality of policies, programmes and service delivery. The first eight innovations promote citizen engagement, with the first five requiring active participation of citizens and the next three reflecting more passive engagement. The ninth is designed to enhance the usefulness and accessibility of the information collected, and the final two present progressive methodologies for more credibly measuring and interpreting results. Most of the innovations are not mutually exclusive. For example, mobile data collection can be used with micro-narratives to provide different perspectives on a particular initiative.

1. Crowdsourcing

What is it?

- A large number of people **actively** report on a situation around them, often using mobile phone technology and open source software platforms; and
- ‘Citizen reporting’ or ‘See something, text something’.

Why is it innovative?

- While traditional M&E is sometimes perceived as intrusive and extractive, citizen reporting is a monitoring and evaluation technique that results in a win-win situation for M&E, potentially leading to greater citizen participation and civic engagement (**process improvement**);
- Allows data collection: a) on a scope usually not feasible through traditional M&E tools; and b) on sensitive issues that more traditional tools would struggle to cover (**catalytic**); and
- A great variety of open source software platforms already exist and the approach is implemented in a number of countries and projects (**concrete**).

How and when best to use it

- Where requirements for data collection go beyond the scope of more traditional monitoring or evaluations, or when quantitative information is required; and
- For sensitive issues where anonymity is preferred (e.g. corruption).

Advantages

- Can gather massive, location-specific data in real-time with lower running costs than more traditional methods;³⁸
- Can boost civic engagement by establishing direct channels of communication from the ground up; and
- If systems are set up right, crowdsourced data tends to be more difficult to manipulate and less vulnerable to biased interpretation, therefore potentially increasing independence and credibility.

Disadvantages

- Requires incentives for citizens to continuously participate; and
- Requires tailoring a crowdsourcing platform.

Tools

- **Ushahidi platform:** a crowdsourcing mapping tool;³⁹
- **SeeClickFix:** a communications platform for citizens to report non-emergency issues and for governments to track, manage and reply;⁴⁰
- **FrontlineSMS:** an open source software to distribute and collect information via text messages;⁴¹
- **RapidSMS:** a open-source framework for dynamic data collection, logistics coordination and communication, leveraging basic mobile phone text technology;⁴² and
- **Ideascale:** a platform for stakeholders to share, vote and discuss feedback.⁴³

38 E.g. checking all the defects in the traffic lights by having city officials patrol is expensive, while asking citizens or taxi drivers to report defective traffic lights is cheaper.

39 See ushahidi.com.

40 See seeclickfix.com.

41 See frontlinesms.com.

42 See rapidsms.org.

43 See ideascale.com.

2. Real-time, simple reporting

What is it?

- A means to reduce to a minimum the formal reporting requirements for programme and project managers and free up their time to provide more frequent, real-time updates, which may include text, pictures, videos that can be made by computer or mobile devices.

Why is it innovative?

- Can overcome an often-voiced dissatisfaction with excessive, detailed and frequent reporting requirements that may result in unread and under-used reports (**impact**);
- Through mutual agreement, the tendency towards more and more data collection and analysis is reversed; long-established but possibly outdated reporting practices are eliminated or complemented, allowing instead for real-time reporting (**catalytic**); and
- Concepts and digital platforms for real-time, simple reporting exist and are in use (**concrete**).

How and when best to use it

- Can be used for all types of public policies, services and programmes if the minimum information needs can be covered by the short reports.

Advantages

- Near real-time updating of progress and results;
- The voices and faces of citizens become more and more directly visible through photos, video and audio recordings;
- Works well with organizations or units with a large number of programmes, projects and partners; and
- Short but real-time reports are more likely to be used by management for decision-making.

Disadvantages

- Reports are limited to key information and do not go into much detail; and
- Potential tendency to collect the most easy-to-measure data, resulting in a reporting bias.

Tools

- **Akvo Really Simple Reporting** is a Web-based system that brings complex networks of projects online and instantly shares progress with everyone involved and interested on multiple websites.⁴⁴

3. Participatory statistics

What is it?

- An approach in which local people themselves generate statistics; and
- Participatory techniques (e.g. participatory mapping, ‘ten seeds technique’, pairwise ranking, proportional piling, matrix ranking) are replicated with a large number of groups to produce robust quantitative data.

Why is it innovative?

- Participatory statistics change the paradigm that data collection is a top-down, centralized process by decentralizing statistical data collection and empowering citizens who are most familiar with local information (**catalytic**);
- Can make it possible to collect statistics on sensitive topics that are largely inaccessible to standard surveys (**process improvement**); and
- Participatory approaches to M&E are well-tested; aggregation to produce statistics—even on the national scale—is increasingly tested and applied following methodological breakthroughs in the 2000s (**concrete**).

How and when best to use it

- Particularly suitable for social and census mapping, household listing and scoring, well-being ranking, trend and change analysis, seasonal diagramming, preference ranking, causal linkage analysis and problem trees; and
- If empowerment is part of a public policy, service or programme.

Advantages

- When carefully aggregated and triangulated, participatory statistics can produce more valid, reliable, and accurate data for M&E;
- Can empower citizens through an M&E process that has traditionally been highly extractive and externally controlled; and

44 See akvo.org/web/akvo-rsr.

- Generating and aggregating local data can make statistics more accurate, especially on sensitive issues, thus increasing accuracy, reliability and ultimately credibility and potential use of data.

Disadvantages

- Can be time-consuming if citizens are asked to collect the necessary data; and
- Needs to be built into a policy, service delivery or programme from the very beginning.

Tools

- **Participatory Methods website**, Institute of Development Studies.⁴⁵

4. Mobile data collection

What is it?

- The targeted gathering of structured information using mobile phones, tablets or PDAs using a special software application; and
- Differs from citizen feedback or crowdsourcing, which mine unstructured digital information; instead, mobile data collection systems run designed surveys which collect specific information from the target audience.

Why is it innovative?

- In addition to an incremental change from paper-based surveys, mobile data collection can include **completely new information** in designed surveys: geographic location through automatic geo-tagging, photographs and video (e.g. as additional evidence that corroborates information obtained through a questionnaire) and audio (to record survey responses as proof and for further analysis) (**catalytic**); and
- Availability of inexpensive mobile phones and specialized software platforms (to build a mobile data collection survey) are widely available (**concrete**).

How and when best to use it

- Where the advantages of mobile data collection outweigh the advantages of a more traditional paper-based survey; and
- Where data collection requires or significantly benefits from audio, video or geographic information.

45 See participatorymethods.org.

Advantages

- Can improve the timeliness and accuracy of the data collection; and
- Platforms allow one to customize the survey to include photographs, voice recordings, GPS coordinates and other information usually not collected through paper-based surveys.

Disadvantages

- Technology alone will not improve the survey design or instrument; and
- Potential bias in favour of well-educated or well-off citizens.

Tools

- Numerous platforms and tools.⁴⁶

5. The micro-narrative

What is it?

- The collection and aggregation of thousands of short stories from citizens using special algorithms to gain insight into real-time issues and changes in society.

Why is it innovative?

- Information collected in the shape of stories is interpreted by the person who has told a story, therefore removing the need for—and the potential bias of—a third party to interpret the data; this meets a core challenge for M&E by reducing or eliminating potential biases of monitoring staff and evaluators (**process improvement**);
- By using a large number of stories, this approach turns previously mostly qualitative data (e.g. in the form of a limited number of not representative case studies included in an evaluation) into aggregated statistical data; the approach has the potential to replace traditional monitoring tools like surveys and focus groups (**catalytic**); and
- Pattern detection software for analysing micro-narratives exists, and the approach is already implemented in a number of countries and projects (**concrete**).

How and when best to use it

- When real-time quantitative information from a large number of beneficiaries is required and cannot otherwise be collected.

⁴⁶ For a detailed list, see “NOMAD Preliminary List of Mobile Data Collection Technologies,” Annex 1, in *Mobile Data Collection Systems: A review of the current state of the field*, June 2011, NOMAD; available at humanitarian-nomad.org/wp-content/uploads/2013/03/NOMAD-MDC-Research.pdf.

Advantages

- Provides governments, for example, access to real-time data for faster, more informed decision-making;
- Allows evaluators to collect independent quantitative information from a potentially large number of citizens, potentially increasing the credibility of data collected;
- Makes it possible to design, monitor and evaluate evidence-based policies and programmes under conditions of uncertainty;
- By detecting weak initial signals in the stories collected, this approach can provide early warning signs for policy or programme implementation in the communities they are trying to effect; this introduces the possibility for the first time of predicting future developments and building foresight into decision-making; and
- Lower running costs once set up compared to repeated surveys.

Disadvantages

- High initial investment in pattern detection software (e.g. proprietary software like Sensemaker) and information campaigns to inform and motivate participants; and
- Citizens must have the skills and continuous incentives to participate.

Tools

- **Sensemaker**, a proprietary pattern detection software for analysing micro-narrative;⁴⁷ and
- **GlobalGiving Story Tools**.⁴⁸

6. Data exhaust

What is it?

- Wherever citizens use mobile phones or access web content, they are leaving trails behind in the form of transactional data called 'data exhaust'; and
- Data exhaust is massive, passively collected transactional data from people's use of digital services like mobile phones and web content such as news media and social media interactions, which distinguishes it from other elements of big data such as citizen reporting, crowdsourcing or physical sensors.

47 See sensemaker-suite.com by Cognitive Edge (cognitive-edge.com).

48 See globalgiving.org/story-tools.

Why is it innovative?

- The availability of passive transactional data has increased exponentially; the private sector is already using innovative technologies to analyse data exhaust from commercial services to understand customers, identify new markets and make investment decisions; for monitoring and evaluating public policies, services and programmes, analysing existing data exhaust can dramatically change how M&E is done and what data is available for M&E (**catalytic**); and
- Commercial services have demonstrated that making use of data exhaust is possible and useful (**concrete**).

How and when best to use it

- When analysed in bulk, data exhaust makes it possible to calculate the current status of entire communities and identify changes happening in real-time through Web-based and social media search queries; and
- This conversational data can also be used to predict human behaviour.

Advantages

- Data is already collected and available; and
- Can allow mining of massive qualitative data to distil quantitative information that would otherwise be beyond the reach of traditional M&E, thereby increasing the potential credibility of monitoring or an evaluation.

Disadvantages

- Potential bias that makes digital data skewed in favour of better-educated, well-off citizens while neglecting those less articulate or with less access to digital services.

Tools

- **CellCensus** makes use of cell phone records, which show the social network of a person or his/her mobility patterns and are strongly predictive of socio-economic factors;⁴⁹
- **Google Trends**, a free tool to track the level of Google search requests over time;⁵⁰ and
- **Recorded Future**, a commercial service that scans tens of thousands of digital sources to explore the past, present and predicted future of a wide variety of things.⁵¹

49 See vanessafriasmartinez.org/CenCell.html.

50 See google.com/trends.

51 See recordedfuture.com.

7. Intelligent infrastructure

What is it?

- Equipping all—or a sample of—infrastructure or items (e.g. roads, bridges, buildings, water treatment systems, hand washing stations, latrines or cook stoves) with low-cost, remotely accessible electronic sensors.

Why is it innovative?

- Automatization of data collection can radically change how and how frequently data is collected in cases where policies, programmes or service delivery include infrastructure or items (**process improvement**);
- Involves unconventional partnerships between high-tech research departments, start-up enterprises, governments and development organizations (**catalytic**); and
- Inexpensive electronic sensors have recently become commercially available, but there are only a few examples where they have started to be used for M&E (**concrete**).

How and when best to use it

- When monitoring or an evaluation attempt to measure and track over time the value of infrastructure or public services to the people (e.g. to determine whether the infrastructure is actually used enough to justify the cost);
- Low-cost, low-power, reliable electronic sensors attached to infrastructure relay usage or operational data in near real-time to the Internet via cellular phone technology, feeding into an automated, remote monitoring system; and
- When data is actually required for a certain purpose, and not simply because the technology exists.

Advantages

- The massive amounts of data generated can be used to better understand programmatic, social, economic, and seasonal changes and behavioural patterns that influence the quality of a policy or a service;
- Real-time data on infrastructure or public service use makes faster, more informed decisions possible;
- Potentially lower running costs once system is set up compared to repeated sample surveys using experts and enumerators; and
- More objective and real-time operational data on the usage and performance of infrastructure or services may result in greater credibility and use of monitoring information and evaluations.

Disadvantages

- Initially expensive, high-tech monitoring option which requires special technical expertise;
- Lack of maintenance or malfunctioning equipment can ‘contaminate’ data; and
- Potential privacy concerns if users, or user groups, can be identified.

Tools

- **SWEETSense**, a technology and concept tested and demonstrated by the Sustainable Water, Energy and Environmental Technologies Laboratory (SWEETLab) at the Portland State University.⁵²

8. Remote sensing

What is it?

- Observing and analysing a distant target using information from the electromagnetic spectrum of satellites, aircraft or other airborne devices; and
- Passive sensors detect natural radiation (e.g. reflected sunlight through film photography); active remote sensing involves the emitting of energy in order to scan objects and areas.

Why is it innovative?

- Since the early days of satellite remote sensing in the 1950s, it has been applied to many disciplines in natural science; applying remote sensing to social research and monitoring and evaluations of social public policies and programmes can have a potentially great impact for large-area monitoring (**process improvement**);
- Allows remote monitoring in areas previously inaccessible due to physical barriers or security concerns (**catalytic**); and
- Passive and active remote sensing information and commercial technology for collecting information (e.g. mini-drones, pattern recognition software) is available (**practical**).

How and when best to use it

- When access is limited due to physical barriers or security concerns;

52 See sweetlab.org/sweetsense.

- For observable changes on the Earth's surface (such as agriculture, deforestation, glacial features and oceans) and natural resource management in general, but also for monitoring social public policies and programmes related to urban areas, demography, land use and land cover, humanitarian conflicts or disasters, or as a proxy for wealth; and
- For social policies and programmes, remote-sensing data might be at its most valuable when used in combination with traditional methods, such as surveys, public records, interviews and direct observation.

Advantages

- Possible to collect data on dangerous or inaccessible areas; and
- Observed objects or people are not disturbed.

Disadvantages

- Privacy concerns over government misuse of information; and
- Potentially high costs for obtaining images or for primary data collection using remote sensors.

Tools

- **Sensefly** operates autonomous mini-drones and related software solutions for accurate mapping of mining sites, quarries, forests, construction sites, crops, etc.⁵³

9. Data visualization

What is it?

- Representation of data graphically and interactively, often in the form of videos, interactive websites, infographics, timelines, data dashboards⁵⁴ and maps.⁵⁵

Why is it innovative?

- Previously heavily reliant on text to communicate monitoring or evaluation findings, the increasing use of suitable data visualization tools in M&E changes the way data is analysed and represented (**catalytic**);

53 See sensefly.com.

54 Visual displays of the most important information consolidated on a single screen or page.

55 Spatial and conceptual representations of important physical elements of an area to assist in identifying patterns (e.g. demographic mapping, GIS Mapping, Geotagging, social mapping, interactive mapping).

- The graphical and interactive presentation of data has the potential to dramatically increase the accessibility of complex data sets and, in turn, the use of the data (**process improvement**); and
- A great variety of free and commercial data visualization tools are available and increasingly used for monitoring, reporting and evaluations.

How and when best to use it

- To better identify trends and patterns of complex or large data sets during the analysis phase of monitoring or of an evaluation; and
- To better communicate information resulting from monitoring or from evaluations.

Advantages

- Effectively visualized data is more likely to be understood and used; and
- Visualized data can identify trends and patterns that could otherwise be unclear or difficult to discern.

Disadvantages

- Visualization needs to fit the purpose of analysis and the intended target audience of communication; and
- Identifying and putting together data visualization can be time-consuming, or costly if outsourced.

Tools

- **DevInfo**, a database system for organizing, storing and visualizing data in a uniform way;⁵⁶
- **Tableau**, a set of software solutions to combine, analyse and visually show data;⁵⁷
- **Google Fusion Tables**, a tool to combine, visualize and share data;⁵⁸
- **Visual.ly**⁵⁹ or **Easel.ly**,⁶⁰ tools to get inspired by and/or commission infographics; and
- **TimelineJS**, a tool to establish visually-rich, interactive timelines.⁶¹

56 See devinfo.org.

57 See tableausoftware.com.

58 See google.com/drive/apps.html#fusiontables.

59 See visual.ly.

60 See easel.ly.

61 See timeline.verite.com.

10. Multilevel mixed evaluation method

What is it?

- While parallel or sequential mixed methods have long been a typical design for development evaluations, this approach includes the deliberate, massive and creative use of mixed (quantitative and qualitative) methods on multiple levels for complex evaluations, particularly for service delivery systems.

Why is it innovative?

- With multilevel mixed methods rapidly becoming the standard method in evaluations, this leads to a paradigm change in evaluation methodology; evaluations using a single method or only nominally applying a mixed-method approach (e.g. a largely quantitative evaluation complemented with a limited number of focus group discussions) may stop being acceptable to governments and development organizations (**process improvement**); and
- While not yet widely used for evaluations, tools and guidelines exist that describe multilevel mixed methods, and some evaluations have experimented with the approach (**concrete**).

How and when best to use it

- Particularly suitable for the evaluation of service delivery systems (e.g. district education departments, state-level health services, a national programme to strengthen municipal governments) that require description and analysis of links between different levels; and
- For very complex and potentially expensive evaluations where multilevel mixed methods can provide valid and credible findings on the basis of smaller and more economical samples.

Advantages

- The multiple mix of quantitative and qualitative methods can lead to more validity, reliability and variety of findings, insights into sensitive subjects, and the revealing of unexpected findings with policy implications; and
- Multiple options for triangulation between different quantitative and qualitative methods and data sources.

Disadvantages

- Requires careful and deliberate planning of an appropriate methodological mix to be credible; and
- Usually requires a team of evaluators with experience in quantitative and qualitative methods and in how to combine them at multiple levels.

Tools

- “Introduction to Mixed Methods in Impact Evaluation”⁶²

11. Outcome harvesting

What is it?

- An evaluation approach that, unlike some evaluation methods, does not measure progress towards predetermined outcomes, but rather collects evidence of what has been achieved and works backward to determine whether and how the project or intervention contributed to the change; and
- An approach inspired by ‘outcome mapping’.⁶³

Why is it innovative?

- Allows the evaluation of policies or programmes where relations of cause and effect are not fully understood and that have previously been difficult to evaluate (**catalytic**);
- Is suitable to search and identify unintended results that frequently escape more traditional evaluation methods (**process improvement**); and
- The outcome harvesting approach has been tested in evaluations since 2010 (**concrete**) (Wilson-Grau 2012).

How and when best to use it

- When relationships of cause-effect of public policies or services are unknown; and
- In situations where complexities are high and outcomes are ill-defined or unclear (e.g. advocacy work, networks, research centres and think tanks).

Advantage

- Can be used for complex policies, services or programmes that are not based on a clear results chain or theory of change.

62 Bamberger 2012, InterAction/The Rockefeller Foundation, Impact Evaluation Notes, No.3. August 2012. Available at interaction.org/document/guidance-note-3-introduction-mixed-methods-impact-evaluation.

63 Outcome mapping (outcomemapping.ca) is a related but broader approach that includes: a) intentional design; b) outcome and performance monitoring; and c) evaluation planning. Outcome harvesting is more narrow in scope, roughly the equivalent of steps 8, 9, 10 and 11 of outcome mapping (Wilson-Grau 2012).

Disadvantages

- Participatory process to reach a consensus can be time-consuming; and
- A potential bias by evaluators in interpreting the expected outcome of public policies, programmes or services might skew findings.

Tools

- **Outcome harvesting**⁶⁴

POLICY OPTIONS AND ACTIONS

Trends in innovation in monitoring and evaluation

Analysis of the 11 key innovations identified above—which can result in process improvements, are catalytic for M&E and are sufficiently concrete—leads to some observations on current trends:

Frequency of feedback increases

With better data-collection tools, information that was traditionally only collected occasionally through planned M&E activities (e.g. through a baseline, mid-term and final survey) now becomes available on a continuous basis.

Innovations depend upon increased citizen engagement

Many of the interventions identified rely on increased citizen participation. Several open up direct communication channels with citizens or beneficiaries (e.g. crowdsourcing, real-time simple reporting, micro-narratives and participatory statistics).

Innovations are being applied throughout the development cycle

There is increasing demand for real-time information throughout the development process, and many of these innovative tools can be applied just as constructively for planning as for monitoring (e.g. crowdsourcing, micro-narrative, mobile data collection, data exhaust and data visualization).

ICT sparks innovations in M&E

Most of the key innovations identified have a strong information and communications technology (ICT) component (i.e. crowdsourcing, real-time simple reporting, mobile data collection, micro-narrative, data exhaust, intelligent infrastructure, remote sensing and data visualization). It appears that the sudden supply of sophisticated ICT-based M&E tools has sparked a wave of innovations in monitoring and evaluation that would not have been possible only a few years ago.

64 See outcomemapping.ca/resource/resource.php?id374.

Academia, private sector and development organizations innovate

Many of the innovations mentioned are initially developed or adapted by non-governmental organizations or bilateral or multilateral development organizations (many mobile data-collection platforms, real-time, simple reporting, for example, Akvo, data-visualization tools such as DevInfo and Gapminder, and crowdsourcing tools such as the Ushahidi platform, Frontline SMS or RapidSMS). Some are originating within the communities of academia and development practitioners (such as the multilevel mixed method, participatory statistics, outcome harvesting). A number of innovative tools are also coming from academia but using a private-sector approach (algorithm and software for micro-narrative such as Sensemaker, some crowdsourcing applications, and intelligent infrastructure like SWEETsense). Finally, a surprising number of innovative tools are coming from the private sector (remote sensing such as senseFly; data exhaust and data-visualization tools such as Tableau, Visual.ly, Easel.ly, TimelineJS; data exhaust such as Recorded Future or Google Trends; and some crowdsourcing tools such as SeeClickFix), which might indicate a greater reliance of development on commercial, private-sector innovation.

How to apply innovations: planning and programming considerations

by facilitating the increased frequency of input and heightened citizen engagement, innovations in M&E have a series of implications for the entire development process. The innovations presented in this paper connect the M&E function to the planning and implementation processes, in that many of the innovations can just as easily be used to gather useful information for the design of more relevant policies, and that programmes, and their results frameworks, should reflect the fact that information can now be collected more frequently. These innovations also allow for more timely adjustments to policies, programmes and service delivery. Taken together, these aspects transition the M&E process into a more holistic management and assurance function that has implications for institutional capacities and processes:

Build M&E into the planning phase

Much more than with traditional M&E, which still often gets away with only vague statements on how M&E will be implemented in planning documents, many of the innovative approaches to M&E need to be built into the planning process of public policies, services or programmes. The current practice of tacking on M&E at the end of a plan is not sufficient for most innovative approaches, many of which require a lot of preparation (with the exception of the outcome harvesting tool, which is designed exactly to handle a situation where little thought was given to planning outcome-level M&E).

Design theories of change to monitor intermediary outcomes

Gathering real-time feedback allows for more frequent measurement of results. Theories of change should incorporate intermediate outcomes and indicators, so information on results at lower levels can be collected through fast feedback loops and used to make course corrections in programme implementation and service delivery.

Keep institutional planning and programming processes flexible

Institutional processes need to be flexible enough to screen, pilot, scale up and absorb innovations—and, importantly, the insights they generate—in M&E systems. Rigid planning and programming frameworks and systems that are focused on budgets, activities or outputs are less likely to allow experimentation and adoption of innovative approaches to M&E. To promote innovation, governments may need to introduce incentives for institutions to pilot and scale up new approaches to M&E that go beyond existing national M&E requirements.

Strengthen internal capacities or partner with third parties

Experts in planning, monitoring and evaluation do not necessarily understand innovations in information and communications technology. Conversely, the person developing information and communications technology tools does not necessarily understand managing for results. A key implication is that governments and organizations need to have the know-how to decide which information and communications technology solutions are appropriate for their needs and select the right tools for the job and the user. Especially for technological innovations, innovative M&E may require national or international technical expertise or services from the private sector, academia or elsewhere outside government (software for big data analysis, micro-narratives, mobile technology using SMS, sensors, etc.), or significant investment to increase national capacities for technological innovations.

Close the loop with citizens

Many of the innovations discussed above will only work effectively if incentives for citizens, service users or programme participants to provide feedback and mechanisms for closing the loop are built right into the design. Participatory statistics, mobile data collection and micro-narratives, for example, require us to give information back to people for the approach to be sustainable in the long run. Citizen reporting is particularly dependent on fast, visible responses to information provided by a citizen or programme participant.

Ensure privacy needs are met

As with more traditional tools and approaches for M&E, privacy needs have to be addressed. While innovative approaches typically pose more challenges with regard to privacy needs as data (and data sources) becomes more readily accessible (or sharable), the setting of privacy policies and ethical standards often lags behind technological advances. It is critical for policy makers as well as planners and implementers of innovative M&E to ensure the privacy of participants, so that they feel comfortable about engaging, and that the information collected is unbiased.

REFERENCES

- Bamberger, Michael. 2012. "Introduction to Mixed Methods in Impact Evaluation". InterAction/The Rockefeller Foundation, Impact Evaluation Notes, No.3. Available at interaction.org/document/guidance-note-3-introduction-mixedmethods-impact-evaluation. Last accessed 5 February 2014.
- Bamberger, Michael. 2013. "The Mixed Methods Approach to Evaluation". Social Impact Concept Note Series 1. Available at socialimpact.com/press-releases/MME613.pdf. Last accessed 5 February 2014.
- Batliwala, Srilatha, and Alexandra Pittman. 2010. "Capturing Change in Women's Realities: A Critical Overview of Current Monitoring and Evaluation Frameworks and Approaches". Association for Women's Rights in Development (AWID). Available at awid.org/About-AWID/AWID-News/Capturing-Change-in-Women-s-Realities. Last accessed 5 February 2014.
- Centers for Learning on Evaluation and Results (CLEAR). Undated. "Mobile-Based Technology for Monitoring and Evaluation: A Reference Guide for Project Managers, M&E Specialists, Researchers [and] Donors." Available at theclearinitiative.org/mobile-based-tech.pdf. Last accessed 5 February 2014.
- Wilson-Grau, Ricardo. 2012. "Outcome Harvesting". PowerPoint presentation. OM Lab, Beirut, Lebanon. Available at outcomemapping.ca/resource/resource.php?id=363. Last accessed 5 February 2014.

COUNTRY PAPERS

This section presents 31 country papers on the theme of the conference that provide a national perspective on building national evaluation capacities. The papers are organized by region.

COUNTRIES	TITLE
REGION: ASIA PACIFIC	
1. AFGHANISTAN	USE OF EVALUATION: LOCAL GOVERNANCE M&E SYSTEM IN AFGHANISTAN
2. AFGHANISTAN	POLITICAL PARTICIPATION FOR DEVELOPMENT EVALUATION IN SOUTH ASIA: AN AFGHANISTAN PERSPECTIVE
3. CAMBODIA	EVALUATION OF DEVELOPMENT ACTIVITIES AT THE MACRO LEVEL: CHALLENGES AND EXPERIENCES IN CAMBODIA
4. INDONESIA	MONITORING AND EVALUATION SYSTEM IN INDONESIA
5. SOUTH ASIA	WHY NATIONAL EVALUATION POLICIES MATTER IN SOUTH ASIA
6. SRI LANKA	COUNTRY-LED NATIONAL EVALUATION SYSTEM: INDEPENDENCE, CREDIBILITY AND USE OF EVALUATIONS; CHALLENGES AND SOLUTIONS
7. MALAYSIA	AN EVALUATION OF PUBLIC PERCEPTION TOWARDS GOVERNMENT PROJECTS: A CASE STUDY OF CONSTITUTION XYZ
8. NEPAL	INSTITUTIONALIZATION AND USE OF EVALUATIONS IN THE PUBLIC SECTOR IN NEPAL
REGION: LATIN AMERICA AND THE CARIBBEAN	
9. ARGENTINA	POLICY EVALUATIONS OF NATIONAL PUBLIC ADMINISTRATION: OVERVIEW
10. BRAZIL	A MODEL TO EVALUATE THE MATURITY OF THE BRAZILIAN PUBLIC ADMINISTRATION'S EVALUATION SYSTEMS
11. BRAZIL	EVALUATING SOCIAL DEVELOPMENT IN BRAZIL: ANALYSIS OF THE CYCLE OF PRODUCTION AND DISSEMINATION OF EVALUATION STUDIES BY THE MINISTRY OF SOCIAL DEVELOPMENT AND FIGHT AGAINST HUNGER
12. DOMINICAN REPUBLIC	ANALYSIS OF LESSONS LEARNED BY THE DOMINICAN REPUBLIC IN LOCAL AND INSTITUTIONAL/GOVERNMENT-LEVEL MONITORING, EVALUATION AND PARTICIPATORY FOLLOW-UP
13. DOMINICAN REPUBLIC	EXPERIENCE OF THE DOMINICAN REPUBLIC IN THE EVALUATION OF THE SOLIDARIDAD (SOLIDARITY) PROGRAMME
14. MEXICO	OVERCOMING MISTRUST IN POVERTY AND EVALUATION FIGURES: THE CASE OF CONEVAL IN MEXICO
15. MEXICO	EVALUATION USE AND ITS INSTITUTIONALIZATION IN THE FEDERAL EVALUATION SYSTEM IN MEXICO
16. MEXICO	INCENTIVES IN THE USE OF EVALUATIONS AS A TOOL FOR IMPROVING PUBLIC POLICY
17. PERU	EVIDENCE-BASED PUBLIC POLICIES: THE EXPERIENCE OF THE MINISTRY OF DEVELOPMENT AND SOCIAL INCLUSION
18. TRINIDAD AND TOBAGO	BUILDING A NATIONAL PERFORMANCE FRAMEWORK FOR EVALUATION

REGION: AFRICA	
19. BENIN	THE PROCESS OF INSTITUTIONALIZING EVALUATION IN BENIN: PROGRESS ON QUESTIONS OF USAGE AND INDEPENDENCE
20. GHANA	BUILDING M&E CAPACITIES TO ENHANCE THE NATIONAL M&E SYSTEM IN GHANA: THE WAY FORWARD
21. KENYA	CREDIBILITY OF EVALUATION: THE KENYAN CASE
22. MALAWI	MONITORING AND EVALUATION OF THE NATIONAL BUDGET IN MALAWI
23. MALAWI	PROMOTING USE OF EVALUATION RESULTS BY SENIOR GOVERNMENT OFFICIALS IN MALAWI
24. SOUTH AFRICA	REFLECTIONS ON THE SOUTH AFRICAN EXPERIENCE WITH EVALUATION AND THE USE OF EVALUATIVE EVIDENCE TO ORIENT PUBLIC POLICY FORMULATION
REGION: ARAB STATES	
25. EGYPT	INSTITUTIONALIZING AND STREAMLINING DEVELOPMENT MONITORING AND EVALUATION IN PRE- AND POST-REVOLUTIONARY EGYPT
26. MOROCCO	GENDER EVALUATION OF PUBLIC POLICIES FROM A HUMAN RIGHTS PERSPECTIVE
27. MOROCCO	EVALUATION OF THE NATIONAL INITIATIVE FOR HUMAN DEVELOPMENT: INDEPENDENCE AND USE
OTHERS	
28. CANADA	ALPHABET SOUP: LINKING THE LETTERS TO PROMOTE EVALUATION USE
29. ALBANIA	CHALLENGES OF A NEW EVALUATION NATIONAL SOCIETY
30. USA	THE DIVERSITY OF THE EVALUATION ENTERPRISE INFLUENCES ITS USE IN THE UNITED STATES
31. WORLD BANK	SCIENCE OF DELIVERY: IMPLICATIONS FOR MONITORING AND EVALUATION

1. AFGHANISTAN

USE OF EVALUATION: LOCAL GOVERNANCE M&E SYSTEM IN AFGHANISTAN

MOHAMMAD HAMED SARWARY

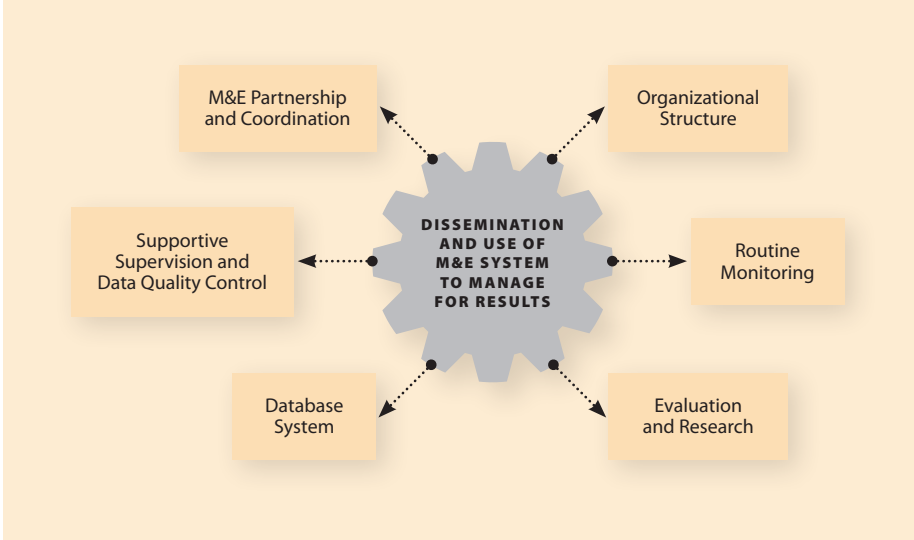
Monitoring and Evaluation Director,
Independent Directorate of Local Governance

INTRODUCTION

Established by a Presidential Decree in 2007, the Independent Directorate of Local Governance (IDLG) is Afghanistan's lead government agency for local governance. IDLG is mandated to administer and manage local government institutions (34 provincial governors' offices, 34 provincial councils, 150 municipalities and 370 district governors offices). IDLG and relevant ministries/entities developed the Afghanistan Sub-National Governance Policy, which lays the foundation of strengthening local governance (the policy was approved by the cabinet in 2010). To prioritize the Afghanistan National Development Strategy, the Government of Afghanistan and development partners developed national priority programmes for various sectors. IDLG, being the lead entity for local governance, developed the National Priority Programme for Local Governance in 2012, which will serve as a strategic priority document for local governance for a period of three years. To ensure effective implementation of the National Priority Programme for Local Governance and Sub-National Governance Policy, IDLG developed a comprehensive monitoring and evaluation (M&E) framework to measure the results, effectiveness and efficiency of its activities and performances based on national strategic documents for local governance.

IDLG's newly developed M&E framework has four components: monitoring, evaluation, research/surveys and data management/reporting. The framework was developed to establish a national M&E system for IDLG and its subnational entities. In addition, the new M&E framework also focuses on measuring the results and performances of the IDLG's national development programmes, which are funded by various donors and implemented by implementing partners.

FIGURE 1: COMPONENTS OF IDLG M&E SYSTEM FOR LOCAL GOVERNANCE

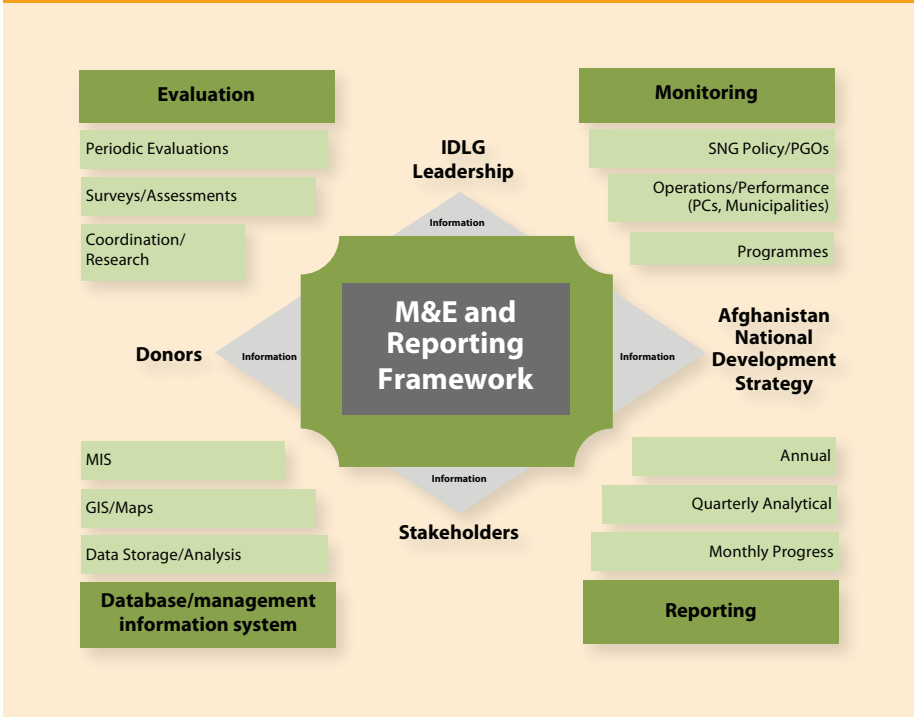


The new M&E framework is being implemented, and as a new practice for the organization and its entities, there are issues in institutionalizing the system at both the national and subnational levels. Major issues include government employees' limited understanding and capacities of M&E and its values. This is because the concept and use of M&E has been very limited in government agencies and ministries over the past years. M&E has been mainly (and somehow successfully) used in programmes and projects but has been limited in common government functions and activities. In particular, leadership, decision makers' use of evaluation and their recognition of its importance have been limited and not systematic. Therefore, this paper mainly focuses on the challenges and proposed solutions for the use of evaluations for local governance in Afghanistan, particularly by leadership and decision makers.

MONITORING AND EVALUATION IN AFGHANISTAN

The focus of this paper is on the use of evaluation. M&E is a relatively new practice in Afghanistan, especially within government institutions. The culture of undertaking professional and systematic evaluations, and consequently using the results and information as inputs for effective decision-making and planning, is still poor in the government institutions of Afghanistan. The concept of proper and effective M&E systems has been mainly consid-

FIGURE 2: IDLG M&E FRAMEWORK FOR LOCAL GOVERNANCE



ered for off-budget,⁶⁵ individual projects and programmes, which are typically implemented by non-governmental organizations and donors.

However, in the case of Afghanistan (which is a post-conflict or still-in-conflict country), during the past decade, donors and development partners have not fundamentally focused on applying M&E mechanisms when implementing projects and programmes. In part, this was because of limited capacity and understanding of M&E within the government to push towards more systematic accountability and transparency. This resulted in low-quality implementation of programmes and projects and did not contribute to building a culture of having M&E integration.

IDLG’s newly developed M&E framework for subnational governance was based on key strategic documents, which both the government and development partners agreed to implement. Institutionalizing M&E practices has been challenging, particularly because the system is being implemented while IDLG is still building systems for local government,

65 Programmes/projects implemented not through the national financial system but through contracts to non-governmental organizations/companies.

because of the lack of existing capacities and resources and because of a lack of demand for M&E systems. The government is less interested in applying systematic M&E for its common functions and public administration, while there remains a huge demand by development partners for doing period evaluations and for gathering reports on the performances and results of government functions, activities, programmes and projects.

There are various factors that limit government's demand and use of evaluations, particularly the lack of understanding by the senior management and leadership of the importance of M&E information. There are various factors that affect the process of decision-making based on evaluation findings and reports, such as political willingness, not completely relying on the reports, lack of technical understanding and awareness of the issues raised, and interventions and favouritism. These factors disturb an organization leadership's proper decision-making in planning and taking action. Further, government institutions are typically not very open to sharing information with civil society organizations or the public. This practice of limiting reports and information to the government limits the use of evaluation information and curbs actions on the findings and recommendations. In both internal and external scenarios, there is limited technical understanding of evaluation and the information that is shared.

As with development programmes that are implemented by non-governmental organizations or donors directly, government ownership of evaluations is limited; when there is limited government involvement in the evaluation process, the results and reports are often not seriously considered by the government. Specific evaluations undertaken by implementing partners or by donors are not widely shared for consideration with the government. However, IDLG has taken steps in this regard and, hopefully, by implementing its new M&E framework, IDLG will be in a position to conduct more objective evaluations of both its subnational entities and implementing partner programmes.

IDLG is adopting the practice. For example, IDLG implemented a five-year programme for provincial governor offices, which was managed by an implementing partner⁶⁶ although IDLG took full leadership on evaluations in its second phase. The quarterly evaluations of the programme were lead by the government with support from implementing partners and donors. Results of the quarterly evaluations were used by the government to measure the performances of local institutions (in this case, the provincial governors offices). The performance ranking system that was used for evaluating provincial governors offices' performance resulted in significant positive competition among the 34 offices. The incentive mechanism (performance-based funding decreases and increases) was a good practice and contributed to better and improved local government performance.

As a good experience, IDLG conducted an evaluation in 2012 of its six national programmes (which are being implemented by various implementing partners and funded by different donors). The results and findings of the evaluation were shared in the Sub-National

66 See pbgf.gov.af.

Governance Forum,⁶⁷ which was attended by key donors and partners. This exercise presented to all stakeholders that the government is now taking a lead and conducting evaluations of the programmes and projects. The forum also alerted partners to use the findings and to align their activities and programmes with government policies and plans.

Building on the positive results of government ownership of evaluation processes, it was recommended that donors ensure effective government participation and stake right from the design and at the programming stage. This year, donors and the government are jointly working to design programmes and to ensure that the government has an effective role in and ownership of evaluations, including reporting and decision-making, beginning with programme design. This is happening in IDLG with one of the key donors (the United States Agency for International Development) for its subnational governance programmes in Afghanistan. This will ensure the building of government capacities, ensure more ownership and will help in better decision-making by the government. Ultimately, this will contribute to better transparency, accountability and that programmes will be effectively implemented in accord with government policies and strategies.

CONCLUSIONS

Based on the above stated points about the use of evaluation, and based on my experiences in Afghanistan working both with the government and non-governmental organizations over the past years, I summarize my statements on why evaluations are not effectively used in public administration:

- Limited understanding and practical experiences of using effective evaluations within government institutions/systems. This is specifically the problem with decision makers and higher officials who are mostly politically oriented;
- Lack of institutionalization of systematic evaluation processes in government institutions due to various factors described above. In addition, there is limited attention from development partners or counterparts to conducting evaluations jointly or involving government institutions for ownership and capacity-building;
- It is sometimes difficult to apply effective and systematic evaluations to the government's common functions; evaluations are typically more applicable to specific development projects and programmes;
- Evaluation and its effective use are not separate, as both should be part of an M&E system within organizations, programmes and projects; Afghanistan still lacks proper and effective M&E systems;
- Due to lack of professional expertise, evaluations are seen as burdensome and something undertaken at the end of the work, and, because the nature of projects and programmes may change, as an exercise that may or may not be helpful to

67 The Sub-National Governance Forum is chaired by IDLG and the United Nations Assistance Mission in Afghanistan and conducted every six weeks to discuss the progress/achievements and issues regarding local governance.

future activities. Therefore, the findings are not considered as important to learning as focusing on gaining new programmes and resources; and

- Donors, government counterparts, and project and programme beneficiaries lack common or equal understanding of the value of evaluation. This is because the interests of these parties vary, and sometimes politics affects effective use of evaluations reports.

Considering the aforesaid issues, I make the following recommendations for the effective use of evaluation:

- Creating a culture of conducting evaluations and using their results by institutionalizing the process and agreeing on principles and valuation among involved parties will enormously contribute to the effective use of evaluations;
- An evaluation's independence and credibility will facilitate its effective use; conflicts of interest in conducting evaluations and political interference will decrease an evaluation's value;
- Evaluations should be part of a system and clearly communicated to stakeholders from the beginning of a project or programme. This will avoid later surprise or hesitance by the parties that conduct and/or use an evaluation. This will also increase stakeholder and government ownership of the process, which is an important issue in development; and
- Decision makers should be involved in evaluation processes in order to ensure effective use, decision-making and action based on evaluation results.

2. AFGHANISTAN

POLITICAL PARTICIPATION FOR DEVELOPMENT EVALUATION IN SOUTH ASIA: AN AFGHANISTAN PERSPECTIVE

RANGINA KARGAR
Member of Parliament

ACHIEVING SUSTAINABLE SOCIO-ECONOMIC DEVELOPMENT IS ESSENTIAL FOR AN EFFECTIVE STATE

There are growing pressures today 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.

Brenda Johnson Padgitt says, "A truthful evaluation gives feedback for growth and success."

Development evaluation is an objective assurance and consulting activity that adds value and improves ongoing operations. It helps to accomplish objectives by bringing a systematic, disciplined approach to evaluate and improve the effectiveness of risk management, control and governance processes. Validity, reliability, comparability and fairness are not just measurement issues, but social values that have meaning and force outside of measurement wherever evaluative judgements and decisions are made.

The M&E key institutions in Afghanistan are the Supreme Audit Office (SAO), the Ministry of Finance, the Parliament and the High Office of Anti-Corruption.

Supreme Audit Office

The Supreme Audit Office is the supreme audit institution of Afghanistan. It reports directly to his Excellency the President of Afghanistan in an independent and impartial way.

The main objectives of SAO are to:

- Protect public funds and take action against errors and irregularities regarding financial carelessness and the misuse of public property;
- Prevent illegal expenditures;
- Review the systems of control over government receipts and payments;
- Identify fraud and ensure that accused individuals are brought to justice;
- Certify the financial statements of government;
- Identify shortfalls in the government budget and provide guidance to minimize errors, shortcomings, overpayments and poor estimation for construction projects; and
- Guarantee the accuracy of aid and grants provided by donor countries.

Because of political changes in Afghanistan during recent years, SAO's membership in several international auditing organizations had been suspended. Fortunately, SAO has now regained its membership in the Asian Organization of Supreme Audit Institutions, the Economic Co-operation Organization Supreme Audit Institutions, and the International Organisation of Supreme Audit Institutions.

The Auditor General is SAO's most senior official (equivalent of a minister). He establishes policies and guidelines for SAO and is directly responsible to the President of Afghanistan. The Auditor General is assisted by two deputy auditors general, who are responsible for implementing SAO's audit plans and evaluating its audit findings. The deputy auditors also consider proposals to improve SAO's work processes.

SAO's rules and regulations are based on international standards, and its objectives encompass financial, accounting and economic monitoring of institutions such as ministries, public offices and organizations, government commissions, municipalities and banks.

SAO's audit work is intended to ensure:

- Transparency and accountability in accounting;
- Effective performance;
- Standardized audits that comply with international auditing standards;
- Protection of property (goods and cash);
- Reliable accounting and reporting;
- Proper implementation of the government's socio-economic development plan;
- Rules are followed and community and current needs for special sectors are met; and
- Accuracy in the expenditure of aid provided by donor countries.

SAO's methodology for monitoring and auditing is based on the International Organization of Supreme Audit Institutions' international auditing standards.

To achieve its objectives, SAO is committed to auditing more than 670 budgetary units each year.

Civil war in Afghanistan has damaged the core of SAO (and other institutions) and diminished opportunities for effective and continuous training. To overcome this problem and rebuild SAO staff capacities, the Transitional Government of Afghanistan has awarded a project to PKF, an international firm of accountants and auditors. This project is funded by the World Bank and provides technical support to SAO in order to enable it to carry out audit functions in accordance with international auditing standards.

Ministry of Finance (Internal Audit, Investigation and Evaluation General Directorate)

Internal auditing is an independent, objective assurance and consulting activity designed to add value and improve Ministry of Finance operations. It helps the ministry accomplish its objectives by bringing a systematic, disciplined approach to evaluation and by improving the effectiveness of risk management, control and governance process. The mission of the Internal Audit, Investigation and Evaluation General Directorate is to:

- Provide effective internal oversight services in order to enhance economy in acquiring resources, efficiency in mobilizing the resources and effectiveness in meeting the goals and objectives of the Ministry of Finance; and
- Assist the Ministry of Finance and audit committee in the effective discharge of their responsibilities by providing analysis, recommendations, counsel and information concerning activities audited.

The directorate's objectives are:

- Ascertaining the level of compliance with established government laws, policies, procedures and plans (including behavioural and ethical expectations);
- Reviewing and appraising the soundness, adequacy and application of accounting, financial and other controls in place to achieve goals;
- Ensuring the economical, effective and efficient use of Ministry of Finance resources;
- Ascertaining the effectiveness with which Ministry of Finance assets are accounted for and safeguarded;
- Conducting special investigations; and
- Ensuring the integrity and reliability of financial and operational information produced by the Ministry of Finance departments.

The key role of the Internal Audit programme is to develop and implement audit procedures in line with government policy.

Article 61 of the Public Finance and Expenditure Management Law states: "The Ministry of Finance shall establish an Internal Audit administration and appoint auditors to audit the financial and accounting affairs of all state administrations. The auditors referenced in paragraph one (above) may require the relevant administrations to provide all information needed for auditing financial affairs."

Parliament

The following articles of the Afghanistan Constitution authorize Members of Parliament for monitoring and evaluation issues.

Article 89

The House of People shall have the authority to establish a special commission, on the proposal of one third of its members, to review as well as investigate the actions of the government. The composition and method of operation of the aforementioned commission shall be determined by the Regulations on Internal Duties.

Article 90

The National Assembly shall have the following duties:

1. Ratification, modification or abrogation of laws or legislative decrees;
2. Approval of social, cultural, economic as well as technological development programmes;
3. Approval of the state budget as well as permission to obtain or grant loans;
4. Creation, modification and or abrogation of administrative units;
5. Ratification of international treaties and agreements, or abrogation of membership of Afghanistan in them;
6. Other authorities enshrined in this Constitution.

Article 91

The House of People shall have the following special authorities:

1. Decide about elucidation session from each Minister in accordance with Article;
2. Decide on the development programmes as well as the state budget.

Article 92

The House of People, on the proposal of twenty percent of all its members, shall make inquiries from each Minister. If the explanations given are not satisfactory, the House of People shall consider the issue of a no-confidence vote. The no-confidence vote on a Minister shall be explicit, direct, as well as based on convincing reasons. The vote shall be approved by the majority of all members of the House of People.

High Office of Anti-Corruption

Taking into consideration the importance of combating corruption for stability and the sustainable development of Afghanistan, and recognizing the inability of existing institutions to deliver, H.E. President Karzai in July 2008 issued a decree establishing a High Office for Oversight and Anti-corruption. This law has been enacted in light of the provisions of

Article 7, item 3 of Article 75 and Article 142 of the Afghanistan Constitution and in accordance with the United Nations Convention against Corruption in order to oversee and coordinate the implementation of the Anti-corruption Strategy. The creation of this Office has fulfilled the requirement of Article 6 of the UN Convention against Corruption.

The High Office of Oversight and Anti-corruption is the highest office for the coordination and monitoring of the implementation of the Anti-corruption Strategy and for the implementation of administrative procedural reform in the country. This office is independent in carrying out its duties and reporting to the president.

CURRENT SITUATION

More than one-third of Afghanistan's people live below the international extreme poverty threshold of \$1 per day. Decades of war and civil strife have caused widespread human suffering, changed the country's social and political fabric and left formidable challenges to recovery. In 2002, the immediate needs were to restore economic stability, rebuild institutions and provide basic services at a time when hopes had been raised that armed conflict might be ending. As it turned out, the poor security situation resumed after a brief respite, and the hostilities have severely impeded the operations of Afghanistan's development partners and overall progress on socio-economic development. With significant support from the international community, Afghanistan achieved some progress over the decade from 2002 to 2013. For example, it now has an elected government, a great deal of essential social and economic infrastructure has been built or reconstructed, and a large share of rural households enjoys access to schools, health services, water and irrigation facilities.

Nonetheless, based on indicators—per capita gross domestic product, poverty, child nutrition, primary school enrolment and literacy rates—Afghanistan remains one of the world's least developed countries. Despite attention to the problem, gender inequity remains an enormous challenge in all walks of life. The fragile nature of the country's natural environment, if not preserved carefully, poses considerable risks to its future well-being.

KEY ACHIEVEMENTS

Operations in Afghanistan started and have continued under highly uncertain, extremely difficult and risky conditions. On average, people are living much longer; more children have access to education and everyone to health care; the economy has grown substantially; an independent media is in place; and women have made exceptional advances. The country can now make use of a network of rehabilitated and improved roads that allow more travel in less time. Of the seven intended airports, five have been built with good outcomes. A new 75-km rail line was completed ahead of schedule. Electricity is now available almost around the clock, compared to about 4 hours a day in 2002 in many provinces. Telecommunications companies performed the country's expansion plans and reached its subscriber target faster than originally anticipated.

CRITICAL GAPS AND CHALLENGES

High-level international efforts to support Afghanistan have influenced the operations of all development partners. Despite agreements on levels of support, the approaches to be taken and the delineation of tasks, knowledge, human resources and procedures of Afghanistan's institutions are still not up to the mark. This lack of absorptive capacity has revealed itself in slow implementation of projects, frequent recourse to foreign consultants, low sustainability and market distortions.

In addition, Afghanistan's budgetary challenges will be increased by the spending required to take charge of the country's security after the withdrawal of foreign troops in 2014. It is vital that all aid agencies focus on sustainable development during this transitional period. Some new initiatives are being planned by the government, particularly in the infrastructure sectors, to introduce funding for operations and maintenance; however, it is too early to discuss the feasibility of these measures.

The government will need grants for a long time for development investment and recurring costs, which will now also include greater amounts for security. The International Monetary Fund and the World Bank have estimated that the country is likely to achieve partial fiscal sustainability, defined as domestic revenues covering operating expenditures, but not until 2024–2025. The financial squeeze likely to occur in the intervening period may cause further security and sustainability issues.

RECOMMENDATIONS

It will be important to carefully assess the security situation, financial and fiscal sustainability, and institutional capacity while coordinating with key development partners to make complementary efforts. Future programmes should continue to focus on infrastructure and capacity development in the government; sectoral strategies should be based on analyses of industrial and agricultural demand and the population's socio-economic needs.

Developing and closely monitoring a long-term capacity development and governance improvement plan carefully tailored to Afghanistan's unique circumstances is required. The current government's short history, weak cooperation between central and local governments, and public-sector staff with limited education levels and low salaries require that capacity development efforts be shaped to the country's specific circumstances and challenges. Further, the country context requires better coordination by development partners. Operations need to explicitly strengthen governance in order to maximize the effectiveness of development support.

3. CAMBODIA

EVALUATION OF DEVELOPMENT ACTIVITIES AT THE MACRO LEVEL: CHALLENGES AND EXPERIENCES IN CAMBODIA

THENG PAGNATHUN

Director General, Ministry of Planning
Royal Government of Cambodia

MEASURING RESULTS IS PARAMOUNT

*If you do not measure results, you cannot tell success from failure;
If you cannot see success, you cannot reward it;
If you cannot reward success, you are probably rewarding failure;
If you cannot see success, you cannot learn from it;
If you cannot recognize failure, you cannot correct it;
If you can demonstrate results, you can win public support.*

STATEMENT OF THE PROBLEM

An important source of evidence of achievements, evaluation should be a systematic and impartial assessment of a project, programme, sector or development process at the aggregate level; it should quantify accomplishments by examining the results chain, contextual factors and causality to assess outcomes; it should aim to determine interventions' relevance, effectiveness, efficiency and sustainability; and it should contribute to knowledge-building and organizational learning. Evaluation is not independent of monitoring, a process of tracking or measuring performance more frequently, and auditing, which addresses financial accountability. Evaluation is necessarily **evidence-based**. This is the Cambodian Government's position.

This paper is **not** about evolving yet another definition of evaluation. Instead, it examines various nuances that Cambodia (or any small developing country) faces in conducting M&E exercises for periodic reporting to senior management in the government and in

development partner organizations. More specifically, the concerns presented are:

- What should be the structure of a results framework that is applicable to macro-level evaluation, a reporting level that is fundamental to national planning?
- How to generate the necessary data required for populating the results framework forms?

These concerns form a part of the debate within the Ministry of Planning while it prepares an M&E framework for the next National 5-year Development Plan 2014–2018. It is believed that other least-developed countries face similar challenges, and that there is a need to address these challenges in a practical and meaningful manner.

COUNTRY CONTEXT

Cambodia, classified as a least-developed country, had a per capita income of about \$931 in 2011. It went through war and turmoil between the early 1970s and the mid- to late 1990s, when the country's institutions of governance, infrastructure, human capital, social fabric and more were severely damaged. Only in the new millennium were substantial activities relating to socio-economic development taken up.

Cambodia pursues a development strategy through planned development in a market framework. Throughout the last 17 years, the Cambodian economy has grown at an average rate of 7 to 8 percent annually; if the 2009 downturn is deleted from the trend, the rate is yet higher. In the recent four to five years, the global economic downturn and commodity/petroleum-led inflation have resulted in global economic turmoil, leaving few countries unaffected. The Cambodian economy, however, has shown the capacity to bounce back after facing a severe setback, though the turmoil has left its scars: many development programmes have had to be rescheduled or staggered.

International development assistance plays an important role in funding and providing technical assistance. In 2011, international assistance was estimated at about 8.5 percent of the gross domestic product. More than 60 percent of total developmental expenditure in the public sector (typically, agricultural extension, irrigation, education, health and infrastructure) is funded by grants and soft loans from development partners. In this regard, a strong M&E framework takes on increased importance.

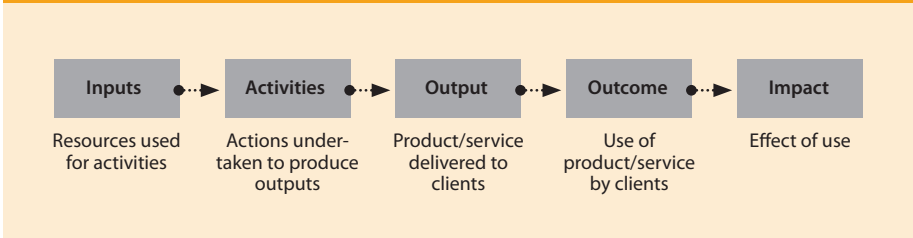
THE RESULTS FRAMEWORK FOR M&E

The conventional wisdom

The government adopted a results framework to account for and evaluate all developmental activities in order to provide vital direction at different stages in project/programme management. This allows senior government staff to answer four key questions:

1. Is the plan being effective?
2. How does one know whether one is on the right path?

FIGURE 1: BASIC COMPONENTS OF A RESULTS CHAIN



3. If not, where is the deviation?
4. How does one use this information continuously for regular corrective action?

The framework assumes that expenditures need tracking in order to ensure that they get translated into outcomes. Consider the education sector, for example:

- The school should have actually been constructed, as per specifications, and on time, and have its hardware, trained and qualified personnel, teaching curricula, etc. (**activity/output**)
- Additionally, the school attracts children from the catchment-area (**outcome**)
- Finally, children attend schools, pass and become literate and educated, and the society moves towards becoming more productive, more jobs are created, poverty reduces, etc. (**impact**)

Generally, development partners have endorsed this framework. A more generalized form of a results framework can be seen in Figure 1.

Practical considerations in using the results framework at the macro level

Measuring outcomes and impacts

In practice, there are issues in identifying and measuring variables at different stages in the results framework. In national planning, impacts (and in most cases, outcomes) are macro-level phenomena, while interventions could be policy or launching projects. Three examples illustrate the macro-micro disjoint.

Case 1: Poverty

Most development projects mention the outcome to be 'poverty reduction', irrespective of whether the projects are of infrastructure, potable water, health, sanitation or education.

- Poverty, a macro-level phenomenon, is reduced by a number of factors that range from individual attributes to performance of the economy. It is difficult to link

poverty reduction with one intervention; a full identification of the influencing factors is required. The results framework is thus required to be made significantly more complicated.

- Next, the relationship between poverty alleviation and its determinants could change in time. A typical case is of potable water, which, when provided, improves health, which in turn is expected to improve educational attainments and incomes. However, after everyone has been provided with clean water, other factors influence poverty reduction, not safe water *per se*.
- Finally, as outcomes (poverty, inequality) are macro-level variables, they are not easy to link to a results framework drawn-up for individual projects or sectoral programmes: the typical macro-micro disjoint.

Case 2: Agriculture

Consider crop yield rate in the agricultural sector. Government inputs, such as agricultural extension work and irrigation facilities, certainly help. But yield rates also increase because farmers adopt scientific agricultural practices, seeing profits. When a new scientific method emerges, farmers learn and then draw upon services that an emerging group of private providers offer, often because government services do not reach everywhere. For example, farmers sink irrigation wells independently of government-created large dams. In many western areas of the country, yield rates began to increase much earlier than dams were completed, thus warping causality.

Case 3: Education

In school education, the government provides buildings, teachers and other means that have a definitive impact on the enrolment rate. However, in Cambodia it is seen that children are sent to schools not only because the government has built schools, but also because parents want their offspring to be educated. If government schools are not available, the private sector would provide services. Recent reports on school enrolments suggest that in Cambodia, even poor households prefer expensive private schools, because children get some quality education there, often scarce in government schools.

To empirically validate this point, a regression equation was estimated to determine whether the supply- or demand-side variables are more important in explaining the portion of children aged 6-17 years in schools, using village-level data for all villages in Cambodia for which data was available. The estimates (given in Table 1) amply demonstrate that supply-side variables, namely the distance of government schools (primary, junior secondary or senior secondary) from villages, are statistically insignificant in explaining school attendance. Instead, it is the demand side variables that are statistically significant (measured by level of affluence, for which the proxy variables include: possession of assets such as motor-bikes and cycles, not having to live in thatch houses, accessing potable drinking water and having access to sanitary latrines).

There is evidence that a conventional results framework could be flawed if it is applied in

TABLE 1: REGRESSION RESULTS EXPLAINING SCHOOL ATTENDANCE IN A CROSS-SECTION OF 11,882 VILLAGES IN CAMBODIA, 2011

DEPENDENT VARIABLE: PERCENT CHILDREN 6 TO 17 YEARS IN SCHOOL	COEFFICIENT	T-STATISTIC	SIGNIFICANCE
(Constant)	73.073	99.406	0.000
Distance primary school (Km)	.002	.302	.763
Distance Junior secondary school (Km)	-.002	-1.127	.260
Distance Senior secondary school (Km)	-.002	-.997	.319
Wet season paddy yield	1.290	5.850	.000
Distance of village to province town	-.035	-6.528	.000
[(Total motorcycles)/(Total families)]X100	.011	1.798	.072
[(Total cycles)/(Total families)]X100	.055	15.664	.000
Percent families living in thatch houses to total houses	-.138	-15.785	.000
[(Number of toilets)/(Total families)]X100	.085	14.778	.000
[(Number of families accessing clean water)/(Total families)]X100	.051	14.257	.000
Percent families farming less than 1 ha land	-.003	-2.683	.007
R2 = 0.376; F = 177.40; n= 11,882			

its generic form to macro-level evaluations. There is need to bring about significant sophistication in the models. Further, one type of model will not fit all situations.

SPECIFICATION PROBLEMS

A second type of problem arises when outputs stand for outcomes (i.e. the output variable, 'increased visits to health centres', is often taken to depict outcomes). This is because outcomes of health interventions are slow and captured in large surveys carried out once in five to ten years. The only tangible indicator for an annual reporting is the number of visits (or other process variables, such as malaria/tuberculosis cases treated). In a conventional results framework, the following hazards emerge:

- Health outputs taken for outcomes could result in moral hazard (i.e. overuse, such as counting the free visits of patients to health centres); and
- Conversely, there could be an underutilization of outputs owing to reasons such as location, access by users, etc. (schools constructed, but children do not go).

Once again, the results framework is required to be sufficiently flexible to accommodate such nuances.

These cases do not deny the importance of a conventional results framework, but make the case for a more realistic framework that can become an evaluation tool rather than being a 'one size fits all' wand.

Alternatives have been suggested, such as constructing approaches based on a theory of change. Putting these into practice, though, is not easy. The models are extremely complex, requiring a great deal of data and resources that are simply not there. Additionally, there are issues of human capacities; typical government officials do not possess the skills needed to construct or interpret complex models.

ALTERNATIVES IN RESULTS FRAMEWORK FOR EVALUATION IN CAMBODIA

Recognizing the issues discussed above, the government has adopted a results framework that matches performance to a stipulated target at the beginning of a five-year plan. These are mainly outcome/impact indicators, though a few output indicators also appear in the list. At the sectoral level though, the government encourages use of the conventional results frameworks with whatever modifications sectoral authorities/ministries prefer to make. At the programme and project levels, it is mandatory for the relevant managers to deploy a near-conventional results framework.

The Ministry of Planning, in consultation with other line ministries and agencies, has identified some 64 core M&E indicators and some 125 auxiliary M&E indicators to assess the annual progress made in the economy (e.g. growth, inflation, trade and balance of payments, debt, government budget), the Millennium Development Goals and other key sectors (e.g. external financial assistance, employment, transport, infrastructure). While the core indicators are mainly multisectoral and cross-cutting (e.g. poverty, growth or child health), the auxiliary indicators are mainly sectoral. The list will expand, depending upon various line departments' needs and requests. Consultations are continuing with different stakeholders. The indicator list includes all the identified Millennium Development Goal Indicators.

The core indicators are divided into five categories:

1. Aggregate outcome indicators (e.g. gross domestic product [GDP], poverty, inequality, inflation);
2. Aggregate output indicators that stand for outcomes (e.g. balance of payments, import/export, structure of GDP and workforce);
3. Sectoral outcome indicators (e.g. infant mortality rate, maternal mortality rate, school completion rates);
4. Sectoral output indicators that stand for outcomes (e.g. crop yield rates, area under crops, roads made, attended births, enrollment rates); and
5. Proxy indicators (e.g. governance, inclusive growth).

Classification of auxiliary indicators is more complex; the indicators are mainly sector-specific and serve the needs of stakeholders in the concerned sectors. The indicators are a mixture of

outcome, output and process. Even non-governmental organizations draw up their own list of indicators. Additionally, the extent of standardization has been more limited than in the core monitoring indicators. This aspect needs strengthening.

REPORTING

The Ministry of Planning used to bring out macro-level evaluation reports of development activities only every three to five years. In 2010 and 2011, it brought out two successive annual Millennium Development Goals Progress Reports. Starting in 2012, the reporting became annual, and was for both the National Five-Year Plan and the Millennium Development Goals. This should remain the case for the 2014–2018 cycle as well.

At present, only the aggregate country-level indicators are being reported upon. Recognizing regional disparity as concern; from 2013 onwards some province-level disaggregated data will be presented (particularly on the indicators developed from administrative statistics). Efforts are also being made to bring in some unofficial project data into the official statistics framework.

Finally, in the last few years, the government, under its subnational governance programme, has begun to collect administrative data at the village and commune levels. Although validity has yet to be established, it is hoped this data will eventually mainstream and strengthen the database.

DATA-RELATED ISSUES

An evaluation exercise requires the right, high-quality data. In most less developed countries, this is the Achilles' heel. Sample or census surveys are conducted with external funds and expertise by agencies that both determine the data generation process and define the variables. To a limited extent, government officials have begun to participate in the exercises. As a result, continuity of surveys and comparability of definitions across surveys (and also the same surveys over time) are not guaranteed. To address this, Cambodia has established a Standing National Working Group on M&E, which is in the process of standardizing the definitions of variables and indicators. The government believes that the recommendations of this committee will bring some significant changes in data systems and, consequently, in the M&E system.

It is now recognized that efforts should be made to strengthen administrative statistics in ministries and departments; because of their broader coverage, they could be more regular and less expensive. The United Nations also recommends strengthening administrative statistics. In Cambodia, almost all ministries collect administrative data, but data quality can be questionable. Among the reasons:

- Central offices are somewhat better staffed than provincial and district offices. At the district level, there is only one official belonging to the Planning Department, and s/he has neither a support staff nor the resources to scientifically collect, collate and present data. S/he relies on village and commune chiefs (who are elected

representatives and, as such, have no formal training in statistics, data management or, for that matter, in any field) to collect data by means of a village book or a commune book;

- Some ministries collect data based on rather small and not necessarily scientific surveys;
- In a few ministries, administrative data is not collected every year for want of adequate resources; instead, linear projections from past data are made;
- Many ministries' data management facilities are weak on many counts (e.g. knowledge of basic statistics, availability of computers and data-storage devices). The situation progressively deteriorates from central to province to district levels. Some communes/villages have no electricity, so no devices work there; and
- Although commune and village chiefs and other local representatives are being trained to an extent, challenges include such stakeholders' weak initial exposure, their other chores and obligations (they are not dedicated staff), and that they could change every five years due to elections.

Development partners make a great deal of effort to draw up forms for measuring success, but leave aspects related to populating these forms with quality data to national governments. This is particularly true for macro/sectoral data.

APPROACH

A key question is how to strengthen M&E statistics? The Ministry of Planning, in its effort to make M&E more useful, is training government staff in line ministries to generate and interpret data, manage Administrative Data Collection Systems (selectively) and conduct small-sample studies/case studies. A standard training module has been developed for this purpose. Several batches of personnel in seven ministries have been trained. Once staff at the centre are trained, it is believed that this knowledge would then be passed on to staff at the provincial and district levels.

The M&E Working Group is central to making the necessary improvements. Following the success of defining a new poverty line in 2012 (a first for the government; previous exercises were conducted by the World Bank), the Ministry of Planning is following the approach of engaging officials from line ministries in task forces. At the central level, it has been possible to change certain definitions, alter/improve the data-collection process (at least in surveys conducted by the Ministry of Planning), and open doors in other ministries for engaging in discussions on their data-collection processes (e.g. agriculture, rural water supply and gender). So far, these processes have been extremely compartmentalized (each ministry is still extremely compartmentalized even now), but a beginning has been made.

Has the government been successful in this effort? It is too early to answer, but preliminary indications suggest that there is positive progress. At the central level, officials are able to devise progress indicators and have taken the initiative to launch a few surveys. It is not

yet evident whether annual reporting has helped, but the dissemination process certainly has; not only are meetings well-attended, there are queries made afterwards as well. At the province and district levels, early results suggest that to a limited extent, officials are able to interpret data and results.

Another question concerns resources. Most governments believe that collecting statistics should be a lower priority than launching 'mega' projects. The Ministry of Planning is making efforts to request earmarking at least three to four percent of development resources towards M&E work.

CONCLUSION

This short paper put together some practical problems and challenges that least-developed countries like Cambodia face in conducting evaluations. This paper, as such, does not delve into new approaches or definitions; instead, it examined two practical aspects that most governments face:

- Applicability of the results framework in its conventional format to evaluate gains of development at the macro level; and
- Availability of quality data at different levels and problems in generating it.

While no definitive answers were provided, the paper outlined the approaches the government is following in Cambodia.

4. INDONESIA

MONITORING AND EVALUATION SYSTEM IN INDONESIA

ARIF HARYANA

Director, Development Performance Evaluation System and Reporting
National Development Planning Agency (BAPPENAS)

INTRODUCTION

As part of the public-sector reforms, Indonesia has enacted Law No.17/2003, Regarding the State Financing, and Law No.25/2004, Regarding the National Development Planning System. These laws provide a regulatory framework on implementing performance-based planning and budgeting, which demand performance information from M&E results. M&E has become an important part of the development management cycle (see Figure 1), because it provides information that presents accountability and provides inputs for improving future planning and budgeting periods. The government also enacted Regulation No.39/2006, which provides the legal basis for conducting M&E and mandates that the Ministry of National Development Planning/National Development Planning Agency (BAPPENAS) develop an M&E system. In order to implement the mandate, the Deputy of Development Performance Evaluation post was established within BAPPENAS in 2007.

The Deputy of Development Performance Evaluation has three substantive roles: formulating and coordinating national development M&E; monitoring and assessing the annual plan and medium-term (five-year) plan; and maintaining national development M&E partnerships.

The enactment of a regulatory framework and institutional establishment for national M&E shows the Government of Indonesia's interest in and commitment to M&E. However, there are still many challenges in operational and policy-level implementation.

The purpose of this paper is to share information and lessons learned from the Indonesian evaluation system and its implementation challenges.

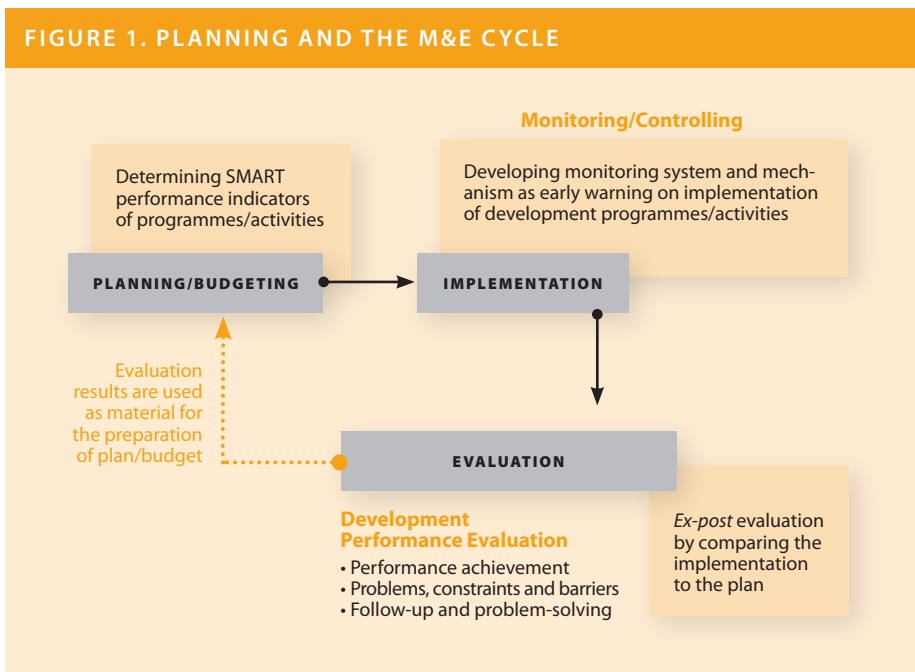
THE M&E SYSTEM IN INDONESIA

Law No.25/2005 describes three time horizons of national development planning: long-term plan (RPJPN, 2005–2025), medium-term plan (RPJMN, five years; the current RPJMN covers 2010–2014), and annual plan (RKP). However, GR 39/2006 only mandates conducting M&E on the RKP and RPJMN. Both RKP and RPJMN contain line ministry-implemented programmes and activities. All programmes conducted by all ministries in all provinces are subject to being monitored and evaluated.

E-MONEY FOR MORE TIMELY MONITORING

The focus of monitoring RKP implementation is to review the progress of projects and activities. Monitoring is conducted quarterly, measures achievement of targeted outputs at year-end and also identifies the constraints and bottlenecks in implementation (if they exist). This information is collected from reports that are submitted quarterly by line ministries and provincial governments. Coordination meetings with line ministries and provincial governments are conducted at least twice a year to discuss progress and to find ways to solve problems and relieve constraints. Information from this mechanism is used as inputs for future planning processes.

There were several problems and constraints in preparing the quarterly monitoring reports, including low reporting rates and lack of reporting promptness (primarily due to



difficulties in collecting and integrating M&E information from programme implementers). Only 30 percent of line ministries submitted reports in 2010; the reporting rate from provincial governments was less than 20 percent. This was not a good input for BAPPENAS to analyse national development performance towards performance-based planning and budgeting. Another cause of low reporting rates is the reluctance of the line ministries to submit the report—there was no punishment or reward for delivering the reports.

In order to improve reporting rates, BAPPENAS has developed an online application system, called e-Money. The application is aimed at assisting line ministries collect and integrate data and submit reports. The application has been implemented and the number of line ministries that submit reports has increased to 65 percent in the first quarter of 2013.

Another purpose of developing the e-Money system was to improve accountability. The system provides a publicly available page that presents basic information, including budget disbursements and programme/activity progress and performance. The e-Money system also provides traffic-light notification colours on the performance progress of every programme and activity. This indicator is based on the gap between achievement and set targets. The notification can be used as a warning for line ministries to improve the implementation of their programmes. For each programme and activity, the notification colour will be green (gap < 0 percent; the programme performance meets the target), yellow (gap < 25 percent; needs more attention and efforts to achieve its target) or red (large gap [gap ≥ 25 percent]; potentially cannot achieve its target by the end the year).

The eye-catching and easily understood traffic-light notifications may be a good way for BAPPENAS and line ministries to monitor the progress of all development programmes.

BAPPENAS keeps improving e-Money to help line ministries and local governments submit their M&E reports. BAPPENAS was developing e-Money for local governments in 2013.

Based on the information collected from quarterly monitoring data, BAPPENAS independently conducted an evaluation of the government's annual plans. Gap analyses were used in the evaluation, comparing the programme's achieved outputs to its targets. There was also a quadrants analysis that compared ministry performance to others.

EVALUATION OF THE NATIONAL MEDIUM-TERM PLAN (RPJMN)

Evaluation of RPJMN implementation is aimed at measuring the outcomes and impacts of programmes implemented during the five years of plan implementation. The GR No.39/2006 mandated conducting M&E on RPJMN at least once in the last year of the implementation period of the plan. The evaluation of each programme was conducted by line ministries as a self-evaluation, and the reports were submitted to BAPPENAS. These reports (and data from other sources) became the input for BAPPENAS to analyse and produce the integrated evaluation report.

In 2012, BAPPENAS conducted a mid-term evaluation on the current RPJMN, even though there was no obligation to do so. The purpose was to measure the progress of achievements on the outcome or impact of the development programmes, especially those programmes that are considered to be national priorities. The mid-term evaluation

FIGURE 2. EXAMPLE OF ANALYSIS FORMAT IN MID-TERM EVALUATION: HEALTH SECTOR

INDICATOR	INITIAL STATUS (2009)	ACHIEVEMENTS			TARGET	STATUS
		2010	2011	2012		
Life expectancy	70.7	70.9	71.1	71.1	72	2
Maternal mortality rate (per 100,000 live births)	228	n.a	n.a	n.a	118	3
Percentage of women giving birth attended by a skilled health worker	84.38	84.78	86.38	43.99	90	2
Infant mortality rate (per 1,000 live births)	34	26 ¹	26 ¹	32 ²	24	3
Coverage of first neonatal visit	80.6	84.01	90.51	33.59 ³	90	1
Prevalence of malnutrition among children under five	18.4	17.9	n.a	n.a	<15.0	2
Total fertility rate (TFR) (per woman of reproductive age)	2.6	2.4 ¹	n.a	2.6 ²	2.1	3

Notes: 1. Population census 2010. 2. 2012 Indonesia Demographic and Health Survey (Provisional Results). 3. Data through March 2012.

1 Achieved/on track/on trend 2 Need greater efforts to be achieved 3 Hard to be achieved

compared the outcomes and impacts achieved through mid-year 2012 with the planned or targeted outcomes and impacts (i.e. it used a gap analysis method). In addition, the mid-term evaluation reviewed the trends of achievement in the last three years. Based on both gap and trend analysis on key performance indicators, we can make a judgement of whether a programme has already achieved its target, is on track towards achieving its target or lags behind. Based on that judgement, colour notifications are attributed to each performance indicator: green (the target was already achieved or is on-track/on-trend to meet the target), yellow (needs more effort to achieve the target) or red (it will be difficult to achieve the target by the year-end of the RPJMN).

The eye-catching colour notifications make it easier for policymakers to grasp and understand the progress of development programmes. For programmes with yellow or red performance indicators, the mid-term evaluation identified the constraints and problems that existed and proposed solutions.

The evaluation conducted by BAPPENAS can be considered independent because BAPPENAS is a planning institution, not a ministry or institution that executes development programmes. Therefore, analytical biases due to conflicts of interest can be minimized. BAPPENAS is also free to use data from any independent credible sources, not only data

from line ministries' reports. Using credible data from independent sources (e.g. Statistics Indonesia [*Badan Pusat Statistik*]) and national and international independent institutions) ensures the evaluation's credibility and the quality of its inputs.

In practice, the information provided by the mid-term evaluation was very beneficial for policymakers, because it was conducted at the right time. Performance information from mid-term evaluation was available just as the process of planning and budgeting for the fiscal year 2014 began. The mid-term evaluation results were presented by the minister and discussed in a cabinet meeting chaired by the President. As a follow-up of the meeting, the President delivered a policy directive to all ministers regarding prioritizing programmes and activities. The President also instructed the three coordinating ministers to coordinate relevant ministries in taking necessary steps to improve the performance of the programmes with yellow or red indicators. This shows how important M&E is to the management cycle.

PUBLIC CONSULTATION FORUM FOR PLANNING AND M&E

The government should provide a public space to give society access to all public activities, including the planning and M&E practices. Article 2 of Law No. 25/2004 requires and provides space for public involvement in this area. The involvement is intended to capture people's aspirations and to increase the sense of stakeholder ownership. In practice, public involvement is accommodated in the planning process. Civil society organizations and academia have been playing an active role in the *Musrenbang* (a discussion forum in the planning process).

Even though there is no obligation to create a formal forum for civil society organizations to deliver their views during the M&E process, BAPPENAS has held public consultation meetings since 2012, involving participants from civil society organizations and universities. Relevant civil society organizations were invited to share and discuss their views regarding specific issues in these meetings. Members of academia and other independent parties acted as facilitators during the discussions. The forum increased the objectiveness of M&E results. It is expected that these meetings will improve the current M&E system and methodology.

CLOSING REMARKS

Indonesia has recognized the importance of M&E. Therefore, Indonesia committed to comprehensively implement it, both in regulation and in institutionalization. With this commitment, Indonesia enacted GR No. 39/2006 and established the Deputy of Development Performance Evaluation post within BAPPENAS in 2007. However, Indonesia faces challenges to improving its M&E system and capacities. The main challenge is building M&E practitioners' capacities. Consequently, Indonesia needs to broaden its networking and partnerships to share knowledge and adopt lessons learned from other countries.

5. SOUTH ASIA

WHY NATIONAL EVALUATION POLICIES MATTER IN SOUTH ASIA

HON. KABIR HASHIM
Member of Parliament

ASELA KALUGAMPITIYA
Parliamentarians Forum on
Development Evaluation in South Asia

INTRODUCTION

Developing and strengthening evaluation policy in South Asia is important for many reasons. South Asia has a high level of poverty. In addition, most of the countries in South Asia depend on donor funds and foreign loans for development projects. Many public-sector development projects do not achieve host countries' goals and objectives. Monitoring of public-sector projects is poor, and political intervention to ensure M&E implementation is limited; evaluation practitioners often blame legislators for not taking action. In this context, a group of committed parliamentarians organized efforts to establish national evaluation policies in their respective countries and to ensure transparency and accountability in public-sector development projects.

The practice of evaluation is yet to be fully institutionalized in South Asia. As compared to other regions, there are some countries that are at an equal or higher level and some countries at lower levels in terms of national level evaluation practices. India and Sri Lanka are good examples of countries where a stronger evaluation culture is being institutionalized. India is currently implementing the Performance Monitoring and Evaluation System, which covers 80 departments and 800 responsibility centres (in addition, 15 states have adopted the system, cutting across political lines). In Sri Lanka, there is a strong evaluation culture, with civil society participating in evaluation through the Sri Lanka Evaluation Association. At the public-sector level, the Sri Lanka Ministry of Planning has a national operations room

that links most of the ministries for monitoring of public-sector development projects. In addition, Sri Lanka was the first country in the region to develop a national evaluation policy. However, due to the lack of an enabling political environment, the national evaluation policy is yet to be legislated.

The South Asian subregion has a growing evaluation culture and strong civil society engagement. The Community of Evaluators is the main regional evaluation network.

There are country-level evaluation networks in many South Asian countries. The Sri Lanka Evaluation Association has the longest history as a country network in the region. The association work closely with the government of Sri Lanka to strengthen evaluation policy in the country and runs professional capacity-building workshops and international conferences. The Sri Lanka Evaluation Association is the first national evaluation body to have advocated for a national evaluation policy, which was brought up to a draft policy at the cabinet level. The Pakistan Evaluation Network also has many years of experience working with evaluation professionals and policymakers. The Network is working with parliamentarians towards national policy development. The Centres of Excellence in Afghanistan, Bangladesh and Nepal and the Development Evaluation Society of India are the other country-level evaluation networks.

Teaching Evaluation in South Asia is another initiative in South Asia to enhance professional development in the region. The Evaluation Conclave is now becoming a regular event that brings together evaluation professionals from around the world to share experiences. The Parliamentarians Forum on Development Evaluation complements this process, contributing value to South Asian evaluation policy development efforts.

The Parliamentarians Forum on Development Evaluation is focusing on attaining achievements in line with EvalYear 2015 goals. EvalYear will be a unique opportunity to focus on diverse approaches to evaluation. The forum will act as a catalyst for important conversations and thinking at international, regional, national and subnational levels on M&E's role in good governance for equitable and sustainable human development. Evaluation can enable leaders and civil society to develop and support better policies, implement them more effectively, safeguard the lives of people and promote well-being for all. Evaluation can enable leaders to consider social, political, technical and financial factors, and help leaders report to their constituents about the impact of their decisions and elicit feedback from every group. However, if evaluation is to play these roles, more dedicated efforts, systems change, capacity building and resources will be required.

The intention of EvalYear is to position evaluation in the policy arena by raising awareness of the importance of embedding M&E systems in the development and implementation of the forthcoming Sustainable Development Goals and all other critical local contextualized goals at the international and national levels. EvalYear is about taking mutual responsibility for policies and social action through greater understanding, transparency and constructive dialogue.

MAIN CONTENTS

None of the South Asian countries have a national evaluation policy in place, although each country has a fairly satisfactory M&E mechanism in its respective public sector. This issue has been discussed in many regional evaluation conferences, but it has not yet been possible to bring policymakers and evaluation practitioners together on the same stage. To address this challenge and create an environment to establish national evaluation policies in South Asian countries, a group of parliamentarians initiated the Parliamentarians Forum on Development Evaluation, a group committed to development evaluations in South Asian Association for Regional Cooperation countries. The forum's goals include advancing an enabling environment for nationally owned, transparent, systematic and standard development evaluation processes that ensure aid effectiveness, achievement of results and sustainable development in line with country-level national evaluation policies. The forum's objectives are:

- National evaluation policies endorsed by the respective South Asian governments are in place and effective;
- A space is created for dialogue between legislators and the evaluation community;
- Improved capacity of parliamentarians who are committed to development evaluation in the country; and
- Established country-level mechanisms that are in line with national evaluation policies, ensuring results-oriented and sustainable development.

The forum has conducted a panel on enabling an environment for development evaluation in Kathmandu, Nepal. A representation of all South Asian country parliamentarians is planned in Colombo, Sri Lanka, to coincide with the Sri Lanka Evaluation Association's international conference. The forum will conduct a panel on Why National Evaluation Policies Matter in South Asia.

The forum has planned several strategies to achieve its intended results. A mapping exercise will help to understand the current global situation, with specific information on South Asia. Based on the mapping's findings, strengths in countries with national evaluation policies and gaps in other countries can be identified. The mapping's findings will be presented to consultation delegates when the development of the model policy and country plans start. The consultation will allow policymakers, evaluation professionals and activists to come together and develop a product to promote the establishment of national evaluation policies. After the consultation, country teams will take the policy development process forward with expert technical support. The forum will work closely with governments and evaluation networks on inaugurating EvalYear at the country level.

The mapping exercise will map out the status of national evaluation policies, mechanisms and guidelines at the country level on a global scale, and will identify success stories in select countries. The mapping exercise will focus on detailed evaluation mechanisms in South Asian countries, which will be presented in a three-day regional consultation. The mapping exercise will be based on a desk review, Internet search, electronic media and communication. Based on the findings, the study will be extended to a comprehensive exercise in the long term.

Expected achievements: A mapping report outlining the status of national evaluation policies, mechanisms and guidelines at the country level, with special focus on South Asian countries. The report will include a table or chart that shows each country's status.

South Asia regional consultation with parliamentarians will develop a model national evaluation policy and country work plans to establish a policy.

Participation will include parliamentarians, representatives from national evaluation networks in each country, a representative from each regional evaluation network, representatives from the United Nations and other donors, stakeholders, EvalPartners and Parliamentarians Forum representatives, and experts from other regions.

Expected achievements: A draft model of national evaluation policy and the development of individual country work plans to lobby for the policy.

A small team of experts will provide technical support to country teams to adapt the model policy to their country contexts and to lobby for legislation.

Country teams include members of the Parliamentarians Forum on Development Evaluation, other interested policymakers, national evaluation network representatives, interested government officials and representatives from donors.

Participating countries are Afghanistan, Bangladesh, Bhutan, India, Pakistan, Maldives, Nepal and Sri Lanka. There is no cost involved for Sri Lanka, as two experts bear the cost (if there is any).

Expected achievements: Country teams receive necessary technical support to take the national evaluation policy forward.

Advocacy support will be provided to country teams to lobby legislation of the draft national evaluation policy. There is a need to identify who can do what regarding advocacy for lobbying national evaluation policies at the country level.

Advocate for the establishment of a national evaluation office hosted by the parliament, president or prime minister's office.

Expected achievements: A national evaluation office is established in at least one country in the five-year period.

Country-level inauguration of EvalYear 2015: Working with governments and evaluation networks to identify concrete activities and achievements for EvalYear (e.g. declaring evaluation week to get public attention and emphasize importance; decision to establish a national evaluation office hosted by the parliament, president or prime minister; development of evaluation guidelines in line with national evaluation policy; organizing public dialogues on national evaluation policy).

Expected achievements: EvalYear is used as an opportunity to get public attention and emphasize the importance of evaluation and evaluation policy.

CONCLUSIONS

There is a need for national evaluation policies in the region to promote the use of evaluation in development interventions in the public and private sectors.

Promote regional political champions to work with other stakeholders in taking forward policy development dialogues.

Parliamentarians' engagement in policy development processes at the country level is a needed and important element that should be supported by civil society.

REFERENCES

Eval Partners. 2013. "EvalYear Innovation Challenge announcement". Available at mymande.org/evalpartners/innovation_challenge. Last accessed 5 February 2014.

6. SRI LANKA

COUNTRY-LED NATIONAL EVALUATION SYSTEM: INDEPENDENCE, CREDIBILITY AND USE OF EVALUATIONS; CHALLENGES AND SOLUTIONS

VELAYUTHAN SIVAGNANASOTHY
Secretary, Ministry of Traditional Industries
and Small Enterprise Development

VAIDEHI ANUSHYANTHAN
Assistant Director, Department of
Project Management and Monitoring

INTRODUCTION

The Government of Sri Lanka has given high priority to ensuring value-for-money in public management. As a result, public management is giving very high focus to concepts such as managing for development results, development evaluation and performance audit.

The August 2012 circular issued by the Presidential Secretariat makes it mandatory for all line ministries to set out a medium-term strategic plan with well-defined outcome-based key performance indicators with baseline and medium-term targets. Further, to enable His Excellency the President to monitor and evaluate progress of the ministries, six outcome-based key performance indicators have been set for each ministry in the medium term. The achievements must be reported on key performance indicators through a scorecard with a traffic light signal-based dashboard. Although managing for development results has been institutionalized in the Sri Lankan public sector, it needs to be further broadened and deepened. National evaluation systems are being strengthened to support these expectations.

COUNTRY-LED NATIONAL EVALUATION SYSTEM

The Paris Declaration, endorsed in 2005 and followed by the Accra Agenda for Action in 2008, focused on aid effectiveness. The 2011 Busan Declaration shifted the focus from aid effectiveness to development effectiveness. Accordingly, there was a parallel shift from evaluating aid to evaluating development effectiveness, which entails a country-based national evaluation system.

In Sri Lanka, development policy and programme evaluations are undertaken by line ministries, the Department of Project Management and Monitoring, Auditor General's Department and development partners. Most of the evaluations during the early 1990s were driven by donors, were heavily confined to aid-funded projects, and were thus focused on accountability rather than learning. Following the Paris Declaration, there was a greater emphasis on country-owned, country-led and country-managed development efforts, and as such national ownership and leadership were recognized as the overarching factors for development outcomes. In this context, country-led national evaluation systems were recognized, supported and strengthened.

INSTITUTIONALIZING EVALUATION IN GOVERNMENT

In the early 1990s, on the request of Government of Sri Lanka, development partners such as the United Nations Development Program (UNDP) and the Asian Development Bank assisted the government to establish a country-owned, country-led, *ex-post* evaluation system.

Presently, at the central level, the Department of Project Management and Monitoring of the Ministry of Finance and Planning undertakes ongoing, *ex-post* and impact evaluations of selected mega projects. In selecting projects for evaluation, criteria such as policy relevance and usefulness, problematic implementation, innovative intervention, projects of a replicable nature and projects that may throw light on new or upcoming policy initiatives are given due consideration. In particular, when new programmes are formulated, the national planning authorities and line ministries like to know the outcomes of previous, similar programmes, including factors of success and failures. Under such circumstances, consideration is given to learning-based demands that significantly help improve the planning and design of new public programmes and policies (see papers published in the proceedings reports of the 2009 and the 2011 national evaluation capacities conferences, available on nec2013.org).

COMMISSIONING AND CONDUCTING AN EVALUATION

In addition to the central agency, the line ministries that are responsible for the overall execution of development programmes and public policy also selectively undertake evaluations. These line ministries' M&E units are responsible for commissioning and conducting evaluations. These units are independent from other management, operational and programme implementation-related functions and report directly to the Chief Accounting Officer, who is the head of the line ministry. For the purpose of routine evaluations, an independent expert evaluation team is appointed with representation from Department of Project Management and Monitoring and independent sector specialists, academia and research institutions.

Similarly, in the case of evaluations of a special nature, where capacity constraints exist, the study is outsourced to an independent academic or research institution.

In order to avoid a conflict of interest or undue pressure, checks and balances are in place, such as management groups, reference groups and peer review systems. Peer reviews are undertaken by specialists who are trained in evaluation both within and outside the government. The M&E unit is expected to manage and undertake evaluations impartially and free from bias. All evaluations require a terms of reference with a methodological framework and with draft evaluative questions based on the Organisation for Economic Co-Operation and Development's Development Assistance Committee (OECD/DAC) criteria of relevance, efficiency, effectiveness, impact and sustainability. The draft evaluation framework with evaluative questions and design matrix is validated at the scoping session with all concerned stakeholders. This process helps to improve independence and minimize conflict of interests or undue pressure.

REPORTING TO HIGH-LEVEL MANAGEMENT: MANAGEMENT RESPONSE

Evaluation findings are presented at a high-level progress review meeting chaired by a minister. Meeting invitees include senior officials from programme implementation agencies, representatives from the Department of Project Management and Monitoring, the Department of National Planning, the Department of National Budget, the Department of External Resources, the Auditor General's Department and other concerned stakeholders. The minutes of the meeting identify 'management responses' that need to be implemented as a follow-up. Further, the findings and lessons are expected to be integrated in the formulation of new public policies, programmes and projects.

The findings and key issues set out in the evaluation reports are taken into the Auditor General's report as part of the annual report and tabled to the Cabinet of Ministers, the Parliamentary Committee on Public Enterprise and the Committee on Public Accounts, and also discussed at the parliament. Therefore, evaluation findings and recommendations get reported to the executive branch and legislative branch of the state.

INDEPENDENCE AND IMPARTIALITY ARE FUNDAMENTAL PILLARS

In order to meet independence criteria, an evaluation has to be impartial throughout the process, from commissioning to conducting to dissemination and feedback. There are possible biases in conducting evaluations, such as spatial biases (e.g. neglect of peripheries), project biases (e.g. failure to look at non-project areas), persons biases (e.g. evaluation covers elite and not vulnerable population), diplomatic biases (e.g. failure to see bad conditions and based on itinerary prepared in advance), and professional biases (e.g. meeting only with well-educated stakeholders). Further, there are possible biases in undertaking evaluations, such as the perception that experts know best and failure to recognize indigenous technical knowledge. When evaluations are conducted by outsiders, they sometimes fail to understand the local context and fail to recognize the real problems. There is a need to conduct decolonized evaluations with the participation of local evaluators that have knowledge of the local context. The national evaluation system recognizes these issues and addresses them at the scoping session.

CREDIBILITY

In Sri Lanka, evaluations are undertaken by the line ministries, central agencies and development partners and are conducted by independent evaluation professionals, research institutions and academia, and in partnership with government officials who are independent from the planning, designing, implementing, managing and monitoring of the evaluated public policy or programme. The evaluation team should include evaluation experts and sector specialists. In selecting the evaluators, the competencies, skills (e.g. analytical and methodological skills, writing and communication skills, management skills) expertise, ethics (e.g. maintaining participants' privacy and confidentiality; sensitivities to gender, beliefs, manners and customs) and independence are given consideration.

The degree of independence from policy, operation and management functions and conflict of interest are examined to ensure independence. Ability to work freely without interference, integrity, honesty, openness and sensitivity to culture, customs, religious beliefs, values, gender and ethnicity are considered to be important qualifications for an evaluator. The credibility of evaluations is further strengthened with the setting up of a management group, a wider reference group and an independent peer review. Moreover, the evaluation team must be authorized and given access to all relevant information that is necessary to undertake the evaluation. The evaluator is recognized as a facilitator and negotiator of diverse options and issues, who arrives at a judgement that is not based on his or her own perceptions, but is based on evidence as seen by the concerned stakeholders. Similarly, findings are derived from logical analysis rather than an evaluator's own perceptions.

STAKEHOLDER CONSULTATION AND QUALITY ASSURANCE

Beneficiary feedback is obtained during the conduct of evaluations, including consultations with all concerned stakeholders; this process increases the credibility of the evaluation's findings. A variety of data-collection methods are used to triangulate and enhance the validity of the data and findings, which helps to improve credibility. Such methods include focus group discussions, community interviews, key informant interviews, intensive review and analysis of existing secondary information and documents, beneficiary surveys and opinion, and field observations.

Further, evaluations follow the OECD/DAC quality standards for evaluation. Standards are followed in the conduct of evaluations (process) as well as in the submission of reports (product). Stakeholder consultation meetings are held to validate the finding and the draft report.

In stakeholder workshops, stakeholders are given the opportunity to comment on findings and conclusions that make the process more transparent, balanced, impartial and credible. Public policies and programmes are rated as 'highly successful', 'successful', 'partly successful' and 'unsuccessful', an approach that enhances credibility.

USE OF EVALUATION

The utility of any evaluation is a prime criterion for judging its worth, regardless of its technical, practical and ethical merit. To have an impact and to ensure behavioural changes and action, evaluations need to be effectively disseminated and communicated.

Evaluations are done for accountability and learning. To have an impact on decision-making, an evaluation should be relevant to stakeholders' needs and interests and be presented in a clear and concise manner.

Evaluations conducted by the central agencies and line ministries cover various stages of a public policy or programme. *Ex ante* evaluations are conducted as soon as the programmes and projects are prepared in order to see whether the programmes and project are bankable and viable. Ongoing evaluations are conducted during implementation for mid-course corrections. *Ex post* and impact evaluations are conducted to assess outcomes, impacts and for learning.

DEMAND-DRIVEN, POLICY-RELEVANT EVALUATIONS

It is important to identify policymakers' information requirements and involve them at an evaluation's early stages, so that their needs will be taken care of, and so that the evaluation will be more demand-driven and responsive to policymaker needs.

Under the country-led evaluation system in Sri Lanka, public policies and programmes are largely selected for evaluation through a demand-driven process. As such, the findings of such evaluations are likely to be more effectively utilized than supply-driven evaluations.

READABLE AND TIMELY REPORT WITH ACTIONABLE RECOMMENDATIONS

In order to enhance the use of evaluation findings, evaluation reports should be clear, concise, readable and not technical. Reports should not be voluminous, similar to academic reports. The findings should be differentiated according to different audiences and users. The findings must be used and made available at a time appropriate for decision-making.

Evaluation should seek answers to all the evaluation questions. Results should follow clearly from the evaluation questions, and incorporate data analysis with clear lines of evidence. Findings should be based on logical analysis. Recommendations and lessons learned should be relevant and targeted to intended users. Evaluation recommendations should be actionable in order to ensure effective usage.

SYSTEMATIC DISSEMINATION

In Sri Lanka, evaluation reports contain an executive summary to give a synoptic overview of the findings and recommendations to policymakers. It is important to link evaluation findings to future activities for planners, policymakers and programme implementing agencies.

To be effective, an evaluation must also be responsive to the needs of different users (i.e. demand-driven and client-oriented), be tailored to different audiences and users; be timely, accessible, user-friendly; avoid information overload; and promote follow-up.

MECHANISM TO PROMOTE EVALUATION FEEDBACK

To ensure effective feedback, dissemination and institutional feedback mechanisms are important. Dissemination mechanisms should identify the means by which evaluation information is prepared and directed to client groups (e.g. abstracts and summaries, feedback seminars, evaluation reports, evaluation information system).

Institutional mechanisms are important to link evaluation units with other agencies (e.g. planning agencies, budgeting agencies and policymaking agencies). Evaluations should link to project concepts and project submission of the planning agency.

MANAGEMENT RESPONSE

In order to ensure implementation of evaluation recommendations, it is necessary to identify management responses whereby management agrees to implement a recommendation within a time frame and through a responsible focal point.

Evaluation reports must research a primary target group or key actors, such as planners, policymakers and donors who are expected to take action on findings. Similarly, evaluation findings must also reach a secondary target group that includes watchdog agencies, media and civil society organizations that will influence the key actors to act on the evaluation.

There are difficulties experienced in drawing inferences of a general nature from one-off project evaluations. Projects are clustered according to sectors in order to analyse common evaluation findings if these recur in several places, and to be able to draw broad conclusions. Such broad conclusions and cluster and sectoral findings are more effective to policymakers in the formulation of public policies.

TIMELY SUBMISSION OF EVALUATION FINDINGS TO MEET DEMAND

Evaluation adds value only if its findings are used. Evaluators must guide decision makers, policymakers and managers to do the right thing. It is important to determine what information is needed by decision makers and users, and to then formulate a terms of reference, evaluation questions and a design matrix that meets these needs.

Policymakers require information at appropriate times in order to cater to their planning and budgeting cycle. The synchronization of evaluation findings with the planning, budgeting and policymaking cycle is vital for the effective use of evaluations. Therefore, evaluators should be time-conscious and apply appropriate methodologies to ensure the timely availability of evaluation information. It is important, though, to find the appropriate balance between time and rigour.

ISSUES, CHALLENGES AND SOLUTIONS

Weak institutional linkages between evaluation and planning make feedback difficult

Evaluation and planning institutions seem to function in isolation and do not have effective formalized feedback arrangements to integrate lessons into the planning and design of new public policies and programmes. An effective institutional feedback mechanism should be established to strengthen institutional linkages.

Lack of demand for evaluation

It is necessary to create local demand for evaluation. Policymakers, planners and other stakeholders (e.g. media, civil society organizations) need to be sensitized.

Supply-side evaluation constraints

Supply-side evaluation constraints include lack of skills, methodological issues, data systems, manuals and guidelines, and national evaluation capacities. It is important to provide technical support in order to strengthen national evaluation capacities and to ensure that evaluations are independent, credible and impartial.

Inadequate evaluation information

Absence of country-level evaluation information on a website has been a challenge. Countries need to develop Web-based evaluation information systems on a sector-wide basis to identify findings, key issues and recommendations. Sector-level evaluation syntheses should be made available for effective policy feedback.

Joint evaluations to replace donor-driven evaluations

Donor-driven evaluations undermine the development of national ownership and country-led evaluation systems. It is important for donors to undertake joint and collaborative evaluations, which will help build national capacities, strengthen country-led evaluation systems and enhance the use of evaluation in decision-making.

Institutionalizing evaluations at the national and subnational levels

Evaluations are not fully institutionalized at the national and subnational levels. More funds and efforts need to be mobilized for monitoring. Awareness-raising and sensitization must be done on the importance of evaluation. Special budgetary provisions need to be made for evaluations. Institutional arrangements need to be strengthened and must report to the highest level of decision makers.

Project evaluations should expand to policy, sector and thematic evaluations

Project evaluations may not have demand at very high policymaking levels. Therefore, it is important to undertake sector synthesis, policy evaluations and thematic evaluations to enable wider policy-level usage of evaluation.

CONCLUSION

Recognizing these issues, Sri Lanka has taken actions to address them with technical assistance from UNDP, the Asian Development Bank and other development partners. UNDP has supported the institutionalization of managing for development results and evaluation in government. Independence, credibility, impartiality and effective use of evaluation are fundamental for successfully institutionalizing evaluation in government. Only such evaluations will be demanded by users and can contribute to development effectiveness.

Successfully institutionalizing evaluation requires that its findings and recommendations are closely linked to planning, budgeting and policymaking processes. It is necessary to create demand for evaluation among policymakers and other stakeholders.

The supply side of evaluation has to be strengthened through a supporting enabling environment, institution strengthening and individual capacity development.

Management responses are key to successfully implementing an evaluation's recommendations and to attaining effective feedback.

7. MALAYSIA

AN EVALUATION OF PUBLIC PERCEPTION TOWARDS GOVERNMENT PROJECTS: A CASE STUDY OF CONSTITUTION XYZ⁶⁸

MOHD KHIDIR BIN MAJID, Deputy Director
SHAHRAZAT BINTI HAJI AHMAD, Director

Outcome Evaluation Division
Implementation Coordination Unit
Prime Minister's Department

INTRODUCTION

The set-up of the Malaysian public administration consists of 24 line ministries, including the Prime Minister Department. On the highest rung, there are five central agencies: the Economic Planning Unit, the Public Service Department, the Malaysia Administrative Modernization and Management Planning Unit, the Treasury and the Implementation Coordination Unit. These central agencies are individually tasked with specific roles with respect to the planning, implementation and evaluation of national policies.

With respect to executing public projects, line ministries are responsible for implementing, monitoring and evaluating their own projects. However, the Implementation

68 For confidentiality reasons, the constitution is referred to as XYZ.

Coordination Unit is mandated as an independent monitor and evaluator. This structure encourages self-auditing by the project-owners (line ministries), and simultaneously infuses a check and balance by an independent party (the Implementation Coordination Unit).

Evaluation in the context of Malaysian development occurs at every level: project (by line ministries), programme and policy levels (by the Implementation Coordination Unit). In this respect, the traditional use of evaluation is for planning policies and budgets, and for improving financial resource allocations to programmes and public services. However, there are instances where a stakeholder conducts an evaluation to address a specific cross-cutting issue.

For example, the Implementation Coordination Unit was tasked to study the perception of government projects (physical and non-physical) implemented from 2005 to 2010 in one constitution. The objective was to evaluate public perceptions (among direct beneficiaries, projects implementers and local leaders) by identifying the following:

- What were the outputs and outcomes achieved during the period?
- Were there any grouses or dissatisfaction that arose from the public?
- Was there any misalignment between the demand and supply of public projects?
- Did it lead to negative impacts on the ruling government?
- What did the public want from the government?

STRUCTURE OF THE STUDY

The study used primary and secondary data to evaluate the outputs and outcome achievements of eight clusters (basic infrastructure, public infrastructure, worship, health, education, security, and recreation and welfare). Each cluster covered the physical aspects and activities of attached programmes. For primary data, a behavioural exercise was employed; a field survey based on stratified sampling was performed with two sets of questionnaires. In addition, a group interview was performed to gather additional information on certain issues.

CHALLENGES AND STRATEGIES USED IN THE EVALUATION PROCESS

A three-month time frame was given to complete the exercise. Coverage included 5,995 projects with an allocation of approximately \$234 million and a population of 72,504 people, with 4,096 direct beneficiary respondents and 128 executing agency respondents. From the perspective of the projects, it cross-cut through several agencies and issues (e.g. health, security, welfare, education, land, humanitarian and transportation). Complications included overlapping and redundancy of projects among agencies, which posed challenges in harmonizing data collection and aggregation. For example, there were extreme variations within groups (e.g. ethnic group, age, educational status, locality and income). Post-completion challenges included convincing stakeholders to take bold and quick actions and communicate findings to politicians, local leaders and implementing agencies.

Because of this diversity, a communication plan was important for follow-up and follow-through of remedial actions. In moderating the challenges, several strategies were adopted.

FIGURE 1. CITIZEN SATISFACTION INDEX

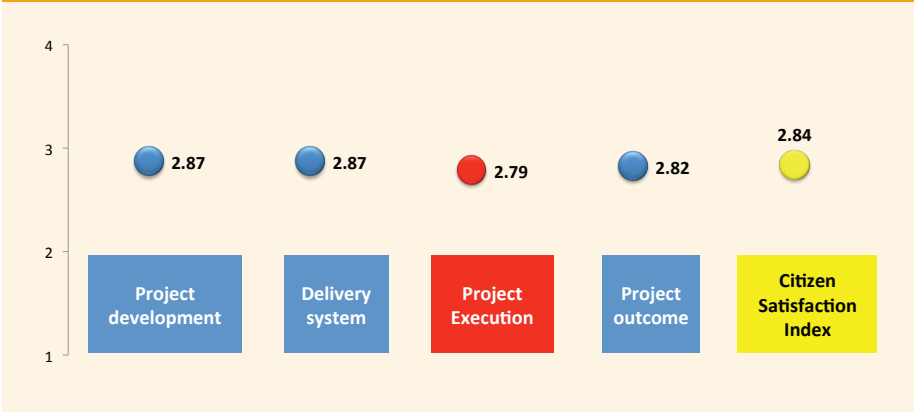
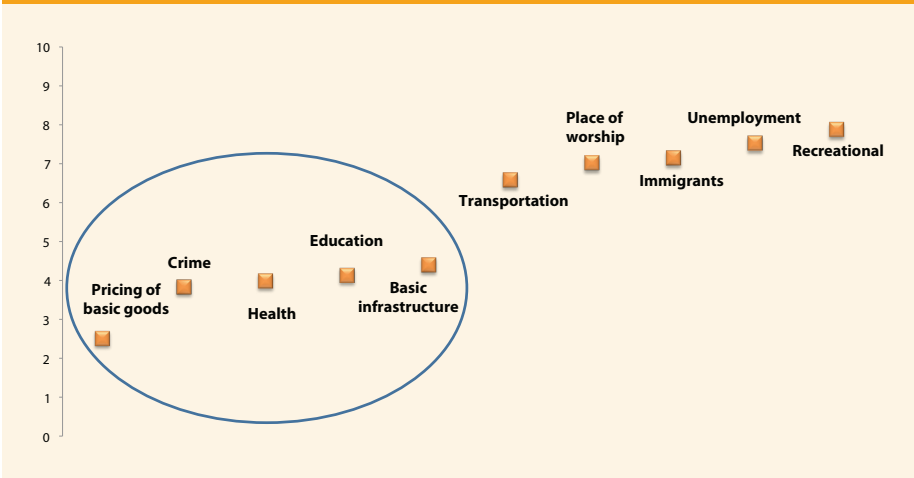


FIGURE 2. PUBLIC PRIORITIES



In this study, a task force comprising representatives from stakeholders' offices was crucial, and a strategic alliance was established with local leaders and local agencies.

MAJOR FINDINGS AND LESSONS LEARNED

The Citizen Satisfaction Index, which stood at 2.84 (on a scale of four) or 71 percent, showed a moderate level of satisfaction. The index is comprised of four elements (evaluation of development projects, delivery system, projects execution and outcome of projects) (see Figure 1).

It is plausible that the moderate Citizen Satisfaction Index resulted from a lack of public engagement, improper planning, sub-standard quality of output, a mismatch between public needs and stakeholders wants, a dialogue gap and a lack of public confidence in government procurement processes (which reflects on their integrity and transparency).

Despite heavy investment, public priorities focus on basic needs, which are public needs (e.g. cost of living, health, education and basic infrastructures) and safety issues (see Figure 2). This makes a good reference point for planning development programmes.

The findings provided many lessons learned, including:

- Implementing agencies must improve their project management planning;
- There is a pressing need for more engagement and dialogue sessions with the stakeholders;
- Immediate reforms are required on certain procedures and processes that slowed down development results;
- Implementing agencies must improve their management information systems to obtain, manage and utilize development information and development data for better planning;
- Implementing agencies must balance production capacities, manpower and resources; and
- Implementing agencies need to instil effective planning and monitoring.

BENEFITS OF THE STUDY

The findings and recommendations from the lessons learned were presented to stakeholders and implementing agencies, helping them re-evaluate and devise a more comprehensive inventory system. A model system was devised and is to be rolled out to the entire country. Identification of mismatched issues helped ease the gaps among leaders, agencies and the public, where public complaints are being prioritized. Additional budget and resources allocations have been deployed to mitigate critical issues. Political leaders also benefited from the study, where the information helped them focus on what the public wants, which led to higher public support.

THE WAY FORWARD

The study paved ways to a new approach in project management, where it prompted the Economic Planning Unit to develop the Rural Happiness Index and the Rural Socio-economic Indicator. A stronger collaboration among implementing agencies has been established, and an effective communication plan is underway to mitigate the misperceptions and misinterpretations among the parties involved. In idealizing a stronger collaboration among implementing agencies, a Blue Ocean Strategy was adopted. Finally, the Public Service Department is undergoing a revision and approval on additional posts for enforcement agency and health officers.

CONCLUSION

The study involved a multidimensional evaluation that examined achieving outputs and outcomes of development programmes in a specified time frame and location. This was followed by an analysis of public perceptions towards development programming, which were compared between two main players—programme beneficiaries and its implementing agency (which plays a dual-role as implementer and user). A pre-presentation was made to local implementing agencies and local leaders. Results were presented to the Secretary General and finally to the Deputy Prime Minister of Malaysia. Relevant parties are conducting a closely monitored post- evaluation to ensure follow-up and follow-through.

Selected to showcase the use of evaluation, the study also depicted the other two themes of the Third International Conference on National Evaluation Capacities (independence and credibility). The Implementation Coordination Unit is indeed independent and free from undue influence, and has full authority to submit reports directly to appropriate levels of decision-making. The Implementation Coordination Unit is credible, as it is mandated as an independent M&E system on behalf of the government and has had its own Outcome Evaluation Division since 2005.

REFERENCES

- Fornell, Claes. 2009. "The Cobalt Citizen Satisfaction Survey, Results for the State of Georgia". Cobalt Community Research: CFI Group International.
- Roch, Christine H., and Theodore H. Poster. 2006. "Citizen, Accountability and Service Satisfaction: The Influence of Expectation". *Urban Affairs Review* 14.

8. NEPAL

INSTITUTIONALIZATION AND USE OF EVALUATIONS IN THE PUBLIC SECTOR IN NEPAL

TEERTHA DHAKAL
Joint Secretary

National Planning Commission Secretariat

OVERVIEW

In Nepal, efforts to institutionalize evaluation of development interventions began in earnest in 1990, though the formal planning process had started in 1956. The National Planning Commission is the apex body that facilitates M&E in the country. The M&E system has been embedded in national planning processes and in all stages of project cycle management. Similarly, starting from July 2002, Nepal has been making efforts to institutionalize managing for development results approaches in its planning processes; the country has designed results frameworks and standardized results indicators at the sectoral and project levels. Since 1995, the National Planning Commission has conducted ongoing and post-completion evaluations of 29 projects in various sectors,⁶⁹ engaging third parties. This paper aims to briefly review the use of evidence generated from those evaluations in planning and decision-making processes.

LITERATURE REVIEW

The use of evaluations of development interventions depends on many things. As per Cracknell (2005), feedback from evaluations can be used at the project, programme, institutional, sectoral and policy or strategy levels. In addition, evaluations can be used in trainings and by beneficiaries and others outside the agency. However, as per Weiss (1999),

69 Available at: npc.gov.np.

policymakers rarely base new policies directly on evaluation results. Weiss gives two main reasons for the low use of such evidences in policymaking processes; competing pressures from interests, ideologies, other information and institutional constraints, and because many policies take shape over time through the actions of many officials in many offices, each of which does its job without conscious reflection (Weiss 1999).

An independent evaluation system that ensures report quality is critical to the effective use of evaluations. However, there is a need to strike a balance between independence and the internal relevance of evaluations. Gaarder and Briceno (2010) “want a system that is independent in order to achieve external credibility and social legitimacy, but not so independent that it loses its internal relevance.” Moreover, management responses to evaluation reports provide a useful basis to ensure the effective use of evaluations by addressing recommendations along with identifying responsibility and timing of implementation. Bamberger and Segone (2011) argue that management responses are a practical means to enhance the use of the evaluations to improve action. The writers also argue for proper dissemination of the report, identifying both direct and indirect users of the evaluation in order to ensure that the findings and conclusions are effectively utilized.

METHODOLOGY

A review of documents (including 29 evaluation reports, five medium-term plan documents and some policies) was conducted in order to assess the use of the evaluations in Nepal. In addition, key informant interviews were conducted with nine individuals (three from the National Planning Commission, two from the Ministry of Finance and four relevant officials from line ministries directly involved in development activities) in order to generate information on the use of evaluation findings. Due to the strict length limits, it is not possible to include key portions of the data and the analysis in this paper.

FINDINGS OF THE REVIEW OF THE USE OF EVALUATIONS

As an apex planning and M&E agency, the National Planning Commission facilitates evaluations, engaging third parties hired through competitive processes. Each year, some programmes or projects are selected for evaluation using specific criteria received from line ministries. Steering committees, formed for each evaluation to facilitate the process, approve the terms of reference, select the right evaluators, facilitate evaluation processes and maintain the quality of evaluations and reports.

From 1996 to 2012, the National Planning Commission conducted evaluations of 29 programmes/projects. Sixteen evaluations were from the agriculture sector, including the irrigation and land reform sub-sectors (nine evaluations were from the irrigation sub-sector). In addition, sectoral disaggregation showed that eight evaluations were from the social sector, three from infrastructure and two from the economic sector. Out of the 29 evaluations, only 60 percent had baseline data, whereas the rest of the projects used the recall method to create baselines.

The evaluation reports mostly focus on implementation processes and outputs delivered, rather than on the upper hierarchies of the results chain, especially the outcomes.

Methodologically, all studies reviewed in this paper were mostly quantitative in nature, though attempts were made to use some qualitative tools. Proper triangulations have not been done on the tools, either in the design of instruments or in interfacing the data in the analysis stages.

The National Planning Commission has established practices to disseminate evaluation findings, inviting a broad range of stakeholders (including policymakers from line ministries and relevant partners) and uploading the reports to websites.

While reviewing the evaluation reports and assessing their uses in the planning processes, it was found that the results have been used instrumentally or directly for several purposes. The recommendations have been used to formulate or refine policies and to make decisions regarding whether to continue, scale up, replicate or modify the project and its implementation modalities. In addition, results have been used in the discussions and decisions regarding annual programme and project budgets.

Moreover, the findings and results of ongoing evaluations of projects have been used for the amendment of programmes and in requesting more funds during the fiscal year. Line ministries have used evaluation results as evidence when responding to concerns that arose in legislative debates regarding budget allocation to sectors or projects. Further, evaluations of projects were documented and used to review and evaluate medium-term plans and relevant policies. However, none of the reports were used to ensure accountability of those engaged in implementation processes, even in cases where projects were not implemented well or were unlikely to realize the initially developed theory of change.

There are some explanations of the low use of evaluation recommendations in planning and decision-making processes. Evidence indicates that ownership and usage are higher among evaluations demanded by line ministries than those initiated by the National Planning Commission. Some evaluations' policy recommendations were used after four years, because they were mostly addressed in a successive medium-term plan.

Some evaluations were done before a programme or project was fully developed and had not yet delivered anticipated results. This lowered usage, because recommended measures were not convincing enough to influence project-related policy decisions. Evaluation and recommendation quality is an important determinant of whether a report will be effectively used. The review found that, even when an evaluation was methodologically sound and captured many facts, if its recommendations were insufficiently based on rigorous analysis, its overall quality and use declined.

When asked about report quality, a policymaker in the National Planning Commission responded, "It is an unhealthy competition among evaluators who bid [a] very low amount to get the assignment even if their proposal is not technically sound." Moreover, a policymaker in a line ministry responded: "It is due to low capacities of both of the parties that facilitate or conduct evaluations."

The Government of Nepal has institutionalized the practice of evaluating public-sector projects. However, evaluations have not been conducted systematically or with a clearly defined purpose (whether for lessons learning, ensuring accountability or both). Unless the agencies that facilitate evaluations do not have predetermined ideas about evaluation use or the decisions for which the evaluation provides evidence, evaluation reports will not be

used effectively. A lack of clear evaluation objectives makes it difficult to frame evaluation questions that will generate evidence in areas of interest to policymakers. The absence of clear-cut evaluation policies and periodic M&E plans correlated with inconsistencies in conducting and using evaluations. Whatever evaluations have been done, the use of the reports is not encouraging.

The review findings of the policy papers and key informant interviews clearly indicate that evaluation usability depends on policymakers' commitment and demand for the immediate or medium-term needs, timeliness and quality of the reports. Policymakers in upper echelons are not always clear about the value-for-money allocated to conducting impact evaluations. A high level M&E official in a line ministry said, "Time and, often, questions come from policymakers about the benefits of investing resources in evaluations arguing that issues in project implementation and the results are visible in the surface, so that there is no need to pour resources into studies." Hesitancies like this at higher levels affect the demand for and use of evaluations in public systems.

In addition to evaluation reports, policymakers have other reference and feedback options. Therefore, it is not always right to expect a one-to-one relationship between evaluation recommendations and policies. Similarly, evaluation users are diverse and range from policymakers in planning, budget and line agencies to project managers and subnational policymakers.

Evaluation report usability has been found to depend on how clearly its recommendations are addressed to the implementing agencies and the relevant implementation action plans. Nepal's experience shows that, in order for recommendations to be effectively used, evaluation design must identify the range of evaluation report users.

Usability questions heavily depend on report quality, which in turn relies on the capacity and independent work of the evaluators. Quality concerns can come from the theory of change of the intervention, its methodologies or from its analysis. In some evaluations, the evaluators did not find a project's theory of change or predefined indicators. In such cases, the evaluators were unclear on the programme or project theory of change (especially impacts or outcomes), and instead focused only on the outputs delivered.

ISSUES AND LESSONS LEARNED

The findings show that there are several issues in the use of evaluation in Nepal. The main one is of taking evaluations as a driver of policy or programmatic changes. It is because questions have been raised about the continuity of funding to weak-performing projects in terms of the results generated by evaluation studies.

Nepal lacks a clear and coherent evaluation policy that drives systematic selection, conduct and use of evaluations; weak capacities to demand, facilitate and conduct impact evaluations result in low-quality studies and limited use of recommendations. Similarly, properly documenting and disseminating reports in local languages have been lacking. Further, because the government's performance management system is weak, lessons learned from evaluations have not been effectively used to ensure accountability.

Based on this review, the following lessons can be drawn:

- Institutionalizing evaluation as a core function in public agencies is crucial to increasing the use of evaluations;
- Evaluation quality requires a sectoral results framework with baselines and a defined theory of change;
- A well-designed, mixed methodology that uses a range of quantitative and qualitative techniques is important to improving the quality and use of evaluation;
- More advocacy is needed at policymaking levels to demand and use evaluations; and
- Dissemination of findings and preparing management responses are important preconditions to improving the use of evaluations.

Finally, Nepalese experiences show that evaluation usability heavily depends on the quality of recommendations; recommendations rely on the evaluator's capacities and independence. Hence, strengthening the capacities of government personnel who facilitate evaluations and of the evaluators who conduct them is extremely important to improving evaluation quality and use. Moreover, the in-country, regional and international communities of practice and knowledge networks are important tools for cross-learning and institutionalizing the effective use of evaluations.

REFERENCES

- Bamberger, Michael, and Marco Segone. 2011. "How to design and manage equity-focused evaluations". New York: UNICEF.
- Cracknell, Basil E. 2000. *Evaluating Development Aid: Issues, Problems and Solutions*. New Delhi: Sage Publications.
- Gaarder, Marie M., and Bertha Briceno. 2010. "Institutionalization of Government Evaluation: Balancing Trade-Offs". International Initiative for Impact Evaluation, 3ie Working Paper 8.
- Government of Nepal. 1997. *Ninth Plan (1997-2002)*. National Planning Commission. Kathmandu.
- Government of Nepal. 2002. *Tenth Plan (2002-2007)*. National Planning Commission. Kathmandu.
- Government of Nepal. 2007. *Three Year Interim Plan (2007-2010)*. National Planning Commission. Kathmandu.
- Government of Nepal. 2010. *Three Year Plan (2010-2013)*. National Planning Commission. Kathmandu.
- Government of Nepal. 2013a. *Thirteenth Plan Approach Paper (2013-2016)*. National Planning Commission. Kathmandu.
- Government of Nepal. 2013b. *National Monitoring and Evaluation Guidelines*. National Planning Commission. Kathmandu.

- Lindquist, Evert A. 2001. "Discerning Policy Influence: Framework for a Strategic Evaluation of IDRC". Research Paper, University of Victoria.
- Patton, Michael Quinn. 2011. *Developmental Evaluation: Applying Complexity Concepts To Enhance Innovation and Use*. New York: Guilford Press.
- Roetman, Eric. 2011. "A Can of Worms? Implications of Rigorous Impact Evaluations for Development Agencies". International Initiative for Impact Evaluation, 3ie Working Paper 11.
- Tiwari, B. N. 2010. "An Assessment of Monitoring and Evaluation System of National Planning Commission". Unpublished draft.
- Weiss, Carol H. 1999. "The Interface between Evaluation and Public Policy". *Evaluation*, 5(4): 468-486. Available at evi.sagepub.com/content/5/4/468.abstract. Last accessed 5 February 2014.
- Weyrauch, V., and Langou, G. D. 2011. "Sound Expectations: From Impact Evaluation to Policy Change". International Initiative for Impact Evaluation, 3ie Working Paper 12.

9. ARGENTINA

POLICY EVALUATIONS OF NATIONAL PUBLIC ADMINISTRATION: OVERVIEW

NORBERTO PEROTTI

Under-Secretary, Secretariat of Budget Evaluation
Office of Ministries of the Presidency

INTRODUCTION

In the past few years, Argentina has made several efforts to consolidate policy evaluations as a key component of the management cycle.

In early 2013, within a context of different institutionalized efforts by organizations dedicated to public-sector evaluation, the Head Department of the Nation's Ministers' Cabinet established the Policy Evaluation Program in order to build awareness on and mainstream evaluation practices within the central management level. The programme gathered success histories of M&E systems within the Argentinean state. With such an aim, the programme jointly coordinates the work of three Under-secretariats of the Head Department of the Nation's Ministers Cabinet.

CONTEXT AND ACTIONS DEVELOPED

Evaluation practices have been part of the region's government agendas for several years. Based on new institutional management needs, evaluation has been positioned as an integral part of the policy management cycle. In Argentina, government policies include these management tools to assess and guide actions towards expected outcomes. The aim is to strengthen social inclusion policies and to expand citizen rights.

Systematizing and mainstreaming public policy evaluation into management processes requires expert human resources, strong political will and a culture of promoting evaluations

as an institutionalized practice within the Argentinean public administration. Mainstreaming processes stem from the need to incorporate evaluation into the government's agenda and policy design as a key driver of quality goods and services. Therefore, the goals of launching a mainstreaming process stem from public agencies' needs. Management requires tools that measure the effective development of past, ongoing and future policies, and that can be used to improve decision-making and product quality.

Three main components form M&E systems in Argentina: a budget-related system (which conducts the physical and financial follow-up of budgetary programmes), a system linked to social programmes (which gathers information about benefits, beneficiaries and programmes' executed budget); and a government programmes' M&E system (which monitors programme management by following-up goals).

The monitoring system involved in preparing, executing and controlling the national budget has the largest regulatory framework and broadest coverage. The Ministry of the Economy, through its Budgetary Assessment Directorate, conducts follow-up of the specific goals of 380 National Budget programmes. Each programme reports quarterly indicators, built jointly among programme leaders and management. Currently, Program for Productive Institutional Strengthening and Provincial Fiscal Management (PROFIP), a programme funded by the Inter-American Development Bank, seeks to develop outcome indicators in 100 programmes budgeted for the next three years. The indicators will help assess results-based management and complement the follow-up of physical goals.

The National Social Policy Coordination Council, under the Nation's Presidency, coordinates social programme efforts with several related ministries. The National Directorate of Social Program Information, Monitoring and Evaluation is the unit in charge of: designing and implementing information systems to study the socio-economic status of the population; conducting useful policymaking analyses; capturing information about existing social programmes; monitoring progress in the compliance of goals established under such programmes; and assessing potential matches between expected and achieved impacts. The systems are implemented at the national, provincial, municipal and local levels for decision-making.

SISEG (the country's monitoring and evaluation system, *Sistema Integral de Seguimiento y Evaluación de la Gestión*), under the framework of the Head Department of the Nation's Ministers' Cabinet, is a management tool that systematizes the information that M&E processes gather about government policies and priority programmes in different jurisdictions. This tool supports the consolidation of the results-based management approach in Argentina by integrating and processing information on expected achievements under the strategic plan.

The Head Department of the Nation's Ministers' Cabinet established the Policy Assessment Program. It was created by linking three of the ministry's under-secretariats, based on its coordination role and mission of understanding policy M&E, the coordination of sector evaluation systems, and follow-up of government programmes. By developing evaluation processes, the programme promotes improved governance, enhances policy quality and improves public management outcomes.

Programme goals for the next three years include:

- Awareness-raising, agenda positioning and mainstreaming policy evaluation in the national public administration;
- Developing evaluation capacities in national public administration;
- Establishing effective evaluation of domestic policies; and
- Developing knowledge through applied research in policy evaluation.

The programme has conducted training workshops for technical national public administration officials. The first evaluations are being designed for several programmes under the Cabinet's Head Department and other ministries. An outcome evaluation has been proposed for the strategic plan of the Ministry of Labor and Social Security. An additional activity, linked to the goal of awareness and agenda positioning, is the International Seminar of Public Policies. In addition, the programme has become the evaluation component of Results-Oriented Management External Pillar Implementation Program (PRODEV), the steering unit within the National Budget's Assessment Under-secretariat.

CHALLENGES

The establishment of this programme is a qualitative improvement in M&E mainstreaming in Argentina. The programme is not intended to replace existing tools and systems, but rather to recover successful experiences and to generate conditions to promote them.

Currently, the three systems operate independently based on their appropriate goals, functions and resources. The strategy of the Public Policy Assessment Program is to rely on its experiences in order to position evaluation in the public agenda so that the national public administration can use existing tools. Therefore, the challenge is to prepare fertile ground for the actions of evaluation stakeholders.

The programme's main challenge is to become a legitimate stakeholder in a context of highly developed but poorly shared experiences. The Public Policy Assessment Program is a venue for coordinating systems and the parties within each national administration organization that have final responsibility for evaluation. Therefore, the programme goals are not limited to developing programme and policy evaluations, but rather aim at promoting awareness, mainstreaming and developing installed capacities in order to improve evaluation processes in Argentina. For this purpose, the programme relies on the political support of the Cabinet's Head, the main party responsible for coordinating the efforts of the national executive power.

CONCLUSIONS

During the past few years, Argentina has been introducing techniques and procedures to streamline policy management efforts and has been establishing required tools to monitor and evaluate main policy decisions. In 2013, the political will of the Cabinet's Head promoted and focused evaluation with the development of the Public Policy Assessment Program in order to complement existing systems, to mainstream the topic and to build an evaluation

culture within the national state.

The programme thus adopts a progressive approach, with adequate and consistent capacity planning and building efforts. In Argentina, experiences with M&E implementation have shown that without the required support and awareness of intermediate management tiers, it is extremely difficult to adopt practices beyond government management or to prevent their removal due to structural public administration changes. The programme thus becomes a key player in the central administration's policy evaluation mainstreaming process.

REFERENCES

- Government of Argentina. 2013. *Resolución 416/2013, de 7 de Junio, de Jefatura de Gabinete de Ministros. Boletín Oficial*. Year CXXI, No. 32.658, p. 30.
- Nirenberg, Olga, Josette Brawerman, and Violeta Ruiz. 2000. *Evaluar para la transformación: innovaciones en la evaluación de programas y proyectos sociales*. Paidós.
- Oficina Nacional de Innovación de Gestión. 2007. *La Gestión Pública por Objetivos y Resultados: Una visión sistémica*, Buenos Aires. Available at sgp.gov.ar/contenidos/onig/planeamiento_estrategico/docs/libro/Gestion_por_resultados_version_completa.pdf. Last accessed 5 February 2014.
- Ortegón, Edgar, and Adriana Prieto. 2005. *Metodología del marco lógico para la planificación, el seguimiento y la evaluación de proyectos y programas*. Vol. 42. United Nations Publications. Available at eclac.org/publicaciones/xml/9/22239/manual42.pdf. Last accessed 5 February 2014.

10. BRAZIL

A MODEL TO EVALUATE THE MATURITY OF THE BRAZILIAN PUBLIC ADMINISTRATION'S EVALUATION SYSTEMS

SELMA MARIA HAYAKAWA CUNHA SERPA, Federal Auditor
GLÓRIA MARIA MEROLA DA COSTA BASTOS, Federal Auditor

Brazilian Court Of Accounts (TCU)

INTRODUCTION: FEDERAL COURT OF ACCOUNTS – BRAZILIAN COURT OF AUDIT

The Federal Court of Accounts – Brazilian Court of Audit is an autonomous, specialized organization that supports the external control conducted by the legislative branch. The court aims to ensure adequate and effective use of public funds by using a range of tools, such as auditing and judgement of the annual rendering of accounts by those responsible for government asset management.

The Brazilian Court of Audit has developed approaches that assess public management tools as evaluation systems, internal control systems, information technology governance systems and risk-assessment tools. These assessments will ensure effective governance and management of public policy and programmes and will contribute to the betterment of public administration.

OBJECTIVE

This paper presents a model that was developed by the Brazilian Court of Audit to characterize the maturity of evaluation systems in Brazilian public administration. Evaluation systems have been disseminated as a key tool to support decision-making processes and organizational learning, to promote transparency in programmes, and to assess policy performance and results. This information is essential to increasing the trust and confidence

of all stakeholders in the policy process, and it contributes to promoting coordination, accountability and more effective governance. Identifying and characterizing the public administration's evaluation systems allows the Brazilian Court of Audit to recognize the gaps and the need for improvement in the systems, and to make recommendations to remedy any deficiencies.

The model was based on the literature of evaluation systems and evaluation capacity building, with a focus on the work of Leeuw and Furubo (2008) on organizational capacities to perform and use evaluation.

The Brazilian Court of Audit has conducted two preliminary surveys to verify the model's adequacy to reach its objectives and goals. The results demonstrated that the model is a strong instrument that characterizes the maturity of ministries' evaluation systems. Therefore, the court approved the extension of the model survey to all Brazilian ministries.

LITERATURE REVIEW

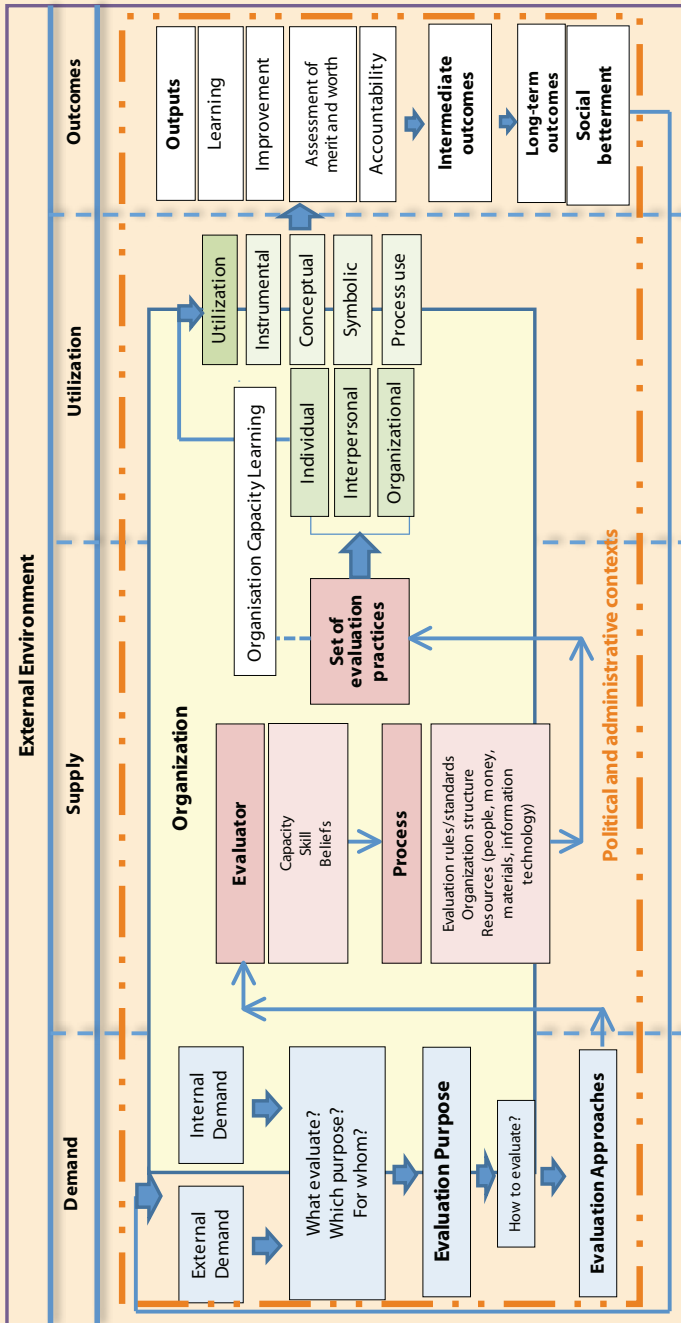
The structure of evaluation systems is related to the institutionalization of evaluation as an instrument to support the functions of planning, control and accountability. For Gray et al. (2002), budgeting and auditing are the main tools used for the exercise of these functions (in addition to the evaluation). Planning, coordination and management of governmental actions are made through the support of budgeting. Evaluation of programmes and policies improves public policies by producing knowledge that subsidizes organizational learning and decision-making to improve public policies. The audit ensures financial control and accountability for the use of public resources.

According to Grau and Bozzi (2008), the growing utilization of M&E systems in the public sector in Latin America creates further transparency and improves the effectiveness of government actions and, in this manner, increases social control capacity and state legitimacy, and facilitates anti-corruption efforts, better use of public resources, and the creation of policies and services that promote social welfare, thus reducing poverty and inequality.

Leeuw and Furubo (2008) applied four criteria in labelling a set of evaluative activities as a system. The first criterion takes a distinctive epistemological perspective; the second criterion is that evaluation activities should be carried out by organizations and institutions and not largely by independent evaluators. The third criterion is the permanence or history in the activities involved; these activities should be part of an organization's initiatives. The fourth criterion focuses on the intended use of evaluation, i.e. information from evaluative activities should be linked to decision and implementation processes.

According to Jannuzzi (2012), M&E systems are articulated parts of a more general system of policy and programme management that supplies customized information and knowledge from formative and summative evaluation. These general systems demand data necessary for decision-making processes from M&E systems. The author states that M&E systems are not independent from the general system once they are developed for the purpose of producing knowledge—because their main reason for existence is to produce knowledge—for the improvement of management, even though an M&E system can also contribute to

FIGURE 1. A MODEL TO EVALUATE THE MATURITY OF THE BRAZILIAN PUBLIC ADMINISTRATION'S EVALUATION SYSTEMS



transparency of government actions and the judgement of programmes and policies.

The institutionalization of evaluation systems can be characterized by the mechanisms that define a regular and continuous stream of demand for evaluation that directs a set of evaluation practices that are formalized, structured and coordinated to produce knowledge that aims to contribute to decision-making processes and organizational learning.

DIMENSIONS OF ANALYSIS OF THE PROPOSED MODEL

Based on this literature review, the proposed model to identify and characterize the maturity of evaluation systems encompasses the following dimensions of analysis (see Figure 1):

- **Demand:** the external and internal contexts in which demand for evaluation is generated. In this dimension, the constructs to be evaluated relate to the external and internal organizational environment (political and administrative), where demand for evaluation is generated, structured and bound by its purpose. Thus, we are interested in examining the adequacy of the planning processes aimed at assuring the consistency of the object to be evaluated, to whom (intended users of evaluation) and for what purpose;
- **Supply (production of knowledge – organizational responsibility):** the structuring of processes and the organization of means to conduct evaluative activities (evaluation capacity). In this dimension, the constructs to be studied are with respect to: the definition and dissemination, within organizations, of the evaluative practices to be used; the organizational support, in terms of education background and training of the professionals responsible for conducting evaluative activities; the formalization of evaluative practices, under the definition of their responsibilities, procedures and instruments; and the allocation of the means required to conduct the activities;
- **Capacity for organization learning:** an organization's attributes and conditions to support learning, with respect to the clarity and support for the organization's mission and vision, the leadership that supports learning, an experimental organizational culture, the ability to transfer knowledge, and teamwork and cooperation; and
- **Use:** the investigation of mechanisms that favour the utilization of knowledge produced by evaluative activities in order to improve programme and policy management.

METHODOLOGY

Two preliminary studies gathered data on the perceptions and opinions of a non-random sample of public managers that are responsible for implementing finalistic programmes and policies. A characterization and evaluation based on the constructs of the model were made regarding how the agencies are structured to attain evaluation knowledge. The studies measured the activities and resources available to support these activities, the purpose of their use and agencies' success in developing an evaluative learning culture.

A questionnaire based on the constructs of interest in the study was developed for a preliminary survey. This questionnaire included four parts: questions about the demand side, the supply side (evaluative production of knowledge), the organizational learning capacity

FIGURE 2. MATURITY PROFILE OF EVALUATION SYSTEMS



and the use of evaluation knowledge. There were 24 questions in total.

Prior to the survey, two Brazilian ministries were invited to take part in the pilot study. Due to the nature and purpose of the study, to evaluate the adequacy of the model and the instrument to measure the maturity of the evaluation systems, it was not disclosed which ministries took part in this preliminary survey. The ministries identified all the decision-making managers in their agencies that are responsible for implementing finalistic programmes and policies (162 in Ministry 1 and 147 in Ministry 2). An electronic survey was sent to all managers; 118 usable responses were received (73 from the first ministry and 45 from second ministry).

Descriptive statistics were prepared using SPSS Statistics12.0. The mean of each construct was adopted to characterize the maturity profile of the two ministries' evaluation systems (see Figure 2).

RESULTS

The results show an intermediate maturity of the evaluation systems in both ministries, with emphasis on the relevance of the knowledge produced for improvement of both programmes and management (as can be seen in the utilization dimension). There are also possible improvements in the dimensions of evaluation demand and knowledge production. In the first case, improvements can be made with respect to the planning of evaluation demand, information sharing and evaluation tools. In the second case, which deals with the development of evaluative capacity, possible improvements deal mostly with personnel

training and allocating necessary resources for the development of activities.

In regard to the organizational environment, the experimental component is highlighted as favourable to the development of an evaluation culture. However, there are opportunities for improvement in the definition of purposes (object and objectives of the evaluation system), as well as in knowledge transfer through information sharing and organizational learning. This reinforces the remarks made about evaluation demand.

The demand of Ministry 1 had been systematized and oriented, mainly by the plurianual plan and the sectorial plans. In Ministry 2, in addition to the pluriannual plan, strategic planning took on an important role in the orientation of evaluation activities.

A heavy competition was observed between internal and external evaluation demand, primarily within the institutions of Control and the Executive Office. This could compromise the evaluative capacity of the ministries, considering the insufficiency of human resources and the gaps in the necessary people skills for completing evaluation activities.

With respect to the flow of information between the suppliers and recipients of the produced evaluation knowledge, it was observed that deficiencies that may compromise the utilization of information still exist (notably in the transfer of produced knowledge).

With regard to the development of evaluation capacity, in Ministry 1 the majority of respondents agreed that the tools used to monitor and evaluate programmes, actions and policies (along with information utilization) are divulged and disseminated in their respective agencies. However, their teams were unable to identify objectives and the tools used, even though they agree that the instruments are formally instituted.

According to the respondents from Ministry 1, the steps for communicating M&E results are not completely formalized, a weakness in the communication system.

Regarding personnel capacity, the majority of respondents in both ministries agreed on the importance of actions that lead to capacity building for the work being developed. However, more than 50 percent of Ministry 1's respondents did not know or disagreed that the ministry provides incentives for employee improvement. More than 50 percent also disagreed that the ministry equitably promotes training for all employees in order to satisfy the needed capacity. In Ministry 2, 45 percent of respondents did not know or disagreed that the ministry provides incentives for the improvement of employees, and 41 percent did not know or disagreed that the ministry promotes equitable training for all employees.

It is relevant to emphasize that both ministries are in the initial stage of supporting the development of leadership skills. There is also space for improving teamwork, because more than a third of respondents did not know or disagreed with this aspect.

It is also worth noting that most respondents of both ministries agreed that the technical knowledge for employees that undertake M&E activities is adequate (although the average of responses did not corroborate this assertion). In contrast, the majority of respondents of both ministries believed that the amount of people to monitor and evaluate programmes is insufficient.

Twenty five percent of respondents of Ministry 1 did not know if an overlap of M&E activities existed between different agencies of the ministry; over 30 percent agreed that such overlap exists. In Ministry 2, almost 70 percent of respondents agreed that some overlap exists.

According to 68 percent of respondents from Ministry 1 and 60 percent from Ministry 2, their respective agencies have the ability to evaluate other studies of M&E of programmes, actions and policies. Furthermore, 58 percent of ministry one's respondents and 61 percent of ministry two's respondents agreed that their agencies have already developed the necessary technical competence to develop these activities.

The main evaluation practices in Ministry 1 are the elaboration of management reports, which obtained the highest percentage of agreement between respondents (82 percent), followed by evaluation planning (80 percent), normative revision of the programme (79 percent), results monitoring (75 percent), and implementation monitoring (74 percent). These practices indicate the nature of the evaluation system in Ministry 1, whose activities are focused on monitoring the implementation and results of programmes and policies. In the respondents' perception, evaluation practices contributed mostly to utilizing the process of programme comprehension, improving management, understanding the importance of these tools, developing abilities and techniques and increasing commitment within the organization.

The main evaluation activities of Ministry 2 are management reports, followed by performance indicators, studies of internal diagnostics and internal meetings. These results make it evident that the structure of the evaluation system is given by performance indicators and the management reports in Ministry 2, which are still segmented by agencies responsible for the implementation of programmes, actions and policies.

With regard to the availability of budgetary and financial resources, more than 50 percent of the respondents in both ministries agreed that their units have access to sufficient resources to fulfil the activities of M&E of programmes, actions and policies; about 44 percent did not know or disagreed in this respect.

It is relevant to highlight that 46 percent of respondents in Ministry 1, and 52 percent in Ministry 2, did not know or disagreed that the tools for evaluation activities (e.g. equipment, software, administrative support) were available in the institution. In addition, 67 percent of respondents in Ministry 1 and 68 percent of respondents in Ministry 2 disagreed or did not know whether the professionals involved in these activities had enough time to reflect on identified successes and failures.

In regard to organizational learning, data analysis showed that Ministry 1 is an institution that favours experimentation and teamwork-based resolution of problems. However, with respect to clarity of purpose, 39 percent of respondents did not know or disagreed that a self-evaluation process exists in relation to the objectives reached by the agency; 34 percent did not know or disagreed that senior managers and their teams shared a common vision of the activities to be developed; and 32 percent did not know or disagreed that all teams shared the ministry's mission and values.

In regard to organizational learning in Ministry 2, according to the respondents, it favours empowerment, a teamwork-based resolution of problems and experimentation. However, with respect to clarity of purpose, as seen in the previous case, almost 60 percent of respondents did not know or disagreed that the ministry's mission identifies values that are shared by all teams. In addition, 33 percent of respondents did not know or disagreed that there exists a shared vision of activities to be developed between senior managers and their teams, and

that there exists a self-evaluation process related to the objectives reached by the agency (27 percent).

With respect to the findings' utilization, it was observed that managers generally have a positive perception of evaluation activities for learning and improving programmes and policies, particularly for promoting change in order to promote the understanding of a programme's function and to identify improvement opportunities.

The proposed model makes it possible to identify the maturity profile of existing evaluation systems and to build a taxonomy specific to Brazilian public administration, enabling more effective control of tools used to aid public policy and programme management.

REFERENCES

- Grau, Nuria Cunill, and Sonia Ospina. *Fortalecimiento de los sistemas de monitoreo y evaluación (M&E) en América Latina: informe comparativo de 12 países*. Vol. 12. Banco Mundial, 2008.
- Kusek, Jody Zall, and Ray C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington: The World Bank.
- Jannuzzi, Paulo M. 2012. Um referencial teórico para análise da institucionalização dos sistemas de avaliação. 3 f. Exame de qualificação em conteúdos específicos (Doutorado em Administração). Programa de Pós-Graduação em Administração, Universidade de Brasília.
- Leeuw, Frans L., and Jan-Eric Furubo. 2008. "Evaluation Systems: What Are They and Why Study Them?" *Evaluation*. 14 (2):157-169.

11. BRAZIL

EVALUATING SOCIAL DEVELOPMENT IN BRAZIL: ANALYSIS OF THE CYCLE OF PRODUCTION AND DISSEMINATION OF EVALUATION STUDIES BY THE MINISTRY OF SOCIAL DEVELOPMENT AND FIGHT AGAINST HUNGER

MARCO ANTONIO CARVALHO NATALINO

Coordinator-General for the Evaluation of Impacts and Results

ALEXANDRO RODRIGUES PINTO

Acting Director of Evaluation

MARTA BATTAGLIA CUSTODIO

Coordinator for the Evaluation of Impacts and Results

Department of Evaluation

National Secretariat of Evaluation and Information Management

Ministry of Social Development and the Fight against Hunger (SAGI/MDS)

OVERVIEW

This paper presents an analysis of the cycle of production and dissemination of evaluation studies by the Brazilian Ministry of Social Development and Fight Against Hunger (SAGI/MDS). It aims to describe and contextualize institutional advances and the standards involved in the evaluation of Brazil's far-reaching social programmes over the last decade, a period when the country's social progress was accompanied by the creation of a secretariat responsible

for the evaluation of many of the country's new social policies. This paper explains the steps involved in determining how and when social programmes should be evaluated, in specifying the research methodology and theoretical perspective and in contracting external evaluation institutions. The paper also discusses how the ministry, through its Secretariat of Evaluation and Information Management (SAGI), monitors and qualifies the conduct of research and knowledge dissemination. The discussion contributes to the understanding of how the creation of credible public institutions and the mix of internal and external evaluation studies can foster the progress of an evaluation culture and can enhance the efficiency of public policies, particularly in developing countries.

NINE STEPS FOR PRODUCING AND DISSEMINATING EVALUATION STUDIES

Evaluating public policies is an important function for improving and qualifying decision-making by public officials by bringing empirical evidence that allows more rational and objective choices. Weiss (1998, p.10) notes that objective information on programme implementation and results can bring "more sensible choices" to public fund allocation and to programme planning. Weiss also highlights that evaluation can and must be used as a way to guarantee accountability, with emphasis on the objectives reached to the internal public and the comprehension of social intervention by the external public (Weiss 1998, 26–28).

With the creation of MDS in 2004, which incorporated actions on nutritional security, cash transfer, social assistance and later on productive inclusion, it became necessary to organize a structure capable of producing relevant and timely information to help the design, implementation, focus and reordering of public policies under its purview. SAGI answers this demand. Vaitsman, Rodrigues and Paes-Souza (2006, 15) consider it an innovation in Brazilian public management, because there was no structure in Brazil with an evaluation organization that shared its hierarchical level of the evaluated units. To reach its goals, SAGI, through its Department of Evaluation (DA/SAGI) uses mainly external evaluation studies, contracting research institutes, private research companies, individual consultants and research groups through the Brazilian Council for Scientific and Technological Development.

After almost 10 years of work, SAGI has produced more than 130 evaluation studies, contemplating a wide group of actions, programmes, benefits and services under the responsibility of MDS. DA/SAGI uses varied methods to evaluate programmes, because the choice of methodology depends on the evaluation questions and the phase of the programme's implementation (Jannuzzi et al. 2009).

This model of evaluation, using evaluators that do not belong to the ministry's structure, can be considered what Worthen, Sanders and Fitzpatrick (2004, 289) define as a mixed evaluation, because it combines the advantages of an internal analysis (e.g. knowledge of programmes, capacity of dialogue with public officials, better comprehension of the organizational culture and the decision-making dynamic) with the advantages of an external analysis (e.g. greater autonomy, external credibility, the possibility to bring specific knowledge on some methodologies). DA/SAGI staff members monitor the attendance of the research schedule and keep a constant dialogue with both the contracted institute and

the officials responsible for the evaluated programme, ensuring the quality of the results in terms of theoretical and methodological consistency. DA/SAGI also reaches out to those that will use the evaluation in order to assess their knowledge demands. Through this method, SAGI aims to accomplish the difficult task of ensuring both the independence and credibility of the research itself, and that the evaluation will be actually used to the qualification of public policies (a common Achilles' heel of evaluation studies).

To achieve the amount of research studies and guarantee the qualities previously described, SAGI has systematized its actions in nine steps that range from validating demand through to the publication of microdata on the Internet.

Step 1: Discussion

Tapajós et al. (2010) note that the nature and the object of the studies that will be done is a result of a collective debate among many sectors of MDS. The debate considers whether a programme has been evaluated yet, the evaluation objectives, appropriate methodologies for the reaching evaluation goals, budgetary constraints and the existence of previous studies done by other institutions. Discussion also considers the needs identified through the periodical review of programmes under the MDS umbrella.

Guided by the ministry's strategic planning, the work group of monitoring and evaluation is the institutional forum that enables the discussion of priorities to take place and organizes the evaluation agenda for the next year. In addition to SAGI, all other MDS secretariats, the Deputy Minister's office and the Ministers Cabinet participate in the working group.

Step 2: Terms of reference

Specifying the objectives of the evaluation and its methodology requires a technical discussion with the demanding area of MDS. This process makes it possible to identify research questions and information demands that can be supplied through other means and strategies, such as the organization of existing databases or the creation of historical series through the manipulation of existing indicators. The type of evaluation study is then defined, along with its objectives, methods and data sources. After the object of study is defined and refined, the Department of Evaluation writes a terms of reference to hire the research team and to specify technical, methodological and administrative requirements to ensure the quality of research.

Step 3: Contract

The hiring of an external evaluator is based on the most advantageous proposal sent to the administration through a process of public bidding, which guarantees wide competition among interested parties. Bidding is usually conducted through an electronic auction (*pregão eletrônico*). It is important that the terms of reference is as detailed, specific and clear as possible. Some questions that do not impact the budget may still be clarified at the first meeting of the research team and SAGI.

Step 4: Dialogue for study operationalization

The first meeting between the research team and SAGI is the time and place for an open dialogue. At the meeting, SAGI and the secretariat interested in the evaluation discuss the administrative terms included in the signed contract, the methodological rigour, the theoretical quality and knowledge expectations that the finalistic area put on the study. External researchers get a better idea of the 'essentials' and 'peripherals', which improves the focus of their analytical effort. This process continues throughout the research schedule, with many discussions occurring as preliminary data is sent to SAGI.

Step 5: Monitoring the study

Research groups or individuals are never solely responsible for executing research activities; SAGI follows each step, giving substantive contributions on each one of them. As any research enterprise is a complex matter, this technical involvement is crucial to identify problems and guarantee the quality of final results. Furthermore, it is important that monitoring is done by a very qualified staff, so that the debate may be kept on a high level with respected academics and researchers. This process also creates a cycle in which SAGI staff are constantly qualified by the close contact with good (and bad) researchers. A true account of the monitoring process must allow for and cope with bad researchers, not suppress their existence. Furthermore, good researchers may do bad research on some cases, as generating new knowledge is always complex and may not be necessarily done within the time frame expected by government officials. The monitoring of researchers and the institutional learning of SAGI allows for quick identification of problems and avoids the worst consequences of good research gone bad.

Step 6: Final results

Once the research is done, a cycle of technical discussions begins with the proposition of a table plan and the presentation of partial results. It often involves, in addition to the research report to be done by the external research group, technical studies written by the Department of Evaluation staff or by an external consultant. As customary and recommended with any research, the analysis aims at the research objectives but frequently reaches out to new and diverse aspects found during fieldwork, giving a rich interpretation of the data that goes beyond the original intent.

Step 7: Dissemination of results

The results are then presented to the finalistic secretariat for a technical discussion and deepening of the analysis. This process may or may not involve the external research team. After discussing the results internally, SAGI may organize public seminars to disseminate the study among stakeholders in government, academia and civil society. It is also common for SAGI to present studies in academic and governmental seminars.

Step 8: Publication of results

The publication of results is a fundamental step that guarantees accountability and gives credibility to SAGI and the MDS. It is important to acknowledge that the results of any study that is done with public funds must be public and open to societal inquiry. The use of the Internet is a powerful tool in this regard, giving the option of a low-cost investment with a great return. The publication involves an executive summary, a research report and a technical paper, in addition to other research papers that may come from the original data, and eventually books. SAGI also organizes a periodical and *Cadernos de Estudos* (study brochures) to publicize its research.

Step 9: Publication of microdata

The publication of microdata is the final step and gives the research community an opportunity to do its own analysis of the data. By institutionalizing this procedure, SAGI became a credible source for statistical data on social policies. Publishing microdata also strengthens research in this area and gives SAGI a central role in the research evaluation epistemic community.

CONCLUSIONS

The nine steps done by SAGI to guarantee the quality of evaluation studies is a case of good management of the evaluation process. The steps combine the advantages of internal and external evaluations and allows for multiple methods and perspectives. The objective of this short paper will be met if this presentation helps other countries and institutions qualify their own processes, strengthening evaluation culture and qualifying the public policies examined.

REFERENCES

- Bamberger, Michael, Jim Rugh and Linda Mabry. 2011. *RealWorld Evaluation: Working under Budget, Time, Data, and Political Constraints*. Sage Publications.
- Jannuzzi, P.M et al. 2009. *Estruturação de sistemas de monitoramento e especificação de pesquisas de avaliação, os problemas dos programas públicos no Brasil são*. Cadernos EIAPP – Avaliação de Programas Sociais. ENAP. Brasília.
- Tapajós, L. et al. 2010. *O Processo de Avaliação da Política de Segurança Alimentar e Nutricional entre 2004 e 2010: a Experiência do MDS*. IN: BRASIL. Ministério do Desenvolvimento Social e Combate à Fome. Fome Zero: Uma História Brasileira. Brasília, DF: MDS, Assessoria Fome Zero, V. 1, 2 e 3.
- Vaitsman, J., e Paes-Sousa, R. 2011. *Avaliação de programas e Profissionalização da gestão pública*. Revista Brasileira de Monitoramento e Avaliação, Número 1, Janeiro-Junho de 2011.
- Weiss, C. 1998. *Evaluation*. New Jersey: Prentice Hall.
- Worthen, Blaine R, James R. Sanders and Jody L. Fitzpatrick. 2004. *Avaliação de Programas - Concepções e Práticas*. Edusp. São Paulo.

12. DOMINICAN REPUBLIC

ANALYSIS OF LESSONS LEARNED BY THE DOMINICAN REPUBLIC IN LOCAL AND INSTITUTIONAL/ GOVERNMENT-LEVEL MONITORING, EVALUATION AND PARTICIPATORY FOLLOW-UP

HERNAN RODRIGUEZ MINIER

Economist, Researcher and Professor
School of Economics and the Centre for Entrepreneurship
and Entrepreneurial Innovation of the
Universidad Autónoma de Santo Domingo

INTRODUCTION

Recently, the merits and features of participatory follow-up and evaluation have been a matter of heated debate. Several academic and empirical trends are evident, both in the fields of development theory and policy design. Donors in Latin America and other regions insist on the inclusion of primary stakeholders in follow-up and evaluation activities, and non-governmental organizations are increasingly doing so.

Programmes and projects are aimed at impacting public policy areas, either at the national, regional or local levels. For this reason, programmes must be subject to a monitoring, follow-up and evaluation plan that, in turn, is part of a follow-up and evaluation system capable of improving the plan's effectiveness and applicability. To be effective, a follow-up and evaluation system must include a participatory component that takes into account both direct and indirect stakeholders eligible to support its implementation.

This paper describes the contributions of a 'community' or 'participatory' follow-up system of planning, follow-up and evaluation efforts that have participatory components and outcomes, analysis of strengths, weaknesses, opportunities and threats (SWOT analysis) and lessons learned resulting from such experiences.

MAIN CONTENTS

In the past, development programmes failed to systematically include evaluation activities. Currently, most programmes—unfortunately not all—include some type of evaluation component. Most programme managers or coordinators agree that evaluations are not only necessary, but are also relevant to inform adequate decision-making regarding the strengthening of programme implementation strategies.

To improve its efficiency, a follow-up and evaluation system must be endowed with a participatory component that considers direct and indirect stakeholders that are able to contribute to the system's implementation. Community follow-up systems play an important role in community capacity-building by empowering communities to monitor and assess efforts aimed at improving their socio-economic status. Participatory follow-up and evaluation provides several innovations regarding conventional follow-up and evaluation methods. Their distinctive features result from the four basic principles applied to project and public policy evaluations: participation, learning, negotiation and process flexibility.

Participation means including populations in process design and data analysis, a significant difference from conventional systems. **Learning**, particularly hands-on learning, entails a basis for continuous improvement and corrective actions. **Negotiation** between the highest possible number of cohorts (e.g. population, donors, civil service) builds consensus on what should be subject to follow-up and evaluation, how and when data should be collected, the meaning of information, how to share results and what actions should be undertaken. **Process flexibility** is essential to adapt to changes in beneficiaries (e.g. number, role, knowledge) and their context. Rather than merely identifying issues and shortcomings, evaluation programmes should focus on gathering lessons learned for future use, both from successes and from challenges.

A participatory evaluation process can help improve programme staff communication during the programme's various implementation levels.

Although the Dominican Republic still has limited experience in participatory and follow-up evaluation, the country is taking the necessary steps to promote social methods that help communities take part in the development and follow-up of public policies, plans, programmes and projects. These efforts have been classified into three categories: local efforts focused on and organized for programmes and projects; efforts focused on local, institutional-type issues; and national efforts focused on the design, follow-up and evaluation of public policies and national development plans.

A SWOT analysis of these experiences has resulted in the following conclusions:

STRENGTHS	WEAKNESSES
Increased community empowerment regarding socio-economic challenges Increased effectiveness in information collection Participation helps improve programme and project performance	Poor civil society participation Need further training in information collection methods Neither local governments nor public policies promote these organizations No budget for their development
THREATS	OPPORTUNITIES
Civil society wants to concentrate these types of initiatives Involvement of political parties in these organizations	Interest of multinational organizations in their development Populations are interested

CONCLUSIONS

While some steps have been taken at the legal level, in practice, participatory follow-up and evaluation is limited in the Dominican Republic; experiences in the field are not being replicated.

Local follow-up committees have specific experience in participatory follow-up and evaluation, based on community follow-up system methodologies and principles.

Institutional participatory venues are an adequate participatory follow-up and evaluation mechanism, but need to be strengthened.

More opportunities for participation at the local institutional level must be developed, but with increased civil society involvement. Currently, 50 percent are formed by representatives of national and local governments.

13. DOMINICAN REPUBLIC

EXPERIENCE OF THE DOMINICAN REPUBLIC IN THE EVALUATION OF THE SOLIDARIDAD (SOLIDARITY) PROGRAMME

IVETT SUBERO ACOSTA, Director of UN System Monitoring
DORIS MELO, Sub-Director of UN System Monitoring
SÓCRATES BARINAS, Evaluation Project Coordinator

Multilateral Cooperation Office (DIGECCOM)
Ministry of Economy

BACKGROUND

With the aim of improving poor social indicators, the Government of the Dominican Republic promoted protection policies for extremely poor or vulnerable groups. In 2004, it launched a new Social Policy Strategy aimed at reducing poverty levels.

In 2005, the government began implementing its social protection net with the Solidaridad Programme, a means to develop a poverty reduction strategy. The Solidaridad Programme supported income-improvement and human capital investment efforts for extremely poor households (identified by a targeting system) through conditional cash transfers. Conditional cash transfers were provided through an electronic card, supported the prevention of inter-generational transmission of poverty, and improved household investments in education, health and nutrition. The Solidaridad Programme is one of the first poverty and social exclusion reduction efforts developed by the Dominican Government, and the programme is coordinated by the Social Policy Cabinet under the Vice-Presidency of the Republic.

In its early stages, the Solidaridad Programme targeted 95,000 pilot households. By 2010, it served 806,000 extremely poor households; it covered 1 million households at the national level. A social protection policy based on conditional and non-conditional cash transfers was implemented through several subsidies: *Comer es primero* (Eating is first), for food purchases up to \$20; the School Attendance Bonus, amounting to \$4 per registered child; Dominicans with a Last Name, facilitating birth and ID registration; a gas voucher, amounting to \$6 for purchases of liquefied petroleum gas; and a power voucher, providing \$11 for electric power.

To be eligible for these subsidies, beneficiaries need to meet requirements related to health improvement, education, basic sanitation and citizenship. The amount of the monthly transfers is based on the conditions met by each beneficiary.

SOLIDARIDAD PROGRAMME EVALUATION

M&E efforts of social programmes and policies are key in the Social Policy Strategy, as supported by the Inter-American Development Bank, the United Nations Development Programme (UNDP) and the World Bank. Programme assessment is crucial to establishing the programme's efficiency and accountability, particularly given the amount of resources it delivers. This assessment is conducted through an M&E system that includes surveys to a sample of 2,000 households to gather information on: beneficiary lifestyles and characteristics; basic education indicators; population features; labour market insertion; access to health care and nutrition; human measurements; and household income and spending. The surveys help measure the programme's direct impact on the households' welfare. Survey application and sampling was developed by the National Accounts expert team of the Dominican Republic Central Bank.

The Solidaridad Programme began the first baseline round of impact assessments in 2010. The purpose of the assessments was to identify programme weaknesses, gather lessons and make improvements to guarantee expected welfare outcomes and impacts on beneficiary versus non-beneficiary households. In 2011, the second round of surveys was conducted as a follow-up to the households interviewed in the baseline assessment.

The assessment identified impacts of the conditional cash transfer programme, including behavioural changes in health and education. Beneficiaries increased the frequency of health care prevention visits, compared to control groups. Improvements are also apparent in vaccination and routine check-ups among beneficiary children under the age of five, with higher rates than non-beneficiaries. The programme has changed behaviours that become an asset to build future capacities, thereby improving family life expectations.

EVALUATION RESULTS

Outcomes on the beneficiaries' quality of living indicators: The outcomes on quality of living indicators can be quantified in health, education and available income of beneficiary households.

Health status of beneficiary household members: The outcomes on quality of living indicators point to reduced spending in health care services among beneficiary households.

Programme beneficiaries account for more than half of the 400,000 visits to the Clinic Management System. The share of women who benefited from prenatal and postnatal health care visits increased by 1.6 and 2.3 percent respectively. Nutrition counselling among beneficiary mothers increased by 7.9 percent, while the share of mothers who breastfeed as the main feeding option went from 76.5 percent to 83.5 percent. Likewise, there were increases in the number of monthly prevention check-ups among children under five years of age and in the share of these children with a full vaccination scheme by 4.8 and 3.0 percent respectively (see Table 1).

Education of beneficiary household members: The Solidaridad Programme has had positive impacts not only on the education of school-age children, but also on all household members. The share of children who received the School Attendance Bonus and completed a school year increased by 0.7 percent. Perhaps the most important impact is the 34.7 percent rise in continued school attendance among beneficiary children. This has overall benefits for

TABLE 1: OUTCOMES OF THE SOLIDARITY PROGRAMME ON HOUSEHOLD HEALTH INDICATORS

INDICATOR	CHANGE 2010–2011		DIFFERENCE (OUTCOME FROM SOLIDARIDAD PROGRAMME)
	SOLIDARIDAD PROGRAMME BENEFICIARIES	NON-BENEFICIARIES	
Prenatal check-up in last pregnancy	1.6	-2.7	4.3
Postnatal check-up in last pregnancy	2.3	-2.8	5.1
Mothers who get nutritional counselling	7.9	1.9	6.0
Mothers who get nutritional counselling by a physician	20.6	17.1	3.5
Regular healthy child check-ups during the last month	1.9	-10.9	12.8
Children from 18 to 60 months of age with a full vaccination scheme	3.0	-0.7	3.7
Last healthy child check-up in a Primary Health Care Centre	9.4	4.2	5.2
Last sick child visit in a Primary Health Care Centre	9.8	0.8	9.0

TABLE 2: CONTRIBUTION OF THE SOLIDARIDAD PROGRAMME TO POOR HOUSEHOLD INCOMES

TRANSFERS AND SUBSIDIES	SHARE OF THE MINIMUM WAGE	SHARE OF THE OVERALL HOUSEHOLD BASKET OF THE BOTTOM QUINTILE	SHARE OF THE FOOD BASKET OF THE BOTTOM QUINTILE
Transfers (<i>Comer es primero</i> + School Attendance Bonus+ Old Age Support) and Subsidies (Gas and Power Voucher)	31%	19%	40%
Transfers <i>Comer es primero</i> + School Attendance Bonus+ Old Age Support	21%	13%	27%
Transfers <i>Comer es primero</i> + School Attendance Bonus	15%	9%	20%

all household members, as literacy among Solidaridad Programme beneficiaries increased by 0.6 percent, and average schooling years went from 4.5 to 4.8 in only one year.

Other improvements are apparent in indicators related to net enrolment rates, school attendance, literacy rates and school-year completion. Enrolment rates among beneficiary children exceed those of non-beneficiary children by 2.4 percent, which suggests that the Solidaridad Programme has prompted poor beneficiary households to prioritize sending their children to school. Continuous school attendance among children beneficiaries of the School Attendance Bonus is 5.7 percent higher than the rate among non-beneficiary children, which suggests that the Solidaridad Programme has also influenced school dropout rates.

Poor household incomes: Conditional cash transfers, *Comer es primero*, and the School Attendance Bonus have had a smoothing effect on household budgets, helping households cope with existing deprivations. These benefits account for 20 percent of the food basket costs for the bottom income quintile. Overall, poor households receive an average of \$50 in cash transfers and subsidies, which accounts for 31 percent of the minimum wage (equivalent to 19 percent of the overall family basket). This improves food security and provides a major complement to household income (see Table 2).

OUTCOMES OF PROGRAMME OPERATIONAL IMPROVEMENTS

- Strengthening cross-institutional links among Sistema Único de Beneficiarios (SIUBEN), Administradora de Subsidios Sociales (ADESS) and Progresando with the Solidaridad Programme for follow-up.
- Improved computer facilities and technical teams in the Programme’s provincial offices through training workshops.

- Design and development of the Integrated Citizen Service System to be used as a tool for electronically managing requests and complaints made by Dominican Social Protection System beneficiaries.
- Design of the automated reporting in Joint Responsibility Verification Processes in Education, Health and Documents.
- Training workshops for more than 100 users of the Solidaridad Programme in the use of the information system, improved version, support tools for decision-making and other features that support the programme's decentralized activities.
- Design, development and implementation of features supports the management of the new Solidaridad Programme component, Bonoluz.
- The programme has been praised by the Inter-American Development Bank and showcased as a regional and global conditional cash transfer model for its integration of capacity building and improved offers and services for vulnerable and deprived groups.

CONCLUSIONS

The evaluations conducted on the Solidaridad Programme (now called Progressing with Solidarity), have measured its impacts on beneficiary populations with the introduction of measures that improve effectiveness. Incentives and tools have been developed to promote coordination among ministries and government agencies involved in promoting human capital in order to improve spending effectiveness.

The first results of the Solidaridad Programme (conducted after three years of implementation), reveal that it has met the goals expected in the first phase, including strengthening social protection and making a sound, efficient and transparent use of public resources. These outcomes have led to the programme's expansion, improvement and replication throughout the Dominican Republic.

The following conclusions emerged as part of the evaluations:

- The programme invests in improving the quality of education and services in public health care facilities;
- The programme supports strengthening health care facilities, especially service quality included in the Ministry of Health's Basic Health Care Plan. The programme provides funding for primary care units and training efforts for health care professionals, in addition to actions that improve the overall quality of M&E. To achieve this goal, the programme has supported the training in health care guidelines and protocols for more than 2,700 physicians and nurses, and has computerized all of the country's health care facilities. Consequently, each patient has a personal electronic medical record;
- Child and youth education have been considerably improved by the programme. Ambitious goals in this area include the completion of, at least, the lower secondary

level and the access to improved quality schooling for all extremely poor youth. This will help break the inter-generational poverty cycle and guarantee better opportunities for these children compared to those available to their parents; and

- The poverty gap was partly closed, and local structures were strengthened at the local, provincial and regional levels. The management, financial, planning and M&E systems of the programme were also improved.

In sum, the Solidaridad Programme continues to be a benchmark in the development of conditional cash transfer programmes in Latin America and the Caribbean. The Solidaridad Programme is a safe bet to reduce poverty and improve the living conditions of the most vulnerable groups of the population.

14. MEXICO

OVERCOMING MISTRUST IN POVERTY AND EVALUATION FIGURES: THE CASE OF CONEVAL IN MEXICO

GONZALO HERNÁNDEZ LICONA, Executive Secretary
AGUSTÍN ESCOBAR LATAPÍ, Academic Researcher

National Council for Social Development
Policy Assessment (CONEVAL)

INTRODUCTION

Usually, Latin American governments are not reliable about the information they deliver. Why? Because usually, Latin American governments lie about the information they deliver. We can modify the previous statement by saying that this is not the case for all Latin American governments in all periods, or that governments from other continents have also lied; this is true. But it seems difficult to disagree with the fact that, when it comes to official figures, the first reaction of citizens in many Latin American countries is disbelief. This holds for a range of indicators, including inflation, poverty, electoral results or the total wealth of politicians according to their public asset declarations.

The problem with lying even once (and being caught) is that it may take ages and a lot of effort to restore credibility. Perhaps for this reason, mistrust of official figures is difficult to eradicate.

However, it is also true that several things have improved in many Latin American countries, especially since the arrival of democracy. One of those changes, albeit slow, has been accountability. Authoritarian regimes have, by definition, more power to say whatever they want with almost no pressure from the society.

"What's the time right now?" asked the General.

"Whatever time you wish, General," replied the faithful assistant.

Direct democracy, and even today's social networks, sometimes transforms society into a more active one (and in a way this also means democracy), which increases the pressure on governments to be more transparent and accountable.

This paper is about one example of the effect of democracy on the credibility of poverty figures and social evaluation reports in Mexico. This story is a good example of overcoming mistrust of official figures. We cannot guarantee this will remain the case forever, because political phenomena change and transform suddenly, and we cannot predict what will happen. We can only examine the story as of today, and we hope that it will still be a success of transparency and accountability for coming decades.

SEVEN DECADES OF MEXICAN POLITICS

There are many things we can say about the period between the Mexican Revolution (1910 to 1921) and 1997. But here, we focus on one: not only did a single party control the government during this period, but, more importantly, this same party had a clear majority in congress. This means that the president had very few political obstacles. It might be difficult to establish that this was an authoritarian regime (we had elections, after all), but it is true that the degrees of freedom the president had were large. Congress could not—and did not—act as a check on the executive's power.

"Yes, Mr. President, it is 2 o'clock right now, if you say so."

This means, among other things, that there was no pressure to check whether official figures were the right ones, or if some figures were hidden from the public. The statistical office and the central bank worked under the supervision of the president. Civil society had neither enough power nor the organizational skills to ask for true figures or to question the ones published.

But things changed in 1997. The most important one for this paper was due to several changes in the electoral process. For the first time since the revolution, the party of the president did not have a majority in congress. For the first time, Mexico had a political balance of power between congress and the executive.

Congress began to demand transparency from the president. It is difficult to imagine today's institutions without an active congress: Due to these political changes, congress created the National Council for Social Development Policy Assessment (CONEVAL) in order to have more reliable figures for poverty and evaluation reports.

WHY THE CREATION OF CONEVAL?

The origin of external evaluations

As early as 1999, the elections to be held in 2000 were expected to be very tight. In the past, the government had resorted to clientelistic discretionary spending to win elections.

Opposition political parties in congress were asking for stricter control on the use of public resources by the executive for electoral purposes. One of those controls was to have external evaluations for all social programmes of the executive. The main idea was that an external observer could confirm that programmes were performing correctly and that fund allocation was driven by technical criteria, such as poverty or malnutrition, and not by political concerns. For the first time in Mexico's history, there was a clear mandate for external evaluations.

The election in 2000 was so tight that a new political party actually won the election. The majority of parties in congress did not belong to the new president, and Mexico had a plural congress (which it still has today).

The origin of official poverty estimates

In terms of poverty figures, before 2000 there were no official estimates; the new government decided, rightly, to determine poverty figures. For this purpose, the government hired excellent academics to elaborate a methodology to estimate poverty. With this method, official figures were released in 2002, which showed that poverty was reduced between 2000 and 2002. In 2004, the government announced that poverty had again fallen.

The power of congress

Despite these two achievements—the existence of external evaluations and the creation of an official poverty measurement—political parties in congress and opinion leaders did not believe in either (the reduction of poverty took place during an economic downturn). Due to the mistrust of social development figures (and other rationale), congress designed an independent institution. In 2004, congress issued the General Social Development Law. The law states that a new independent institution, CONEVAL, is responsible for both coordinating the evaluation of social programmes and creating a new method to estimate poverty at the national, state and municipality levels.

Congressional power in terms of accountability reaches beyond the creation of CONEVAL. Today, Mexico has an autonomous statistical office, an independent office of Transparency and Access to Information, a similar electoral institute, a quite rigorous law for budget and fiscal responsibility, and an autonomous central bank (the latter since 1994).

CONEVAL

CONEVAL was created in 2006 as a result of the Social Development Law. Perhaps the most important element of the institution is the way congress structured its governance bodies. CONEVAL has a board with eight seats; two belong to officials from the Ministry of Social Development and the Finance Ministry, and six are for independent researchers who are elected by a body of 44 votes, where the states, congress, municipalities and the executive are represented.

This means that the main decisions on the evaluation of social programmes and the measurement of poverty are taken, in fact, by the majority of researchers who are elected, not appointed by the president or by the Minister of Social Development. This characteristic

has given CONEVAL effective independence from both the executive and congress.

Since 2006, CONEVAL has published around 1,500 evaluations of specific federal social programmes, a number of comprehensive reports of the state of social development policy, two reports analysing M&E at the state level, and five sets of poverty estimates spanning federal, state and municipal levels.

CONEVAL estimated that poverty decreased nationally between 2004 and 2006, but it increased between 2006 and 2010.⁷⁰ The latest set of figures, for 2012, shows a decrease in the percent of poor persons, but a small increase in their absolute number.

The way CONEVAL proceeds in order to assure citizens that the figures are credible is as follows: The basic information is taken from a household survey produced by the autonomous statistical office. The entire survey (including the micro data) is public. CONEVAL applies the poverty estimation programme to the survey (the publicly available programme can be run on STATA, SPSS and R statistical packages). With this procedure, everyone can replicate the figures. Anyone can check for themselves whether the poverty formula may suffer any change.

CONEVAL has hired external institutions twice (Mexican universities, poverty institutes and UNDP) to carry out independent mirror runs of the poverty estimates before publication. CONEVAL felt that these mirror runs, in addition to helping detect mistakes, would also serve to certify that the measurement did what it was meant to do without any sleight of hand. Today, however, there are hundreds of 'mirrors' all over the world, and individuals routinely carry out modifications to check what would happen to a measurement if they modify the script. CONEVAL provides advice to anyone wishing to test the programme or to perform additional statistical tests and calculations.

Our hypothesis is that without institutionalizing this process, and without a majority of independent researchers on CONEVAL's board, it would be difficult for the government to accept the release of these figures, especially when poverty rises.

MANAGING THE PROCESS OF CREATING A RELIABLE EVALUATION SYSTEM

The same is true for the evaluation of social programmes. CONEVAL believes that programme evaluations have two important objectives: programme improvement and accountability. Taxpayers have the right to know if the social programmes they finance have results or if improvements are needed. For this reason, the evaluation reports coordinated by CONEVAL publicly show programmes' strengths and weaknesses. Without the autonomy of CONEVAL, it would be unlikely that governments would willingly publish evaluation reports that include programmes' flaws. This actually happens in countries and regions where the evaluation process is controlled by the president or local governments.

The level of independence provided by elected academics is certainly a factor in the increasing international recognition of CONEVAL. However, CONEVAL has also managed the process of strengthening the credibility of information. In 2007, CONEVAL detected a

70 See [coneval.gob.mx/Medicion/Paginas/Medición/Pobreza 2012/Pobreza-2012.aspx](http://coneval.gob.mx/Medicion/Paginas/Medición/Pobreza%202012/Pobreza-2012.aspx).

number of earlier evaluations in which consultants had written the same text to evaluate two different programmes over the course of three years—one text for six evaluations, six contracts, six large checks.⁷¹ No one had noticed until a single, technically competent, independent agency, such as CONEVAL, took charge of the evaluation system. CONEVAL had to ask a ministry to refrain from working with some evaluators, but essentially we had to convince evaluators, the evaluated and the public that the process was in fact rigorous, transparent and useful: that the rules had changed. Evaluation cannot proceed without everyone's cooperation.

Initially, CONEVAL restricted the freedom of both public officials and consultants. It produced very restrictive evaluation syllabuses with very little room for interpretation. There were two objectives: to lay a level ground for all programmes, and to convince everyone that spinning the facts into 'acceptable' assessments (or not doing the work) was no longer allowed.

At the same time, public officials and consultants were only allowed to use documents that had been posted on a special Web page. Positive or negative assessments had to result from documentary evidence that could be checked by anyone involved. In keeping with its transparency policy, CONEVAL initially designed this Web page as entirely public, but ministries noted some documents were restricted and public officials demanded protection, so access to the page was restricted. Later, however, many more documents were made publicly available.

Because CONEVAL had insisted that even the simplest of evaluations be performed by top academics specializing in the appropriate fields, our evaluation staff had to manage tensions between the three main parties involved. Public officials often felt consultants were not adequately expressing certain positive aspects of evaluated programmes. Academics working as consultants felt that they needed more freedom to express their expertise and their opinions. Both felt CONEVAL was placing a straitjacket on evaluations. CONEVAL discouraged private meetings between them.

We often heard complaints in the sense that evaluation was just another bureaucratic hurdle providing nothing useful. Because communication between officials and consultants was often formal and tense, officials sometimes blacklisted good academics in order to bar them from evaluating their programmes ever again. In addition to being patently unfair, blacklisting further narrowed the very small evaluation market in Mexico. Fortunately, the situation has evolved.

Evaluations can be funded and hired by ministries or by CONEVAL. Ministries have, in fact, retained their evaluation budget. Nevertheless, they have to plan their evaluations and submit their terms of reference documents to CONEVAL, which then reserves the right to oversee the process and to participate in progress meetings. Also, all evaluations hired and paid directly by CONEVAL are not only overseen by CONEVAL, but also all meetings between consultants and federal agencies take place in the presence of a CONEVAL representative. This

71 The minister receiving these evaluations had complained that he believed they were not useful. Clearly, if the procedures that produced the evaluations were faulty, the resulting evaluations were, in fact, useless.

has helped reduce the occurrence of private agreements between consultants and agencies and has avoided the appearance of false allegations against evaluators who honestly point at significant areas of opportunity or aspects requiring urgent attention.

This is not a definitive solution; powerful ministries can still influence evaluators. CONEVAL staff read all evaluations (approximately 160 annually) to suggest improvements and, if necessary, to point out inconsistencies among evidence, analysis and conclusions. Negotiation is necessary in some cases, when ministries would rather not publish an evaluation. Nevertheless, we believe the situation is entirely different from that of 2006. Evaluators feel they are backed by an institution that asks them to be constructive but critical whenever necessary. Evaluators are currently far more independent. Institutions, on the other hand, feel they have an arena in which their opinions matter, if correctly argued. Institutions can be supported by CONEVAL if they have a case. A positive byproduct is that ministry staff has to acquire and exert skills in creating and using evaluation information.

Because CONEVAL knows all the heads of evaluation teams, it has become a powerful market agent. A CONEVAL recommendation is highly valued, and we believe this has influenced the quality and impartiality of evaluations in a positive manner. Nevertheless, we acknowledge the fact that the market still needs to grow much more, particularly because state governments are now significant clients and skilled evaluation teams have become scarce.

CONEVAL does as it preaches: it has asked consultants and officials to assess the evaluation process. Feedback has resulted in substantial improvements, and CONEVAL's marks have risen continually among both. CONEVAL has also carried out a series of meta-evaluations in order to highlight areas that are in need of improvement. These meta-evaluations have also helped modify evaluation processes.

Institutionalizing evaluation processes was possible because of the creation of an independent institution with a strong mandate, but two other elements should be acknowledged. First is the full respect the executive has had for CONEVAL work. We have not perceived, so far, an attempt to convince CONEVAL to change figures or ideas. The second element is the high professional standards for the social programme in charge of the evaluation process. We believe that it would have been difficult to finish and use so many evaluations without support from the programmes themselves.

CONCLUSION

Democracy in the form of a balance of power between congress and the president contributed to the credibility of information in the area of social development in Mexico. But the task is not done. There is still more to achieve among state and local governments, where independent evaluations and transparency, with very few exceptions, have not yet arrived.

We also have to be aware of changes in the future. Political powers can sink any vessel. We hope civil society can keep an eye on this one.

15. MEXICO

EVALUATION USE AND ITS INSTITUTIONALIZATION IN THE FEDERAL EVALUATION SYSTEM IN MEXICO

GABRIELA PÉREZ-YARAHUÁN

Visiting Scholar, Center for Learning and
Evaluation Research (CLEAR) Latin America, and
Professor, Universidad Iberoamericana

INTRODUCTION

A key element of improving the quality of a democratic government is the adequate performance of its public sector's interventions. In this respect, a country's evaluation capacities play a significant role, as evaluation is central to informing debate on policy decisions, to improving programmes' results and to enhancing public accountability (Chelimsky 2009, Independent Evaluation Office of UNDP et al. 2011). The Mexican Government and civil society have engaged in important efforts to consolidate and strengthen institutions for government evaluation, transparency and accountability during the past decade.

Though there are still important challenges, there has also been unquestionable progress. It is important, then, to point to fundamental characteristics in the development of the Mexican Federal Government's social programme evaluation system and to describe why and how evaluation studies have affected programmes' changes and decision makers' perceptions of evaluation importance and its use.

THE MEXICAN SOCIAL PROGRAMME EVALUATION SYSTEM

The first nationwide and internationally well-known social programme evaluation studies in Mexico were those of the Programme for Education, Health and Nutrition, today known as Oportunidades. This was one of the first conditional cash transfer programmes implemented

on a national scale in Latin America. National and international experts were involved in the evaluation design and study. Several documents regarding this programme's evaluation have circulated extensively.⁷²

The fluster generated by the Programme for Education, Health and Nutrition's external evaluations triggered a national interest in the evaluation of social programmes in Mexico within an environment of increasing demands for government accountability. Access to information and government evaluation became newly acquired undertakings for the new party in government in 2000. But it also became an issue of interest for the former political elite, then the opposition, as there were new incentives to search for mechanisms for control and oversight.

In 2001, the Annual Federal Budget Bill included the requirement that all federal subsidy programmes present an annual external evaluation to serve as an assessment of the programme's implementation and its compliance to its operation rules, an analysis of the results on its stated objectives and a cost-effectiveness analysis. From 2002 to 2006, 326 evaluations studies were conducted, an average of 80 programmes a year. However, the use of the evaluation reports to improve on programme design and operation was largely ignored or undocumented.

The General Law for Social Development, enacted in 2004, contained several guidelines regarding evaluation use. According to the law, the purpose of evaluation was to periodically review the fulfilment of social programmes' objectives, targets and actions, in order to correct, modify, reorient or suspend the programmes. In its article 80, the law stated that, based on evaluation results, the National Council for Social Development Policy Assessment (CONEVAL) was to elaborate suggestions and recommendations to the Federal Executive's programmes, and that these suggestions were to be made public.

The Annual Federal Budget Bill established that evaluation studies were to be used for the process of analysis and discussion of the Budget Bill. However, there was no precise mechanism on how the information was to influence budget discussions. Furthermore, there was no incentive to motivate the use of evaluation reports to promote any programme change. It is important to say that evaluations varied greatly in their quality, extent, information and methods—and even in the issues that were discussed and analysed.

In 2006, the new Federal Law of Budget and Financial Responsibility gave permanence to many of the established guidelines of the Annual Budget Bill regarding evaluation implementation and procedures. With respect to evaluation use, the new law was equally brief and ambiguous, establishing only that agencies were to follow up on the recommendations of evaluations and link to the Performance Evaluation System and budget decisions.

In March 2007, the Department of Finance, the Department of Public Service and

72 For example: (1) Coady, David P. 2000. "Final report: The application of social cost-benefit analysis to the evaluation of PROGRESA". November. Report submitted to PROGRESA. Washington, DC: International Food Policy Research Institute. (2) Behrman, Jere R., and Petra E. Todd. 1999. "Randomness in the experimental samples of PROGRESA (education, health, and nutrition program)". February. Report submitted to PROGRESA. Washington, DC: International Food Policy Research Institute. (3) Gertler, Paul J. 2000. "Final report: The impact of PROGRESA on health". November. Report submitted to PROGRESA. Washington, DC: International Food Policy Research Institute.

CONEVAL jointly issued the General Guidelines for the Evaluation of the Federal Public Administration Programmes. These established a set of rules with the intention to clarify and give a more permanent and viable structure to the programme evaluation system. The guidelines served to clarify three important aspects regarding evaluation and its use. First, the document established different types of evaluation studies that were to be conducted for federal programmes; it also established the obligation of the federal government to issue an annual evaluation programme that enumerated types and subjects of evaluations assigned within the federal administration's programmes. Second, it established the requirement for evaluation dissemination and publication by all federal agencies in charge of programmes subject to evaluation. And third, it confirmed the compulsory use of evaluation results by federal agencies on those aspects conducive to enhancing programme performance.

Since then, a new focus on evaluation use has been evident in the legal and administrative frameworks. Aspects such as communicating and publishing evaluation results are present, as well as the intent to make evaluations more applicable and relevant to programme improvement. However, even though the General Guidelines for Evaluation issued in 2007 confirmed the required use of evaluations, there was no clear process to make this requirement effective. One of the main problems perceived was that evaluations' recommendations entailed the joint efforts of diverse actors within federal and state governments, which most often did not take place. It is also important to recognize that there is a clear need for evaluators and other stakeholders to interact in order to transform evaluation findings into real programme improvement.

Independent academic research that used a sample of social federal programmes between 2002 and 2008 found little relationship between changes in a programme's operational rules and evaluation improvement recommendations (Pérez-Yarahuán 2012).

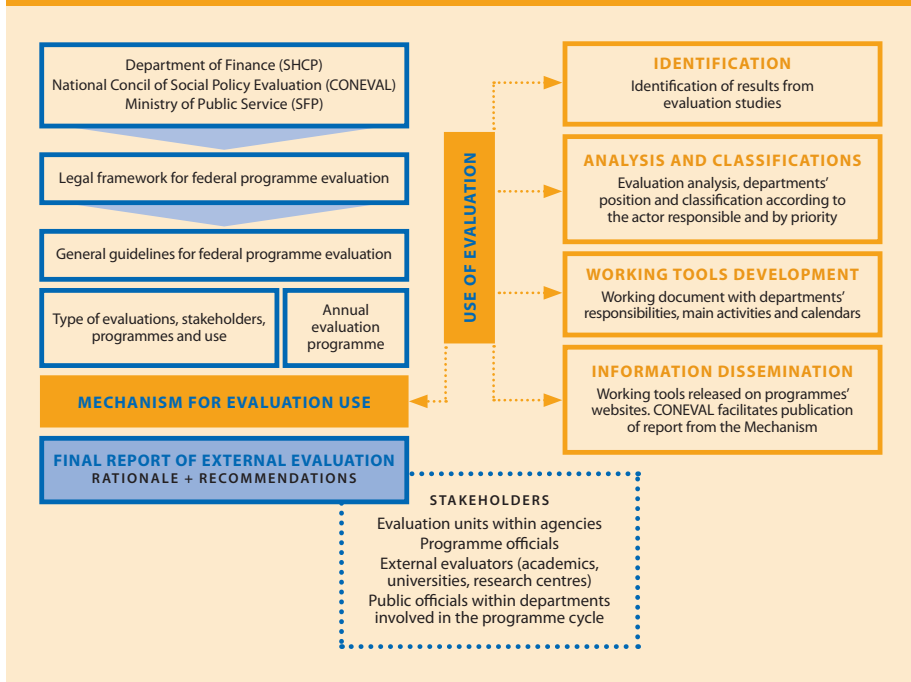
INSTITUTIONALIZING EVALUATION USE

By October 2008, officials from the Department of Finance, the Department of Public Service and CONEVAL, aware of this problem, issued a formal mechanism for federal programmes and their corresponding agencies to follow up on programme evaluation reports findings.⁷³ This mechanism is a process by which the main evaluation stakeholders analyse and comment on the evaluation reports and findings and then propose specific actions to enhance programme performance. This process is public and transparent through the publication of documents on federal agencies' Web pages. CONEVAL publishes an annual report online that shows the agencies and programmes that followed up the process, the type of improvement actions that were to be undertaken and the stakeholders involved.⁷⁴ This report puts in motion actions that involve federal and state governments and that are reviewed in the Intersecretarial Commission for Social Development.

73 SHCP, CONEVAL, SFP, *Mecanismo de Aspectos Susceptibles de Mejora*, 14 de octubre, 2008, Oficio No. 307-A-1981, No. VQZ SE 273/08, No. SCAGP/200/821/2008.

74 CONEVAL, *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de los Programas y Acciones Federales de Desarrollo Social 2012/2013*. México, CONEVAL, 2013. Available at web.coneval.gob.mx/Informes/Evaluacion/Informe ASM/Informe ASM 2012-2013.pdf.

FIGURE 1: MECHANISM FOR THE FOLLOW-UP ON PROGRAMME EVALUATION REPORTS FINDINGS WITHIN THE CONTEXT OF THE NATIONAL EVALUATION SYSTEM IN MEXICO



The formal mechanism, renewed each year, puts in motion a process by which different stakeholders have to become familiar with the evaluation studies, have to engage in active debate regarding evaluation findings and have to propose specific courses of actions. This process permits a follow-up on evaluation findings.

Empirical research on evaluation use and public officials' perceptions shows that evaluation use has increased in Mexico and that federal public officials at different levels of responsibility are aware of government programme evaluations being performed each year by independent evaluators. Public officials have started to perceive evaluation as an instrument for specific programme improvement, but unfortunately less so as a public accountability instrument (Pérez-Yarahuán 2012). Evaluation use has improved in the government sector, particularly in the executive federal branch and in professional civil society organizations. But evaluation use needs to extend to congress and to the citizens in order for the evaluation effort to realize its full potential and purpose.

In this respect, the evaluation process needs to strengthen its communication strategies, particularly those focused towards other non-governmental stakeholders. Since 2008, CONEVAL's communication strategy has not had much change. In this respect, the

quasi-independent nature of CONEVAL (its budget and some of its appointments depend on the federal executive), may pose some obstacles to increasing its outreach and influence on policy debate, thus affecting potential use.

CONCLUSIONS

During the past decade, the evaluation system in Mexico has been steadily institutionalized into a set of rules, organizations and procedures by which hundreds of programmes are effectively evaluated each year. Evaluation activity in Mexico has been perceived as a highly legitimate task under a political climate of democratic transition, and in which instruments for attaining accountability and control of those holding power is deemed a *sine qua non* for political change. An important challenge, then, is to assure that evaluation efforts are used in fact for different purposes, such as debating public policy, improving government programmes and enhancing public accountability.

This paper described the mechanism by which the federal evaluation system in Mexico has been formed and strengthened. Today, evaluation appears to be highly ranked as an instrument for government change among public officials in the federal government, particularly so for the change of rules and the acquisition of information about programme performance. This use is perceived by government actors within the executive. The perception of the use by actors outside the government, such as congress, beneficiaries or citizens, remains yet to be achieved.

REFERENCES

- Chelimsky, Eleanor. 2006. "The Purposes of Evaluation in a Democratic Society". *The Sage Handbook of Evaluation*. London: Sage Publications.
- Independent Evaluation Office of the United Nations Development Programme and the Public Service Commission of South Africa. 2011. *Proceedings from the Second International Conference on National Evaluation Capacities*. Johannesburg, South Africa.
- Secretaría de Hacienda y Crédito Público (SHCP), National Council for the Evaluation of the Social Development Policy (CONEVAL) and Secretaría de la Función Pública (SFP). 2008. *Mecanismo de Aspectos Susceptibles de Mejora*. Oficio No. 307-A-1981, No. VQZ SE 273/08, No. SCAGP/200/821/2008.
- CONEVAL. 2013. "Informe de Seguimiento a los Aspectos Susceptibles de Mejora de los Programas y Acciones Federales de Desarrollo Social 2012/2013". México, CONEVAL. Available at web.coneval.gob.mx/Informes/Evaluacion/Informe_ASM/Informe_ASM_2012-2013.pdf. Last accessed 5 February 2014.
- Pérez-Yarahuán, Gabriela. 2012. "Politics, Accountability and Evaluation: Toward Institutionalization and Utilization of the National Evaluation System in the Social Policy Sector in Mexico". Paper presented at the 2012 Meeting of the Midwest Political Science Association Meeting, Chicago Illinois. More information available at mpsanet.org/Conference/ConferencePaperArchive/tabid/681/q/Perez/year/2012/Default.aspx - sthash.oVP239pt.dpuf. Last accessed 5 February 2014.

16. MEXICO

INCENTIVES IN THE USE OF EVALUATIONS AS A TOOL FOR IMPROVING PUBLIC POLICY

THANIA DE LA GARZA, Adjunct General Director of Evaluation
RASEC NIEMBRO, Monitoring and Evaluation Specialist

National Council for Social Development Policy Assessment (CONEVAL)

INTRODUCTION

Given the current challenges of social and economic inequality and poverty faced by developing nations, governments must continuously improve public policies in order to secure the conditions for the population to exercise their social rights.

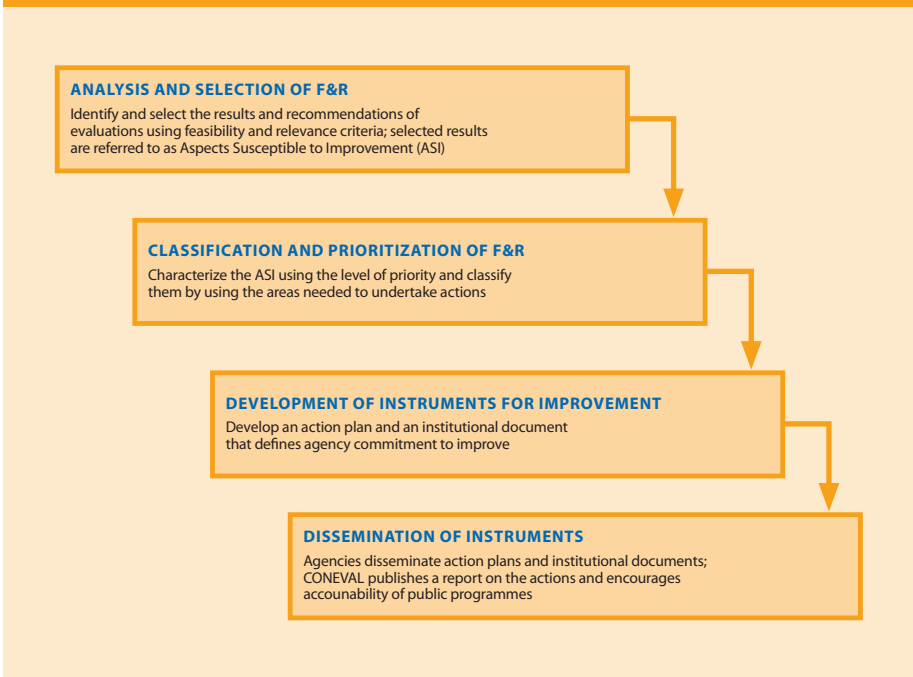
One of the primary tools adopted by the Mexican federal government is the use of evaluations as a key element of improving government performance and measuring the results of public action. Therefore, since 2008, Mexico has built an enabling environment with one principle: to create the right incentives for the programmes' operators to use the evaluations.

For this reason, the National Council for Social Development Policy Assessment (CONEVAL), the Secretary of Finance and Budget, and the Secretary of Public Administration issued the Follow-up Mechanism for the Use of Evaluations, with the objective of institutionalizing the use of external evaluations' findings and recommendations. This paper aims to demonstrate the positive incentives created by the Follow-up Mechanism for using evaluations to improve social programmes in Mexico from 2008 to 2012.

M&E IN THE MEXICAN SYSTEM

In 2005, the General Law of Social Development established the creation of CONEVAL, a public organization with autonomy and the technical capacity to generate objective information on the social policy in order to improve public programmes, decision-making and accountability. The main functions of CONEVAL are to conduct national, state and municipal

FIGURE 1. FOLLOW-UP MECHANISM FOR THE USE OF EVALUATIONS



Source: Based on CONEVAL, 2008, 2009, 2010-2011, 2012-2013.

poverty measurements and to regulate and coordinate the evaluation of social development policy and programmes implemented by public agencies.

In order to achieve its evaluation responsibilities, CONEVAL has established, in partnership with the Department of Finance and the Department of Public Service, an M&E system. The purpose of the system is to generate information from the results and the overall performance of government programmes. Since institutionalizing the M&E system, CONEVAL identified the absence of clear incentives for using evaluation results. The evidence of improvement of public policies as a consequence of evaluations was almost non-existent.

FOLLOW-UP MECHANISM TO EXTERNAL EVALUATION RECOMMENDATIONS

CONEVAL, the Department of Finance and the Department of Public Service designed the Follow-up Mechanism for the Use of Evaluations⁷⁵ to institutionalize the process of

⁷⁵ The Follow-up Mechanism for Use of Evaluations identifies aspects that are susceptible to improvement. These include the findings, weaknesses, opportunities and threats identified by the evaluator, the external evaluation and/or reports that can be addressed to improve the programme.

FIGURE 2. EVOLUTION OF THE PARTICIPATION OF GOVERNMENT AGENCIES IN THE FOLLOW-UP MECHANISM

PERIOD	NUMBER OF SECRETARIES	NUMBER OF PROGRAMMES	NUMBER OF ASPECTS SUSCEPTIBLE OF IMPROVEMENT DEFINED IN THE ACTION PLANS
2012–2013	13	113	641
2011–2012	14	108	413
2009–2010	15	150	871
2008–2009	11	91	930

Source: Based on CONEVAL, 2008, 2009, 2010–2011, 2012–2013.

results follow-up to facilitate evaluation use. The mechanism’s main objectives are to establish a general procedure to monitor improvement areas identified in the evaluations, and to disseminate the commitments made by programme operators to improve public programmes. The mechanism applies to all federal agencies and has an annual implementation cycle.

The Follow-up Mechanism for the Use of Evaluations encourages the participation of several actors, including those in programme units, evaluation units, and programming and budget departments inside agencies. Furthermore, as part of the feedback process, continuous meetings are held between evaluators and programme operators in order to consider the opinions of all stakeholders involved.

As observed in Figure 1, the mechanism is integrated by four stages. In the first stage, stakeholders select aspects they will define as susceptible of improvement by using feasibility, capability and explicitness criteria. Stakeholders do not have to address all the recommendations. In this stage, the main assumptions are that, having the results of the evaluation, programmes operators have more knowledge and information about the most effective actions that can improve the programme, have certainty regarding constraints (e.g. budgeting, human and technological) to define actions to improve, and have the most information about the political arena that will affect the effectiveness of the improving actions.

In the second stage, stakeholders define the responsible areas that must take action in order to improve the programme. Because implementing any programme entails interactions among several areas (e.g. operation, budgeting, targeting), the improving actions cannot be undertaken by a single area.

The third and fourth stages consist of the elaboration of the action plans and its publication on CONEVAL Web pages and those of each agency accountable for the plans. The last stage encourages the operation responsible for performing the commitments defined in

**FIGURE 3. ACTIONS IMPLEMENTED TO IMPROVE
SOCIAL DEVELOPMENT PROGRAMMES**

ACTIVITY	2011	
	ACTIONS	PARTICIPATION
Improve activities or programme processes	163	71%
Modify programme's services or goods	27	12%
Substantially reorient the programme	35	15%
Increase or reallocate the programme	3	2%
TOTAL	228	100%

Source: CONEVAL, 2010-2011.

the action plans to improve the programmes and report the advance of each action during specific dates in the cycle.

RESULTS OF THE FOLLOW-UP MECHANISM FOR USE OF EVALUATION

The Follow-up Mechanism for Use of Evaluation, designed under an incentive scheme, has motivated agencies to take concrete actions to improve programmes. In Mexico, evidence suggests that evaluations are becoming a key factor driving public policy change. In 2011, according to the classification established in the mechanism regarding the priority level (high, medium, low), 55 percent of improvement areas identified were listed as high priority concerning the purpose of the programmes, 32 percent were classified as medium priority and 13 percent as low priority.

In addition, since 2012, CONEVAL has measured the improvements' characteristics using only the aspects susceptible of improvement that were fulfilled at the time. Figure 3 shows the outcomes of the changes implemented in 2011 as a result of using the findings and recommendations of evaluations.

The greatest achievement of this exercise is to show the increase and systematic use of evaluations. However, there are several challenges that Mexico still needs to overcome in order to enhance the culture of the use of evaluations, such as increasing the quality and relevance of the evaluations themselves.

CONCLUSIONS

Countries must develop incentives tailored to their own context in order to achieve the efficient use of evaluations. All actors involved in the evaluation process must be committed to considering the information derived from the M&E process and attaining continuous

improvement in the cycle of public policy. Mexico's methodology includes positive incentives that can be used by other M&E systems. In Mexico, recommendations derived from evaluations are not legally mandatory.

There are two main reasons that explain the mechanism. First, there is the assumption that programme operators know the programme best, and if they are unwilling to adopt the necessary recommendations, there will be fewer possibilities for success. Second, internal and external evaluators can easily undermine the importance of change in the political arena. As evaluation does not often demonstrate these conditions on time, programme operators will be unable to implement proper modifications.

In essence, it is important to find the right incentives that will motivate operators to obtain the necessary information and implement the changes they consider most appropriate. So far, Mexico has provided proper incentives by raising awareness of evaluations and by going public in the media with a call for action in evaluation use. With constant effort, these strategies may continue to be successful and fulfil the expectation of higher participation.

REFERENCES

- Consejo Nacional de Evaluación de la Política de Desarrollo Social. 2013. *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de los Programas y Acciones Federales de Desarrollo Social 2012-2013*. México, D.F. CONEVAL.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social. 2012. *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de los Programas y Acciones Federales de Desarrollo Social 2011-2012*. México, D.F. CONEVAL.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social. 2011. *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de los Programas y Acciones Federales de Desarrollo Social 2010*. México, D.F. CONEVAL.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social. 2010. *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de los Programas y Acciones Federales de Desarrollo Social 2009*. México, D.F. CONEVAL.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social. 2009. *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de los Programas y Acciones Federales de Desarrollo Social 2008*. México, D.F. CONEVAL.
- CONEVAL, the Department of Finance, SFP. 2011. *Mecanismo para el Seguimiento de los Aspectos Susceptibles de Mejora para los Programas de la Administración Pública Federal 2011*. México.
- Ley General de Desarrollo Social*. Diario Oficial de la Federación el 20 de enero de 2004. México.

17. PERU

EVIDENCE-BASED PUBLIC POLICIES: THE EXPERIENCE OF THE MINISTRY OF DEVELOPMENT AND SOCIAL INCLUSION

ANIBAL VELÁSQUEZ VALDIVIA

Lead Evaluation Expert

United States Agency for International Development

URSULA MARTÍNEZ ANGULO

General Director of Monitoring and Evaluation

Ministry of Development and Social Inclusion

FABIOLA CÁCERES PAURINOTTO

Evidence Management Specialist

Ministry of Development and Social Inclusion

INTRODUCTION

The creation of the Ministry of Development and Social Inclusion (MIDIS) in October 2011 marked a milestone in the institutionalization of social inclusion as a state priority for the Peruvian government. It consolidated the beginning of a reform towards an evidence-based development and social inclusion policy that focused on focalization, inter-agency and inter-governmental coordination, and rigorous evaluation of the results.

In this context, MIDIS created the General Direction of Monitoring and Evaluation (DGSE) as the unit in charge of monitoring and evaluating policies, programmes and projects related to development and social inclusion in all levels of the Peruvian government. The mission of the unit is to improve the impact, efficiency, quality, equity and transparency of development and social inclusion policies and programmes and, therefore, to provide citizens and policymakers with the results of evaluations and the evidence generated by the M&E system.

Aware that the use of evidence in social policy is not an automatic event but rather a

process conducted in a political and institutional context influenced by various visions, actors and interests, MIDIS designed an M&E system that emphasizes developing instruments to ensure that the evidence generated not only fulfils the highest quality standards, but also is translated in a learning source to continuously improve development and social inclusion interventions.

THE EXPERIENCE OF MIDIS IN THE USE OF EVIDENCE

Traditionally, M&E systems have been focused on the timely production of high-quality and independent information and evidence. Those efforts have resulted in great improvements on the professionalism and academic rigour of the scientific evidence generated about the efficiency and efficacy of public interventions on social issues. It is increasingly common to identify a crisis in the use of evidence in public policy decisions. This situation suggests that even though the generation of evidence is a necessary piece of the puzzle, it is insufficient to ensure informed public policies capable of improving the quality of social expenditure, because the use of evidence depends on multiple factors that can increase or decrease the probability of influencing public policy.

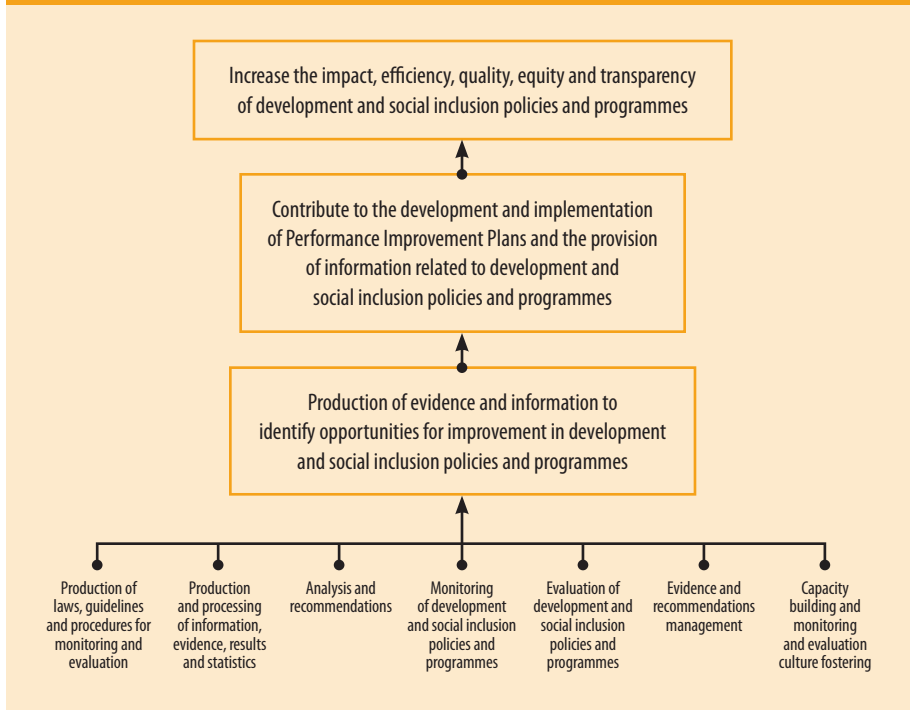
According to Weiss (1999), there are four factors that influence the use of the results of evaluations: interests, ideologies, institutional culture and information sources. If the evidence generates conflicts with any of those four factors, the probability of being used in public policy decisions decreases considerably.

From a different perspective, Innvaer (2012) argues that the ‘two-community thesis’ explains the limited use of evidence in public policy decisions. According to the author, there is a collision between science and politics that hinders the use of evidence in decision-making processes. Innvaer argues that scientists see themselves as rational, objective and innovative, and perceive politicians as interest-driven actors immune to innovation and scientific evidence. In contrast, politicians and policymakers see themselves as responsible, action-oriented and pragmatic leaders and perceive scientists as naïve and commonly disconnected with pragmatic reality. Innvaer suggests that, in order to promote the use of evidence in public policy decisions, it is mandatory to create conditions that increase the probability of use. This can be done, for example, by fostering spaces of dialogue and interaction between science and politics. However, in our opinion, the responsibility for promoting further dialogue between both fields cannot be placed on either side—a third actor is necessary to intermediate and facilitate dialogue.

In designing the M&E system in this context, MIDIS considers DGSE both as a unit that generates high-quality evidence and as a mediator between evaluators and policymakers. In that sense, through the dual role of DGSE, MIDIS intends to close the gap between the scientific world of academia and the policymakers’ perceptions of reality, translating evidence into clear, timely and viable recommendations that can be understood and, mostly, used by policymakers.

Therefore, unlike the traditional M&E systems, the DGSE concept constitutes an innovation; the final goal that justifies its existence goes beyond the production of information and evidence and aims to improve the impact, efficiency, efficacy, quality, equity and

FIGURE 1: DGSE LOGICAL FRAMEWORK

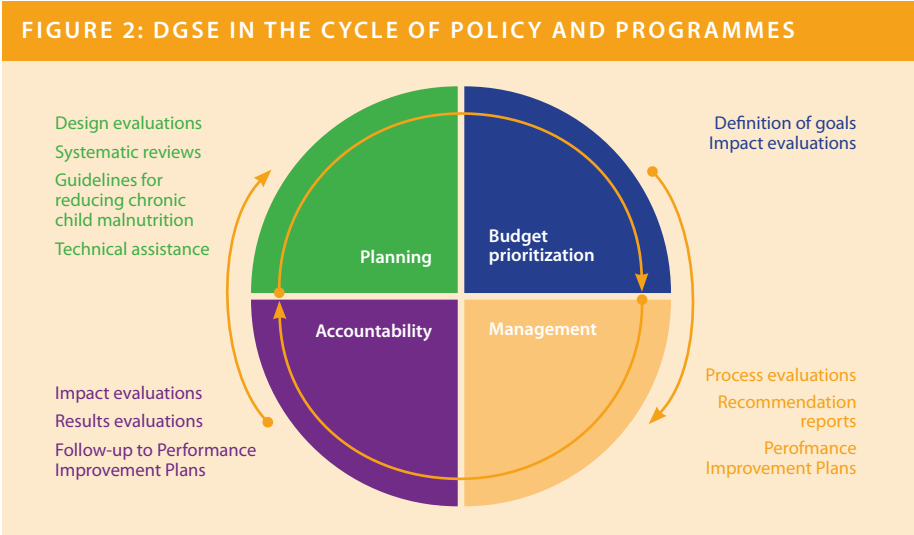


transparency of social and development interventions.

As shown in the logic framework developed by DGSE, the unit has conceptualized several products and services to produce systematic information and evidence regarding opportunities to improve social interventions (see Figure 1). Those results will allow DGSE to contribute to the development of performance improvement plans and to inform about the effectiveness of social policies and programmes with the objective of increasing the impact, efficiency, quality, equity and transparency of governmental interventions. The design of the DGSE model was validated and supported by all of MIDIS's internal stakeholders, who made several contributions to the 'MIDIS's Guidelines for Monitoring, Evaluation and Use of Evidence'.

In order to ensure evaluations' independence, impact evaluations of MIDIS's social programmes and policies are funded by either the Ministry of Finance or multilateral agencies. However, DGSE actively participates in evaluations and works closely with the Ministry of Finance, providing technical assistance to ensure the quality of the evidence generated.

In particular, it is important to note that the most innovative element of DGSE regarding traditional evaluations units relies on the evidence and recommendations management component. This specific line of action influences Performance Improvement Plans and the



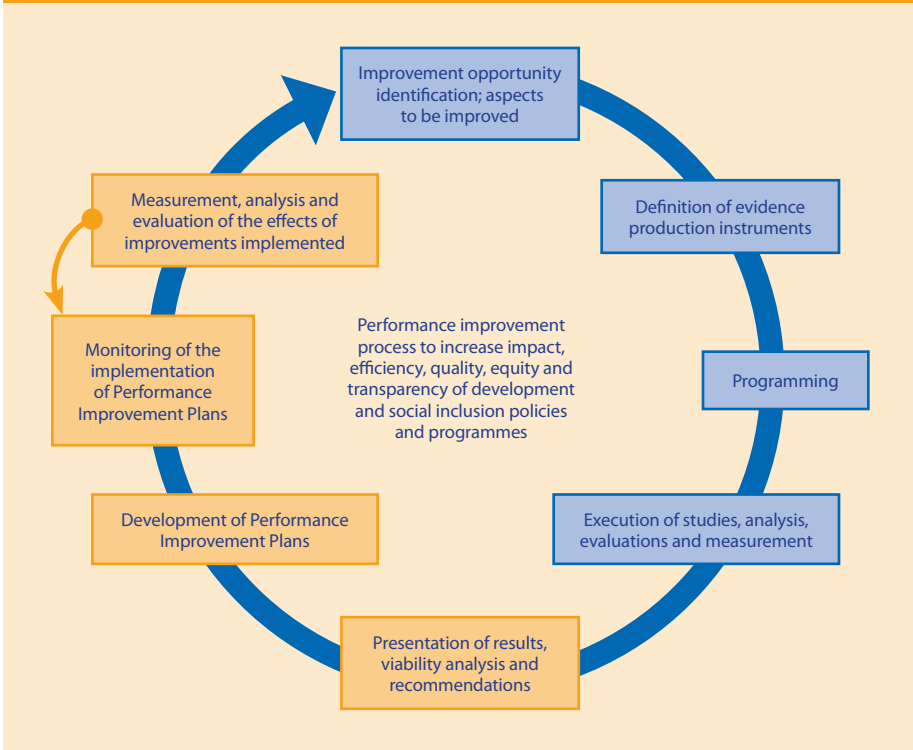
use of evidence. Therefore, it allows evidence to translate into greater impact, efficiency, quality, equity and transparency of public interventions on social and development issues.

DGSE has evolved from the production of information to the use of evidence and has learned that what justifies the existence of an M&E system is the final goal: improving the quality, efficiency, efficacy and equity of policies, programmes and services through evidence and results-based management. In this context, DGSE is determined to be a part of the decision-making process of policies and programmes, providing timely, reliable evidence and information in order to feed into planning, design and operational decisions.

DGSE's model is not only centred in the evaluation cycle, but also considers the political context and the policy, programme and public administrative system cycles (see Figure 2). Therefore, DGSE provides evidence and information to programmes and organic units according to the phase of the cycle they are going through. For example, in the case of programmes in a design or redesign phase, DGSE provides evidence to identify and implement adjustments if necessary. In the case of operationally focused programmes, DGSE provides products designed to identify and solve specific problems that affect the intervention's efficiency and efficacy.

In terms of the products and services designed around the evaluation cycle, DGSE has developed a Performance Improvement Strategy (see Figure 3). The cycle begins by identifying potential areas of improvement in policies and social programmes. Either the programmes themselves or DGSE can start the process. In the second phase of the cycle, DGSE and the programmes decide together on the best instruments to produce the expected evidence in a timely manner.

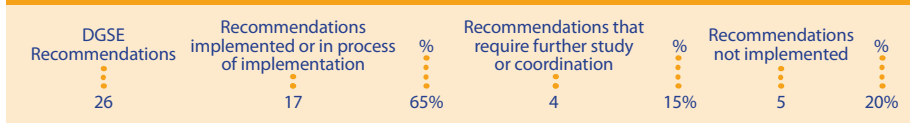
FIGURE 3: DGSE PERFORMANCE IMPROVEMENT CYCLE



Once the evidence production phase concludes, DGSE has conceived of an intermediate step between the production and use of evidence, where the Recommendations Technical Reports are prepared and presented. These management reports have been designed with the aim of providing policymakers in charge of the design and/or operation of public interventions with clear and timely recommendations that consider both the political and economic viability of implementation. Through this innovative design, MIDIS expects to connect scientific evidence with the pragmatic reality of the operation of social interventions, which should increase the use of evidence in decision-making processes.

Another innovative element in the Peruvian M&E system relies on the development of the Performance Improvement Plans, which demand close coordination and negotiation between DGSE and the programmes or units in charge of the operation of the evaluated interventions. The Performance Improvement Plans have been designed as a management tool that is based on the opportunities for improvement as identified in the Recommendations Technical Reports. The plans also consolidate the commitments assumed by the operators

FIGURE 4: USE OF DGSE'S RECOMMENDATIONS – 2012



of the evaluated interventions as well as the mechanisms for monitoring progress on implementing the reforms.

The next phase in the cycle is related to implementing the commitments assumed in the Performance Improvement Plans. Even though the implementing actions depend mostly on the operation of the evaluated programmes and interventions, DGSE is expected to provide technical assistance during the process in order to ensure correct and timely implementation.

Finally, the DGSE model considers a final stage, where the effects of the evidence-based improvements that were implemented in accordance to the Performance Improvement Plans are measured and evaluated on the impact, efficiency, quality, equity and transparency of the development and social inclusion interventions. This final step is also intended as a means to evaluate the success of DGSE as an evaluation unit. This component constitutes an innovation as well, because in the traditional evaluation unit model, success is commonly measured in terms of the number of evaluations performed or the scientific rigour and quality of the evaluations. That paradigm overlooks that real success goes beyond the ability of a unit to generate high-quality evidence in a timely manner, and needs to be measured in terms of the unit's contributions towards greater impact, efficiency, quality, equity and transparency of the policies and programmes evaluated.

Even though MIDIS's existence has been relatively short, the DGSE model design has already shown encouraging results. In the context of the redesign process of social programmes, in 2012, DGSE developed 16 evaluations that resulted in seven technical reports containing evidence-based recommendations for social programmes. Furthermore, 65 percent of the recommendations made by DGSE have been implemented (or are in the process of being implemented); 15 percent requires further studies or coordination with other sectors to be implemented.

A noteworthy example is the use of DGSE evidence in the decision to close a food assistance programme formerly named PRONAA (*Programa Nacional de Asistencia Alimentaria*). In 2012, DGSE commissioned seven evaluations regarding PRONAA, which showed that the programme was poorly targeted and was not available to a significant part of the population that required the most nutritional support. Additionally, the evaluations pointed out that PRONAA delivered a service with very low standards and had scarce transparency mechanisms, which resulted in corruption. The results and lessons learned from the evaluations conducted by DGSE over PRONAA were summarized in a Recommendations Technical Report and presented to the relevant decision makers. PRONAA was eliminated in May 2012 and ceased operations on 31 December 2012.

Other clear examples of how the DGSE model had an impact on policy are the evidence-based improvements introduced to the national *Cuna Más* programme in 2013 and the evidence provided for the Policy Guidelines against Child Chronic Malnutrition developed by MIDIS in 2012. In the first case, DGSE made evidence-based recommendations regarding the evaluations conducted in 2012 over the previous *Wawa Wasi* programme, which were used as an input to introduce several improvements to the quality of the day-care service provided. In the second case, DGSE provided national and international evidence that served as a basis to identify effective interventions against child malnutrition and to develop a policy tool to guide national and subnational government agencies in the design and implementation of their social policies.

REFERENCES

- Horton y col. 2008. "Evaluación del desarrollo de capacidades. Experiencias de organizaciones de investigación y desarrollo alrededor del mundo". Centro Internacional de Agricultura Tropical (CIAT) – Colombia.
- Innvær, Simon, Gunn Vist, Mari Trommald and Andrew Oxman. 2002. "Health Policy-Makers' Perceptions of Their Use of Evidence: A Systematic Review". Health Services Research Unit, National Institute of Public Health. Oslo, Norway.
- Weiss, C.H. 1999. "The Interface Between Evaluation and Public Policy". *Evaluation*. 5(4):468-486.

18. TRINIDAD AND TOBAGO

BUILDING A NATIONAL PERFORMANCE FRAMEWORK FOR EVALUATION

DINESH D. RAGOO

Programme Director, Ministry of Planning
and Sustainable Development

INTRODUCTION

Trinidad and Tobago is a small but rapidly developing country. It is generally considered as the financial capital of the Caribbean region. A cursory glance at the economic stability of the twin-island state shows that from 2011 to 2012, gross domestic product stood at \$13,632 million. Foreign Direct Investment now totals approximately \$1,110 million, and unemployment is below 5 percent.⁷⁶ However, over the last decade, there has been increasing concern among citizens, stakeholders and international financial and donors agencies about the efficiency and effectiveness of public-sector performance. For citizens, the issue is one of accountability and transparency as it relates to government expenditure of public funds and, of course, value for money. Though Trinidad and Tobago does not receive foreign aid, for international financial institutions such as the Inter-American Development Bank, the concern is focused on whether loan agreements (and in some cases, grant funding) are being used to achieve the objectives and outcomes for which they were intended.

The government, and by extension the public sector, is being challenged to become more effective in providing services such that there are positive impacts on the lives of citizens. In order to ensure this, the government's performance must be measured. In this regard, there has been the recognition that there is a need for developing a national M&E system with a clear understanding of the benefits it brings. There is also a general consensus as to the

76 Central Bank of Trinidad and Tobago, 2012.

crucial role of an effective and efficient public sector in facilitating economic growth, sustainable development and the well-being of all citizens.

However, in order to ensure effective service delivery, performance must be managed. Therefore, the M&E approach not only focuses on measuring performance, but also seeks to build a culture of performance management in public service that will be evidenced by good governance, transparency, efficiency in government operations and accountability for public spending. These outcomes, if managed properly, are expected to contribute to improved public-sector service delivery and greater stakeholder satisfaction.

Since 2005, the Government of Trinidad and Tobago has been involved in creating the infrastructure for a results-based performance management system in the public sector. The move towards M&E began in 1992, with capacity building in project cycle management and the use of logical frameworks, which became a requirement for submitting and approving capital projects for funding.

In 2008, as support and buy-in for M&E increased, the cabinet agreed to establish M&E units in all government ministries and relevant departments in order to provide support for evidence-based decision- and policymaking. In 2010, the cabinet further agreed to a re-designating the then Vision 2020 Office (the office with oversight for reporting on the implementation of the then National Strategic Plan of Trinidad and Tobago, *Vision 2020*). The Office was re-designated as the National Transformation Unit, now responsible for national M&E policy and oversight of M&E practice in the public service, and was tasked to provide general guidance to M&E units in line ministries. It is important to note that this unit has always been placed within the Ministry of Planning and Sustainable Development.

In 2010, the government introduced a new policy direction: Prosperity for All, Manifesto 2010. This document was translated into the Medium Term Policy Framework 2011–2014 and is meant to bring focus to government work from 2011 to 2014. The framework identifies the following five priority areas: crime and law and order; agriculture and food security; health care services and hospitals; economic growth, job creation, competitiveness and innovation; and poverty reduction and human capital development.

MAIN CONTENTS

Based on this policy document, a measurement strategy was developed in 2012. The National Performance Framework 2012–2015 outlined the measurement strategy for assessing and evaluating the implementation of the Medium Term Policy Framework. The performance framework document is the first of its kind in Trinidad and Tobago and represents a significantly new approach to planning and a distinct shift in thinking and policy, in that it incorporates the element of measurement.

It is common to develop extensive policies and plans, but translating these into actions that achieve results is a new experience for the country. The National Performance Framework shifts the focus from reporting on activities and outputs to the actual impacts of government interventions on the lives of our citizens. The National Performance Framework has outlined a results chain for each priority area, inclusive of outcomes and impacts, and

will serve as a measurement tool to track the achievement of the goals and targets that have been outlined in the Medium Term Policy Framework 2011-2014. It is also important to note that the National Performance Framework was developed using an extensive collaborative approach with all ministries and relevant departments and agencies.

Essentially, the National Performance Framework has detailed eight key result areas, 23 national outcomes and 52 national indicators in order to assess the Medium Term Policy Framework. Based on this new performance measurement framework, the first report on performance was produced in 2012. It is expected that the report will be produced annually by the National Transformation Unit and will report against the national outcomes by tracking national indicators. Both the National Performance Framework and the Annual Report on Performance 2012 were officially launched in February 2013 and subsequently laid in the Parliament of the Republic of Trinidad and Tobago.

To further gain support for M&E in the public sector and to strengthen the network of local M&E practitioners, the Development Evaluation Forum was initiated in 2011. This is a two-part initiative. The first part is an annual conference to exchange information in the area of local M&E, to motivate and to build a community of practitioners. The first meeting was held in November 2011; consultant Robert Lahey delivered the keynote address on using M&E in Trinidad and Tobago. The second annual meeting was held in November 2012, and featured a keynote address by Professor Christo de Coning, Board Member of the South African M&E Association, where he shared the 28 Lessons Learned from the South Africa Experience. Attendees were also updated on the year's local M&E-related activities.

The second part of the initiative is an online forum that was established to improve communication among the local, regional and international community of development evaluation practitioners. The forum seeks to expand the dialogue on the use, practice and benefits of development evaluation, particularly as it relates to the Caribbean and local contexts.

Trinidad and Tobago is also in the process of developing its first national policy on monitoring and evaluation. This policy will: assist in setting the infrastructure for an evidence-based decision-making culture in Trinidad and Tobago; facilitate the improvement of an integrated, all-inclusive participatory M&E system that has the capacity to carry out development studies; and aid Trinidad and Tobago's development stakeholders in acquiring timely and strategically focused national performance-based information for evidence-based decision-making.

BUILDING CAPACITY

In 2009, Trinidad and Tobago developed a medium-term action plan for building M&E capability. The plan's agenda for effective M&E development included: establishing an M&E centre in the Ministry of Planning and Sustainable Development to support the development and operation of results measurement and reporting; building M&E capability in all ministries through training and development; and producing and investing in a data-development strategy.

Despite these efforts, M&E capacities remain weak. At present, about 75 percent of all ministries (30) have established M&E units, based on the cabinets' decision in 2008. However, staffing of these units continues to be a challenge as a result of an insufficient number of M&E practitioners. Building capacity, particularly in evaluation, continues to be slow. Without the expertise available through local or regional institutions, ministries often have to look towards international training institutions to provide training to public officers. Nevertheless, the government remains committed to finding a solution to this dilemma. Dialogue continues with the regional bodies, such as the Caribbean Development Bank, as well as international organizations, such as the World Bank and the Inter-American Development Bank, to develop a comprehensive training programme—not only for Trinidad and Tobago, but for the Caribbean region at large.

CONCLUSIONS

The usefulness of any M&E system depends upon the credibility, utility and timeliness of data. Data management in Trinidad and Tobago is officially coordinated by the Central Statistical Office, which is currently undergoing a strategic restructuring in order to ensure its relevance and effectiveness in delivering its mandate. It is expected that this office will provide the requisite support for the emerging M&E system in Trinidad and Tobago. Insufficient data is a significant problem, particularly as it relates to building a robust and sustainable M&E system.

Another constraint, though one that does not present a significant challenge at this time, is the independence of the National Transformation Unit, which essentially serves as the centre for monitoring and evaluating the government's performance. As the national system is advanced, consideration may have to be given to making this unit a statutory body or an independent authority that reports to either the president or the parliament.

Trinidad and Tobago is at the point of moving towards an advanced national M&E system that provides timely reports. This system will influence decision-making, which will in turn determine how national budgetary allocations are made. To this end, an integrated public management framework is being worked out that incorporates national policy, strategic planning, reporting, budgeting and M&E. Moreover, it is anticipated that over the next three years, the system will be fully integrated within an information technology platform, with all the necessary user requirements and access defined.

As of 2013, Trinidad and Tobago is well on its way to becoming a leader in developing a results-based M&E system at the national level. Moreover, Trinidad and Tobago recognizes the value in M&E and continues to work towards building an environment that values measurement, performance and service delivery.

19. BENIN

THE PROCESS OF INSTITUTIONALIZING EVALUATION IN BENIN: PROGRESS ON QUESTIONS OF USAGE AND INDEPENDENCE

ARISTIDE N. DJIDJOHO

Coordinator of Public Policy, Evaluation Office
Cabinet of the Prime Minister

INTRODUCTION

Benin first became aware of the importance of evaluation in 2007, when it incorporated this function within a ministry. The inadequacies noted in the practice of evaluation in Benin, particularly the low level of institutionalization, led authorities to address the problem by creating the Public Policy Evaluation Office (BEPP).⁷⁷ The office currently operates under the authority of the Prime Minister.⁷⁸

Following the 2010 audit of national evaluation capacity, the choice was made to institutionalize evaluation at three levels (institutional, organizational and operational), in particular by clarifying institutional measures, creating tools, carrying out evaluations and national-level capacity-building. This specific experience, driven by the strong political will of the government, made it possible to improve the evaluation process and overcome many obstacles.

This paper summarizes the developments in the institutionalization process and analyses the use of evaluation in relation to questions of independence.

77 Bureau d'Évaluation des Politiques Publiques.

78 BEPP operates under the supervision of the Prime Minister's Cabinet in charge of the Coordination de l'Action Gouvernementale, de l'Évaluation des Politiques Publiques, du Programme de Dénationalisation et du Dialogue Social (coordination of government action, the evaluation of public policy, the denationalization programme and social dialogue).

DEVELOPMENTS IN THE INSTITUTIONALIZATION PROCESS

The government has mandated that BEPP evaluate and develop the evaluation of public policy in Benin.⁷⁹ Within this framework, the approach to institutionalization has been to clearly define the vision of the government in the area of evaluation: devise an M&E system to assist in decision-making and transparency, shared by the public administration, the institutions and all the other development players with the aim of improving the effectiveness of public policies.

In order to clarify the organizational structure of evaluation within the public administration, a formal system has been put in place. The Institutional Framework for the Evaluation of Public Policies,⁸⁰ created by the decree governing the adoption of the National Evaluation Policy, identifies all those involved in evaluation and specifies their roles. The BEPP is at the centre of the system used to carry out evaluations, disseminate information, capitalize on knowledge and strengthen capacities. This formal system is integrated into the Planning Programming Budget Monitoring and Evaluation process chain, and contributes to the evaluation of state projects, programmes and public policies. However, while this system works well at a national level, it is less effective at the sectoral level, particularly with regards to collecting, processing, analysing, centralizing and publishing data.

Following the implementation of the Institutional Framework for the Evaluation of Public Policies, Benin strengthened its approach to evaluation by drawing up and adopting the ten-year National Evaluation Policy⁸¹ in October 2012. The scope of its application includes the Strategic Development Orientations, the Growth and Poverty Reduction Strategy 2011–2015,⁸² other sectoral policies, the activities of public services and the actions of decentralized municipalities.

In addition, substantial resources have been invested to train and strengthen evaluation capacities of more than 150 managers working in ministries and municipalities. A biennial conference, known as the *Journées Béninoises de l'Évaluation*, was organized in 2010 and 2012 to mobilize the national and international evaluation community around the challenges of institutionalization. Despite making significant progress, substantial efforts are still required at both organizational and operational levels (see Figure 1).

USE OF EVALUATION

Evaluations carried out in Benin are mainly evaluations of public policies or evaluations of programmes and projects; they cover all social and productive sectors. The BEPP is responsible for carrying out evaluations of public policies and programmes that have national impacts; M&E focal points of BEPP within the programming and prospecting departments of each sectoral

79 Decree establishing the Office of the Prime Minister and decree governing the adoption of the National Evaluation Policy (Politique Nationale d'Évaluation).

80 Cadre Institutionnel de l'Évaluation des Politiques Publiques.

81 Politique Nationale d'Évaluation.

82 Stratégie de Croissance pour la Réduction de la Pauvreté.

FIGURE 1. DEVELOPMENT IN EVALUATION CAPACITY IN 2013

LEVEL	COMPONENT 1	COMPONENT 2	COMPONENT 3
MACRO (Strategic/ Institutional/ National)	VISION	POLICY	REGULATORY FRAMEWORK
	YES	YES	YES
MESO (Tactical/ Organizational/ Structural)	ROLE OF EVALUATION	SPECIFIC BUDGET	SPECIFIC SKILLS
	STRONG	AVERAGE	WEAK
MICRO (Operational/ Technical/Tools)	PROFESSIONAL RESOURCES	SPECIFIC TRAINING	QUALITY EVALUATIONS
	WEAK	WEAK	WEAK

ministry are responsible for their ministry’s projects and programmes; the Social Change Observatory⁸³ is in charge of evaluating the impacts of poverty reduction programmes.

Evaluations seek to assess the overall effectiveness of policies, but focus on specific aspects on a case-by-case basis. Evaluations are carried out by consultants and independent firms to guarantee the impartiality of the reports. In this matter, the BEPP has drawn up a code of professional ethics inspired by international standards.

Evaluations are used by the government, the administration, civil society and beneficiaries; they are used by decision makers (Council of Ministers), policy managers (ministries, technical departments) and technical and financial partners to provide information and help with decision-making. Evaluations help in the drawing up of new policy cycles and public programmes, and can help determine resource reallocations.

Because evaluation reports are not published, evaluations are not particularly useful for informing the public. However, the reports are made available to the public administration, technical and financial partners and to the professional associations of the sectors evaluated. BEPP is nevertheless keen to develop a participative approach to encourage better dissemination of information to civil society.

Analysis shows that the results of the M&E system are not sufficiently taken into account, although there have been positive developments in recent years. The quality of the evaluation reports and the M&E of the Growth and Poverty Reduction Strategy have improved, which has had a positive impact on the adoption of recommendations.

83 L’Observatoire du Changement Social.

PRINCIPAL CHALLENGES IN BENIN

In order to meet the needs of development, notably in improving the effectiveness of public action, there are a number of challenges to overcome that relate to the use of evaluation and its independence.

Planning

Evaluation provides information on the public management cycle. This understanding is fundamental with regards to problems related to using evaluation. This is because if evaluation activities remain disconnected from the public management cycle, it will always be difficult to ensure that evaluations are used correctly. It is therefore necessary for the practice of evaluation to become more systematic and for it to correspond to the government's planning and budget programming timetables.

Usefulness

Usefulness lies in the capacity of evaluation stakeholders to ask the right questions and use the results of an evaluation correctly in order to improve the quality of public actions. Absent a clear demonstration of how evaluation can be useful to Benin, evaluation will quickly be seen as a niche for a handful of specialists to whom resources are awarded unnecessarily.

Communication and dialogue

Evaluation must become a tool for dialogue and a means to achieve scientific, social and even political consensus around the policies evaluated. Evaluations must stimulate public debate and not be used by only managers, as is still the case in Benin.

Quality

It is absolutely essential that the development of evaluation practices does not have a detrimental effect on evaluation quality. Evaluations must remain rigorous and objective and must serve development purposes. To this end, evaluations must be carried out by skilled individuals who are independent, have no declared or apparent conflicts of interest and respect the standards of professional ethics.

Role of institutions and civil society

If evaluation is encouraged by the executive branch but not by other stakeholders, it is neither encouraging nor viable. In Benin, the parliament shows little interest in evaluation despite the efforts of the government; there is a lack of dynamism from civil society. This is a trend that needs to be reversed.

POSSIBLE SOLUTIONS

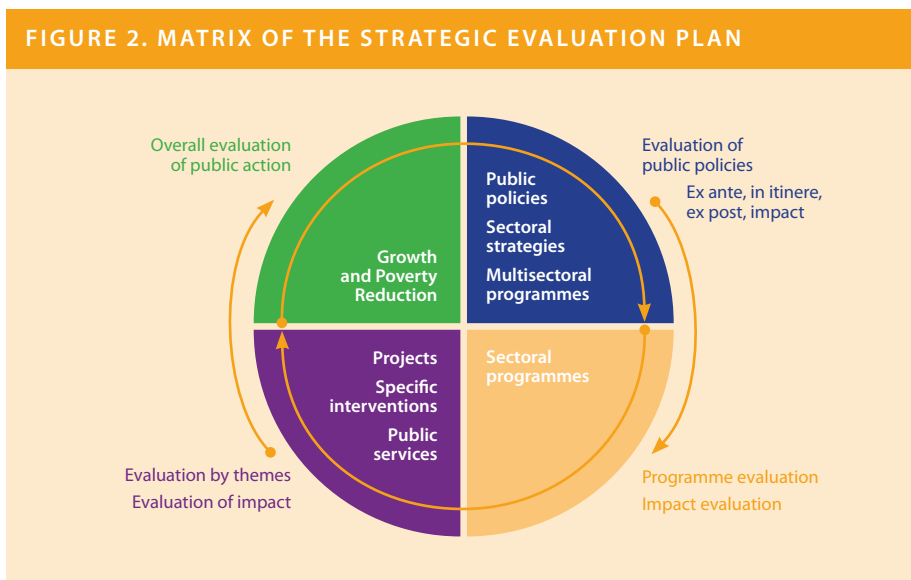
To address these challenges, Benin is working to continue the evaluation development process while putting in place practical solutions.

Strategic Evaluation Plan

The Strategic Evaluation Plan is a national evaluation programme based on the Growth and Poverty Reduction Strategy. The Strategic Evaluation Plan brings together state interventions and technical and financial partners to systematically plan all evaluation activities considered to be a priority at the central level. By putting in place a tool to ensure that evaluation of the Growth and Poverty Reduction Strategy becomes operational, it will be possible not only to mobilize the structures around the priority policies, strategies and programmes to evaluate, but also to find suitable responses to the concerns of the government and the technical and financial partners.

Creation of the National Evaluation Council

The National Evaluation Council⁸⁴ was established in 2012 by a decree passed during a meeting of the cabinet. The council is in charge of: advising the government about evaluation and promoting the development of evaluation at the national, regional and municipal levels; supporting BEPP in drawing up the government's various evaluation programmes; and



84 Conseil Nationale de l'Evaluation.

promoting norms, standards and methodologies. Its creation is a way to resolve questions of independence in evaluation. More perennial approaches, such as constitutionalizing evaluation, are foreseeable in the long term.

Capacity building

Strengthening evaluation practices entails increasing the demand for evaluation. To be intense and sustainable, the demand for evaluation must be fulfilled by evaluations that are independent, impartial and of adequate quality and quantity. Therefore, it is necessary to strengthen the capacities of both individuals and organizations and professionalize evaluation. Professional training is underway to this end, and a university programme was planned from late 2013.

Strengthening international initiatives

Support of international organizations—e.g. Regional Centers for Learning on Evaluation and Results (CLEAR), United Nations Development Programme (UNDP)—remains a formidable lever for developing evaluation practices. In the case of Benin, UNDP's support for BEPP and Benin's participation in forums such as the national evaluation capacities conferences have been very useful for promoting evaluation. It has made it possible to form new cooperative relationships with the Performance Monitoring and Evaluation Department of the Presidency of South Africa, the Evaluation Commission of the Office of the Prime Minister of Uganda, the African Evaluation Association, the CLEAR Initiative and the International Initiative for Impact Evaluation. It would be advisable to intensify this type of cooperation with countries of the global South.

CONCLUSION

Despite some operational difficulties, Benin's experience shows that institutionalizing evaluation is an important factor in improving the quality of governance and, therefore, for the development of African countries. The M&E system in Benin is still a work in progress, but the impetus generated by adopting evaluation at the national level means that it should be possible to gradually develop an evaluation culture.

With regards to the challenges to overcome, Benin's recommendations include:

- Strengthening institutional, regulatory, legal and constitutional measures to guarantee the independence of evaluation, its sustainability, and to require M&E public policies (e.g. an evaluation and results-based management act);
- Fostering the participation of parliamentary and civil society bodies in evaluation processes; and
- Promoting the synergies possible among donors in order to create centres of excellence and 'tool box' windows for countries to develop evaluation.

REFERENCES

Bureau d'Évaluation des Politiques Publiques (BEPP). 2010. *Etude diagnostique des capacités nationales en évaluation au Bénin*.

BEPP and Center for Learning on Evaluation and Results (CLEAR). 2012. *Etat des lieux du système national de suivi et d'évaluation du Bénin*.

PRIMATURE and BEPP. 2012. *Politique Nationale d'Evaluation 2012-2021*.

20. GHANA

BUILDING M&E CAPACITIES TO ENHANCE THE NATIONAL M&E SYSTEM IN GHANA: THE WAY FORWARD

BRUNO B. DERY
Deputy Director

National Development Planning Commission

BACKGROUND

The Government of Ghana is committed to effective public service delivery, strengthening government accountability to its citizens, ensuring that policy formulation and decision-making are based on evidence, and that results are achieved in relation to the country's growth and development targets. The 1992 Constitution and several acts of parliament recognize the need to use monitoring and evaluation to ascertain the extent of progress made towards achieving the objectives of national policies and interventions. There is also a general recognition that M&E are the main instruments for assessing the extent to which the government has done what it pledged to do within the context of its development policies and plans.

Ghana has elaborate institutional arrangements for M&E at the national, regional and district levels. The constitution and supporting legislation have defined M&E roles and responsibilities for government agencies, specific departments of sectoral ministries and units in regional and district administrations. The National Development Planning Commission (the Commission) is the apex body responsible for national M&E. The Commission's M&E guidelines emphasize the involvement of traditional authorities, civil society organizations and development partners to create a participatory approach for effective M&E and feedback mechanisms. The Commission recognizes the fact that both monitoring and evaluation are indispensable tools for measuring performance and development outcomes.

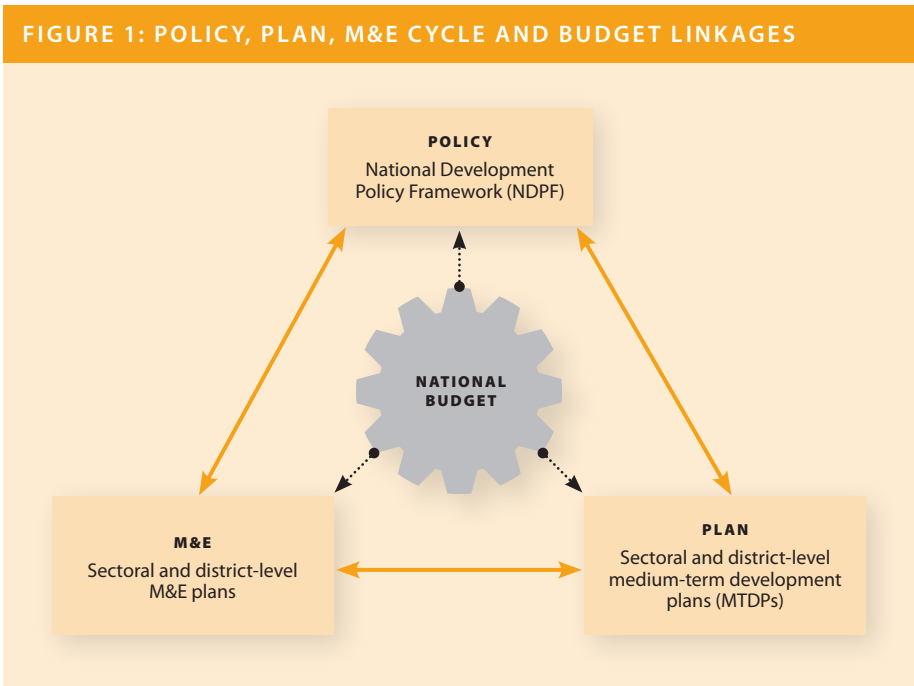
THE NATIONAL M&E SYSTEM

Over the past 10 years, the Commission has increased its efforts to establish a functional national M&E system and to eventually ingrain the culture of M&E in the public sector. The system is hinged on successive four-year medium-term national development policy frameworks and corresponding development plans at the sector and district levels.

National policy frameworks

National development policy frameworks are the principal development policy blueprints that are crafted by the Commission at the national level, with inputs from the public and private sectors. The Commission collates and analyses policy inputs from ministries, departments, agencies and other sources in order to produce a national development policy framework.

Realizing policy objectives depends largely on its effective linkage to development plans at the sector and district levels, and to the national budget. Consequently, effective M&E systems provide valuable feedback and lessons for continuous improvement of the policies, plans and the national budget. The indispensable linkages among these national development building blocks are depicted in Figure 1.



The last three national development policy frameworks were:

- Ghana Poverty Reduction Strategy (GPRS I) 2002–2005;
- Growth and Poverty Reduction Strategy (GPRS II) 2006–2009; and
- Ghana Shared Growth and Development Agenda (GSGDA) 2010–2013.

Sector and district development plans

Plan preparation follows policy formulation within the planning cycle. It is the medium for translating the policy objectives and strategies of the relevant national development policy frameworks into implementable projects and programmes. The process integrates the spatial, social, economic and environmental issues into specific actions within the decentralized planning system. In this respect, guidelines are issued by the Commission that are in line with the national development policy frameworks and prescribe the process, format and content of the development plans to districts and sectors. Medium-term development plans are then produced by all the metropolitan, municipal and district assemblies and ministries, departments and agencies. These plans are prepared for implementation in order to achieve the policy objectives of the national development policy frameworks.

National M&E plans

M&E processes begin after plan preparation. Effective M&E systems are built on policy formulation, development planning and budgeting systems to provide valuable feedback to those systems. M&E is the main instrument for assessing the extent to which the government has done what it pledged to do in its national policy statements and development plans. The Commission has produced three national M&E plans in line with the GPRS I & II and the GSGDA.

Sector and district M&E guidelines

The Commission issues M&E guidelines for all metropolitan, municipal and district assemblies and sector ministries, departments and agencies for preparing an M&E plan for their corresponding development plans. The sectoral and district-level M&E plans form the basis for their M&E systems and serve as roadmaps for the monitoring and evaluation of corresponding development plans. The guidelines also contain formats and timelines for preparing M&E plans and quarterly and annual progress reports.

M&E training and backstopping

The Commission organizes basic M&E training for core M&E staff of metropolitan, municipal and district assemblies and ministries, departments and agencies in order to enhance their understanding and capacity to prepare and implement their M&E plans. The two-day workshops first discuss key concepts (i.e. monitoring, evaluation and participatory M&E). Attendees then participate in group exercises to explore the 10 steps of preparing an M&E Plan:

1. Identifying and analysing stakeholders;
2. Assessing M&E needs, conditions and capacities;
3. Monitoring indicators;
4. Constructing an M&E matrix;
5. Creating an M&E work plan and calendar;
6. Preparing an M&E budget;
7. Collecting, validating and collating data;
8. Analysing data and using the results;
9. Reporting; and
10. Disseminating information.

On request, Commission staff also organizes backstopping meetings and tailored training programmes to further enhance staff abilities to prepare M&E plans.

Annual progress reports

The majority of metropolitan, municipal and district assemblies and ministries, departments and agencies now prepare M&E plans and annual progress reports. Draft copies of these documents are submitted to the Commission for review and approval. The Commission prepares checklists that are used to review all draft M&E plans and annual progress reports submitted in order to ensure compliance with the key requirements of the M&E guidelines and report formats. Meetings are organized to provide feedback to all metropolitan, municipal and district assemblies and ministries, departments and agencies, in the form of general and specific comments on their M&E plans and annual progress reports. By so doing, the Commission has streamlined the national M&E processes, provided a standard yardstick for measuring progress and established the minimum requirements for any M&E system in the country.

The Commission has produced 10 national annual progress reports between 2002 and 2012. These reports have informed the government on its performance, where it is doing well and where there are weaknesses. The reports also provided policy recommendations on minimizing the weaknesses.

National M&E manual

A key recommendation from participants of the M&E training workshops organized by the Commission was the need for a reference manual. The Commission took up the challenge and produced a draft national M&E manual with financial support from the World Bank and the Japanese International Cooperation Agency. Review work on the manual was planned for launching in the country. This document will supplement the M&E guidelines, M&E plans and other M&E documents produced by the Commission in collaboration with

its development partners. The manual covers all the essential elements of a national M&E system within the context of development in the Ghana public sector. The main themes of the manual are discussed in four parts: national M&E rationale and linkages, understanding key concepts, understanding M&E as a system and steps to developing an M&E system. The four parts are subdivided into seventeen chapters.

Other M&E activities

During this process of building the national M&E system, the Commission also:

- Conducted a poverty and social impact analysis in order to assess the consequences of certain government policy reforms on the poor and vulnerable under GPRS I;
- Undertook participatory M&E exercises on select national issues and produced citizens assessment reports in 2005 and 2008;
- Disseminated the national annual progress reports and organized regional dissemination workshops to discuss their content;
- Conducted advanced, tailored M&E training programmes for newly created districts in 2009;
- Conducted a baseline study on budgetary resources spent on M&E and statistics by ministries, departments and agencies in 2011; and
- Prepared biennial Ghana Millennium Development Goals reports (2002, 2004, 2006, 2008 and 2010) with the support of the United Nations Development Programme in order to gauge Ghana's progress towards attaining the Millennium Development Goals.

CHALLENGES

The Commission made steady progress, surpassing numerous challenges in its efforts to establish a functional national M&E system. These challenges included:

Weak demand for and utilization of M&E results

The demand for and utilization of M&E results in policy formulation and decision-making is still very low at all levels in the country. For example, national annual progress reports were produced largely in response to development partners' demands for information in the context of their direct support to the national budget through the Multi-Donor Budget Support process. A higher demand for M&E results by the government would have ensured that all interventions were monitored and the impacts of these programmes on citizens were effectively evaluated. Further, stronger demand for M&E results by parliament in its oversight role over the executive is needed. Civil society organizations also need to increase their demand for M&E results for advocacy and social accountability.

Limited resources and budgetary allocations for M&E

To date, M&E has received little priority in all ministries, departments and agencies, and in metropolitan, municipal and district assembly budgets; and actual disbursements were comparatively insignificant. The funds released for M&E activities are often insufficient to build and maintain effective systems. Some ministries, departments, agencies and metropolitan, municipal and district assemblies lack the physical equipment required to undertake M&E and data-production activities (e.g. computer systems, transport and office tools). M&E expenditures often lose out when budgetary releases are inadequate. Evaluations are often not conducted primarily due to a lack of funds.

Weak capacity

Although the Commission provided some training to ministries, departments and agencies and metropolitan, municipal and district assemblies, intuitional and individual M&E capacities remain weak. There are limited resources to build the necessary M&E skills within ministries, departments, agencies, and metropolitan, municipal and district assemblies or to ensure that M&E information is used to inform budget allocations and policy formulation.

Lack of incentives, rewards and sanctions

Ministries, departments, agencies, and metropolitan, municipal and district assemblies are not held to account for the results they achieve and not required to demonstrate value for money or explain whether they have achieved what they said they would with the resources allocated. There are also no incentives at the individual level to reward staff for carrying out M&E activities.

Non-compliance with M&E reporting timelines and frameworks

The majority of metropolitan, municipal and district assemblies now produce their M&E plans and annual progress reports using the agreed formats in the guidelines. However, the adherence to reporting timelines remains a huge problem at all levels. The differing reporting cycles of ministries, departments and agencies impedes the coordination of M&E results across government and negatively affects the quality and timeliness of producing the national annual progress report.

M&E frameworks that are unaligned to the guidelines issued by the Commission are often due to development partners supporting different M&E systems instead of using the national M&E systems.

Poor data quality, data gaps and inconsistencies

Every year, the production of national annual progress reports must overcome inconsistencies in the data provided by ministries, departments and agencies for the same variables over the same time period, incomplete data in district-level annual progress reports and a lack of data for some indicators.

Inadequate management information systems and networks

Management information systems across ministries, departments, agencies, and metropolitan, municipal and district assemblies are inadequate and, in some cases, non-existent. In

many ministries, departments and agencies, data from the district and regional levels is not transferred over a network.

THE WAY FORWARD

The Commission recognizes that the road to a functional M&E system is very long. The Commission is also aware that it will have to build solid partnerships and continue to refine the roadmap through a very participatory process. To move forward, one of the immediate steps the Commission must take is to enhance the M&E awareness and capacities of policy- and decision-makers at all governance levels. With the support of the Japanese International Cooperation Agency, the Commission will soon organize tailored M&E training programmes for ministers of state, members of parliament, chief directors, district chief executives and presiding members of district assemblies. If the political leadership understood and appreciated the value of M&E, this would translate into greater support for establishing a national M&E system and enhancing investments in evaluations.

Other steps the Commission must take in the near future include developing a national long-term policy on building institutional and individual M&E capacities to increase the pool of skilled M&E specialists in order to support the growth of the national M&E system. The Commission should also develop mechanisms to increase general awareness and recognition of the necessity and benefits of having a strong national M&E system. This will enhance demand and use of M&E information by the office of the president, parliament, civil society organizations, research and academic institutions, development partners, the media and indeed all citizens of Ghana. There is an urgent need to shift from preparing annual progress reports merely to comply with statutory requirements or donor-related demands to generating domestically owned M&E products as an integral part of good governance.

There is the need to institute rewards and sanctions for institutions and individuals that abide by or fall short in the use of time and resources to meet their M&E mandates. An example would be to strengthen the linkages between M&E results and the release of funds to ministries, departments, agencies, and metropolitan, municipal and district assemblies.

The Commission should develop a long-term strategy for a sustainable funding mechanism for all M&E activities in the country—for example, by earmarking 2 to 5 percent of all development budgets for M&E (e.g. investment and service portions of the national budget, budgets of the medium-term development plans, budgets of projects and programmes).

There is a need to develop a long-term mechanism to progressively increase budgetary allocations to the Commission, ministries, departments, agencies, and metropolitan, municipal and district assemblies, as well as the Ghana statistical service, in order to conduct evaluations, to collect credible data and to develop information management systems. There is a need to have database systems readily available to store, analyse, retrieve and properly use M&E information for policymaking and budgetary decisions. The information and communications technology networks between different levels of government also need to be addressed in order to ensure timely and quality reporting of information and data.

The Commission should develop a dialogue mechanism that will ensure that all development partners are aware and support the national M&E system and avoid duplicating efforts.

REFERENCES

- Government of Ghana and Canadian International Development Agency. 2004. "Indicators for Change. Decentralized Poverty Monitoring and Evaluation: An Operational Manual for District Assembly and Regional Staff". District Capacity Building Project (DISCAP). Ministry of Local Government and Rural Development. Accra.
- Government of Ghana. 2009. "Guidelines for Preparation of District Monitoring and Evaluation Plans under GSGDA (2012-2013)". National Development Planning Commission document. Accra.
- Government of Ghana. 2009. "Medium Term M&E Development Plan under GSGDA 2010-2013". National Development Planning Commission document. Accra.
- Government of Ghana. 2009. "Sector Monitoring and Evaluation Guidelines under GSGDA 2010-2013". National Development Planning Commission document. Accra.
- Government of Ghana. 2011. "Resources Spent on M&E and Statistics, Monitoring and Evaluation". National Development Planning Commission document. Accra.
- Independent Evaluation Office, United Nations Development Programme. 2007. *Evaluation of Results-Based Management at UNDP: Achieving Results*.
- Independent Evaluation Office, United Nations Development Programme. 2009. *Handbook on Planning, Monitoring and Evaluation for Development Results*.
- Imas, Linda G. Morra, and Ray C. Rist. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations*. World Bank.
- Organisation for Economic Co-operation and Development (OECD). 2002. *Working Party on Aid Evaluation. Glossary of Key Terms in Evaluation and Results Based Management*. Development Assistance Committee. Paris.
- Osborne, David, and Ted Gaebler. 1992. "A power of measuring performance". *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. Plume.
- Patton, Michael. 1997. *Utilisation-Focused Evaluation*. 3rd ed. Thousand Oaks: Sage Publications.
- United Nations Children's Fund (UNICEF). 1991. *A UNICEF Guide to Monitoring and Evaluation: Making a Difference?*
- United Nations Evaluation Group (UNEG). 2013. "National Evaluation Capacity Development: Practical Tips on How to Strengthen National Evaluation Systems". Report for the United Nations Evaluation Group (UNEG) Task Force on National Evaluation Capacity Development.
- World Bank. 2007. *International Program for Development Evaluation Training Handbook*. International Program for Development Evaluation Training (PDET).
- World Bank. 2008. *Monitoring and Evaluation Handbook for Business Environment Reform: A Handbook for Practitioners*. Washington, DC.

21. KENYA

CREDIBILITY OF EVALUATION: THE KENYAN CASE

SAMSON MASESE MACHUKA

Director, Monitoring and Evaluation Directorate
Ministry of Devolution and Planning, the Presidency

JENNIFER MUTUA

M&E Specialist, Chair of Evaluation Society of Kenya
and East African Region Board Member of
the African Evaluation Association

INTRODUCTION

Kenya is implementing 'Kenya Vision 2030', a long-term plan that seeks to create a cohesive, equitable and just society based on democratic principles. This aspiration is built on the framework of economic, social and political pillars. One of the strategies that the Government of Kenya uses to achieve this is through the promotion of good governance and accountability using the national M&E system coordinated by the Monitoring and Evaluation Directorate domiciled in the Ministry of Devolution and Planning.

This strategy is further supported by the Kenya Constitution 2010 that was promulgated on 27 August 2010. The Constitution provides for the devolution of governance to the county level, transparency, fair representation in parliament, separation of powers and an independent judiciary, among other important pillars that are to be informed by a well-functioning evaluation system.

The national M&E system was established in 2004 through a multi-stakeholder effort to track the implementation of policies, programmes and projects at all levels of governance. Development partners that helped establish the system include UN bodies, the United Kingdom's Department for International Development and the World Bank. Significantly, the government provides an annual budgetary allocation to support national M&E system operations. The system is operated under the strategic direction of a National Steering Committee that is chaired by the Permanent Secretary in the Ministry of Devolution and Planning, which

includes stakeholders from the government, development partners and civil society civil society organizations.

At its formation, the system aimed at encouraging the culture and practice for M&E and at promoting accountability in order to enhance public service delivery. The national M&E system is also expected to provide timely and reliable feedback to budgetary preparation processes through the preparation of M&E reports such as the Ministerial Monitoring and Evaluation Report, Ministerial Public Expenditure Review and the Annual Progress Report. Results from the national M&E system were to also provide evidence on the effectiveness of government programmes and projects by ensuring the active participation of entities such as civil society organizations, non-governmental organizations, academia and the private sector.

The national M&E system has five basic components that include:

1. Capacity development and policy coordination to manage evaluation and to make revisions to the implementation of the framework;
2. Development of qualitative and quantitative integrated data-acquisition and data-management frameworks to support research and evaluation of results;
3. Research and results analysis that interrogates key report findings as a basis of informing policy and budgetary issues;
4. Enhanced project M&E systems at the central and devolved levels that supports public values and results-based service delivery, and is supported by an electronic project management information system; and
5. Integrated dissemination and communication mechanisms for sharing evaluation results and findings with a feedback mechanism. It also supports sharing reports on the government's national and global commitments to a wide cross-section of stakeholders, and endeavours to strengthen partnership engagement with the government, non-governmental actors and development partners for the effective use of M&E information.

MAIN CONTENT

National M&E system operations and implementation are guided by Technical Advisory Groups, teams of experts drawn from the government, non-governmental organizations and development partners. Notwithstanding this support, the national M&E system is challenged by weak capacities and M&E culture. In order to strengthen these capacities, the Monitoring and Evaluation Directorate renewed its efforts to involve non-state actors in refining its national M&E framework. Members were drawn from civil society organizations, UN agencies, the private sector and the Evaluation Society of Kenya.

The national M&E system operates at both the national and devolved levels. At the national level, the framework exists in all the 18 ministries that form the national government structure. At the devolved level, committees exist in all 47 counties. The county committees have the flexibility of devolving to lower-level structures that are closer to the communities.

The framework is expected to provide a transparent platform by which government and development partners can undertake shared appraisal of results.

With coordination from the Monitoring and Evaluation Directorate, second generation national M&E indicators have been consultatively developed, reviewed and revised. The capacities for using the indicators to track progress in order to provide evidence to decision makers still needs strengthening at the individual and institutional levels. The lack of capacities has led the Monitoring and Evaluation Directorate to conduct periodic M&E trainings for national and subnational staff. Similarly, the Monitoring and Evaluation Directorate closely collaborated with UN Women to develop gender-specific and gender-sensitive indicators for the second medium-term plan.

Since 2004, the Government of Kenya has used the national M&E system to regularly produce Annual Progress Reports and Public Expenditure Reviews. These reports have been useful in providing information for designing service provision policies and during the budget-making process, by informing the quantum of allocations for each sector. Products of the national M&E system are also a reservoir of knowledge and learning, useful for refining policies supporting the achievement of Kenya Vision 2030 objectives.

Through a multi-stakeholder effort, particularly with support from the World Bank and the German Development Agency, Kenya undertook a public expenditure-tracking survey in the ministries of agriculture, health and education. Since 2004, the Ministry of Health has conducted five public expenditure-tracking surveys in the health sector.

The public expenditure-tracking surveys conducted by the Ministry of Health have been done on “delays and leakages of funds” and on the “trends and flow of funds and commodities.” Information derived from the public expenditure-tracking surveys has been essential for the analysis of Public Expenditure Reviews and for identifying leakages and inefficiencies in the system. Public expenditure-tracking survey analysis has been useful in providing information to strengthen the systems of funds transfers, which has increased transparency and accountability in decision-making.

The public expenditure-tracking surveys have also provided evidence on the differences in performance across facilities and influenced the improvement of many procedures, including the removal of several redundant stages of central-level approval and authorization processes prior to distributing notifications of Authority to Incur Expenditures, the issuing of cheques that accompany the Authority to Incur Expenditures, the introduction of the 10/20 policy on cost-sharing, changes in the KEMSA distribution system (pull/push system) and the direct transfer of funds to rural health facilities.

Another innovation in which the Monitoring and Evaluation Directorate is involved in is the coordination of the Kenya Community of Practice on Managing for Development Results initiative. This continent-wide initiative on managing for development results is being spearheaded by the African Development Bank. One of its key innovations is the involvement of civil society organizations and the private sector in a public sector-led initiative, with a special focus on regional integration.

The Evaluation Society of Kenya is able to connect with the African Evaluation Association and the International Organisation for Co-operation in Evaluation/United Nations Children's

Fund-led EvalPartners Global initiative. This peer-to-peer learning is based on regional and global synergies that strengthen country-led M&E efforts. It is expected that collaborations between the Monitoring and Evaluation Directorate and the Evaluation Society of Kenya will add value to the implementation of the national M&E system through peer learning and experience sharing.

Similarly, the Monitoring and Evaluation Directorate has been involved in a South-South learning initiative; the Directorate hosted delegations from Sudan, Somalia and the Republic of South Africa, which came to study the institutional arrangements and operational structures of the national M&E system. The delegation from the Republic of South Africa was also interested in learning about how parliament uses M&E information.

The Monitoring and Evaluation Directorate, in close collaboration with other stakeholders, finalized an M&E policy designed to guide the national M&E system's implementation. The draft policy addresses the issue of implementation and coordination with the aim of improving management for development results. A well-coordinated M&E system is expected to improve performance in the execution of public programmes and projects in the country.

This policy will also focus on both state and non-state actors engaged in development activities in the country. It emphasizes building strong linkages with all stakeholders, including Kenyan citizens, to improve reporting and feedback mechanisms in all development aspects. It is expected that a working M&E system will ensure efficiency in utilizing resources to effectively realize results and, hence, to accelerate development. It also espouses the importance of results-based management, transparency, accountability and efficiency as fundamental principles for managing all public programmes and projects in Kenya.

In line with what is espoused in the policy, and as one of the innovations to straighten M&E capacities, the Monitoring and Evaluation Directorate is collaborating with the Evaluation Society of Kenya in professionalizing M&E practices in the country. This close collaboration was evidenced when the Monitoring and Evaluation Directorate and the Evaluation Society of Kenya co-hosted the inaugural National M&E Week in November 2012. Collaborations with the Evaluation Society of Kenya are expected to continue, especially in capacity-building efforts.

CONCLUSIONS

A recent national M&E system needs assessment and capacity analysis revealed that after more than eight years of operations, the country still lagged in uptake of M&E culture and practices. The report revealed the need to strengthen capacities for individuals and institutions, and recommended identifying high-level political champions to lend political weight for M&E uptake and increased funding to support devolved-level operations.

Plans are underway to address weaknesses in the supply and demand for evaluation by identifying and commissioning strategic programme evaluations to inform evidence-based decision-making. The Evaluation Society of Kenya, among other stakeholders, is expected to play a role in this by providing professional input into the processes.

REFERENCES

- Government of Kenya. 2007. "Master Plan for the Implementation of a National Integrated Monitoring and Evaluation System in Kenya". Ministry of State for Planning, National Development & Vision 2030. Nairobi, Kenya.
- Government of Kenya. 2012. "NIMES Needs Assessment and Situational Analysis Report". Ministry of State for Planning, National Development and Vision 2030. Nairobi, Kenya.
- Njunguna, D. 2013. "Public Expenditure Tracking Surveys, Experiences in Health". Ministry of Health. Nairobi, Kenya.

22. MALAWI

MONITORING AND EVALUATION OF THE NATIONAL BUDGET IN MALAWI

WINSTON NYASULU
Assistant Budget Director
Monitoring and Evaluation Section
Ministry of Finance

INTRODUCTION

Budgets are a vital policy tool that governments use to promote their objectives, from ensuring macroeconomic stability to allocating resources and delivering services. Given the inherently technical nature of budgetary processes, and their core political function of raising and distributing public resources, such processes have often been scrutinized and deemed to be characterized by a lack of transparency and scarce participation by external actors (Ministry of Finance 2009).

The only way to counter this thinking is through budgetary M&E, which provides documented observations of processes and activities of government programmes over a time period. It is a way of ensuring that monies voted for various projects and activities in the budget are actually spent and put into good use.

According to the Public Finance Management Act (2003), the Ministry of Finance is mandated to follow up on budget implementation on both the recurrent and the development budget. Currently, the M&E Section of the Budget Division takes up this responsibility. However, there are other institutions that also conduct budgetary M&E: the Ministry of Economic Planning and Development and the Office of the President and Cabinet.

JUSTIFICATION FOR MONITORING THE GOVERNMENT BUDGET

There are a number of reasons why M&E of the budget is important for the Ministry of Finance and other stakeholders. Such reasons include:

- Government budgets involve public finances, and in a democracy, citizens have the right to know what their money is being spent on and how;
- To determine the government's level of performance in terms of service delivery promised to the people;
- To keep ministries and government departments alert, since they know that they are being assessed by others as they perform their activities;
- To enhance transparency in government and to build public confidence in budget processes; and
- To assist in decision-making within the government.

INSTITUTIONS RESPONSIBLE FOR BUDGETARY M&E

Ministry of Finance

The Ministry of Finance is responsible for monitoring implementation progress of the budget on all its categories (e.g. Personal Emoluments, Other Recurrent Transactions, Development). Wage bill trends are analysed under the Personal Emoluments Budget; where huge variations are found, reviews are made for proper recommendations. The Other Recurrent Transactions Budget is monitored by comparing planned outputs (as indicated in the Output-based Budget Document) to the actual outputs delivered by the ministries. The Development Budget is monitored by visiting project sites (e.g. roads, hospitals), and then reports are written based on the progress made at those project sites. The reports are produced on a quarterly basis and submitted to the Secretary to the Treasury and the Cabinet.

Ministry of Economic Planning and Development

The Ministry of Economic Planning and Development is responsible for evaluating projects and programmes as implemented in the National Budget. The basis for evaluating projects is the Public Sector Investment Programme, which is a basket of projects implemented by the government under the mandate of the ministry (Ministry of Economic Planning and Development 2009). In addition, the ministry leads the process of developing national policy documents, such as the Malawi Growth and Development Strategy. The Growth and Development Strategy is the overarching medium-term strategy for Malawi designed to attain long-term development aspirations. The national budget is based on priorities set out in the Growth and Development Strategy. This is why the Ministry carries out evaluations to assess impact of projects and programmes on livelihoods.

Office of the President and Cabinet

The Office of the President and Cabinet has a Programmes and Project Evaluation Unit that monitors budget implementation by evaluating organizational performance agreements. Organizational performance agreements are signed agreements between controlling

officers of ministries and the Chief Secretary in the Office of the President and Cabinet. In organizational performance agreements, ministries commit to achieving specific outputs in a particular financial year, depending on the levels of resources provided. The assessment tries to ascertain if the agreed-to outputs were achieved.

CHALLENGES IN MONITORING AND EVALUATING THE BUDGET

M&E faces a number of challenges, such as:

Duplication of reports

Ministries submit monitoring reports to a number of institutions (e.g. Office of the President and Cabinet, the Ministry of Finance, Ministry of Economic Planning and Development), which creates duplication and heavy paperwork burdens.

Inadequate capacity to monitor the whole budget

The M&E Section in the Budget Division does not have the required capacity to monitor the entire budget. This is because the section does not have enough officers to simultaneously prepare and monitor the budget.

Resistance of ministries subject to monitoring and evaluation

When M&E is perceived of as auditing, ministries and departments are reluctant to participate in M&E activities. Therefore, there is reluctance to submit budget progress reports, which make it difficult for the M&E Section to follow up on outputs being delivered by ministries.

Lack of an M&E database system

There is lack of an M&E system that coordinates information for all the M&E departments in the government. Reports from ministries are usually hard copies, which can be misplaced easily.

Absence of an M&E unit in some ministries/departments

Not every ministry/department has an M&E unit, which hampers information collection from these departments.

Lack of information sharing

Sharing of information among stakeholders (e.g. Office of the President and Cabinet, the Ministry of Finance, Ministry of Economic Planning and Development) involved in M&E of the budget is a challenge.

Unclear output indicators

Since the M&E Section monitors immediate outputs, the Output-based Budget Document is the only reference material; however, its output indicators are not clearly defined, making M&E difficult and of little value.

Lack of feedback

M&E data must be fed back into the ministries in order to guide the implementation of ongoing activities and to plan for future ones. However, implementing agencies receive little feedback.

SOLUTIONS TO IMPROVE MONITORING AND EVALUATING OF THE BUDGET

The following solutions have been proposed to counter the challenges highlighted above.

Enhance monitoring of Other Recurrent Transaction expenditures

The monitoring of the Other Recurrent Transactions Budget needs to be strengthened. The Other Recurrent Transactions Budget gets more resources than the Development Budget. Strengthening monitoring can be achieved by encouraging ministries to indicate clear outputs to be achieved, which would make it easy to follow up on implementation progress.

Harmonizing M&E templates

The M&E departments in the government should develop a standard template that will gather information necessary for all departments. Concerned parties should collaborate to develop standard M&E templates.

Create an online M&E budget database

An Internet-based system should be created in which ministries input their quarterly budget progress reports for the recurrent and development budgets. The information would be readily available to all stakeholders because the reports would have been standardized.

Improve capacity building

There is need to build capacities of the M&E Section through trainings of officers to equip them with skills to conduct proper M&E of the budget.

Redefine outputs and their indicators

To avert the problem of unclear output indicators, it is imperative that the current budget indicators are clearly defined in collaboration with line ministries and with reference to the Malawi Growth and Development Strategy. Therefore, there is the need to conduct training for implementing agencies in order to improve on the development of outputs at the budget planning stage.

Improve on feedback

For lessons learning, relay the results to relevant stakeholders to redirect the course of implementation.

Conduct public expenditure reviews

Public expenditure reviews are a form of M&E that assists the government to make immediate

decisions in view of longer-term trends and helps bring more rationality to budget monitoring processes. Previously, public expenditure reviews were influenced by development partners such as the World Bank, hence the concentration on sectors where it has more interest. However, M&E departments should be able to initiate public expenditure reviews on their own by looking at the whole budget.

Independence of budget monitoring and evaluation

In Malawi, budget evaluation has some level of independence; there is limited interference in M&E activities. Politically, the President and Cabinet ministers are keen to learn from achievements made in the execution of programmes in the budget. However, there remains the need for the M&E units in the departments to be headed by more senior officers, so as to reduce interference.

There are some departments (e.g. State Residences, Office of the President and Cabinet and the Malawi Defence Force) where less M&E of budget programmes is done relative to other departments. This is mainly due to officers' concern over political pressure. This phenomenon has gone down since 2012, as the Office of the President and Cabinet is also working on the M&E activities.

Credibility of budget monitoring and evaluation

There is a degree of credibility of M&E in Malawi on the budget. However, it is being compromised by lack of expertise on evaluation. There is need for M&E officers to acquire specialized evaluation skills that improve evaluation quality. There is less evaluation being done on the budget, as M&E tends to focus on monitoring. The Public Expenditure Reviews have contributed to the evaluation of the budget, but since most of them are done by consultants, there is no skills-transfer to the officers to carry out evaluations of their own.

Use of budgetary monitoring and evaluation

Government decision makers' use of evaluation is limited. Experience has shown that evaluations on the budget are mainly used by the Ministry of Finance to assist in resource allocation and disbursement. To improve on the use of evaluations, there is a need for proper feedback to the implementing agencies. This could be done by having quarterly or annual budget reviews with the ministries in order to discuss evaluation results and to agree on the necessary actions to be taken.

CONCLUSION

Budgetary M&E in Malawi is vital for accountability and transparency in use of public finances and for guiding implementing agencies on budget execution. However, M&E on the budget faces a number of challenges that need to be addressed in order to yield better results. In regard to independence, credibility and use of evaluations, there is still a long way to go to attain the required targets as recommended by the the United Nations Development Programme.

REFERENCES

- Government of Malawi. 2011. "Malawi Growth and Development Strategy (MGDS) 2011–2016". Ministry of Economic Planning and Development.
- Government of Malawi. 2009. "Public Sector Investment Programme and Development Budget Comparison". Ministry of Economic Planning and Development.
- Government of Malawi. 2009. *Budget Manual*. Ministry of Finance.

23. MALAWI

PROMOTING USE OF EVALUATION RESULTS BY SENIOR GOVERNMENT OFFICIALS IN MALAWI

TED SITIMA-WINA
Permanent Secretary
Ministry of Economic Planning and Development

INTRODUCTION

In 2006, the Ministry of Economic Planning and Development in Malawi developed the National Monitoring and Evaluation Master Plan, which provides the main framework for monitoring economic and social development policies and programmes in the country. The government implemented the Malawi Poverty Reduction Strategy from 2002 to 2005, and has been implementing the Malawi Growth and Development Strategy I from 2006 to 2011; it is currently implementing Malawi Growth and Development Strategy II (2011 to 2016) as the main frameworks for implementing national development activities. These national development strategies show that all stakeholders have a role to play in implementing, monitoring, evaluating and reviewing development initiatives. To buttress the National Monitoring and Evaluation Master Plan, the Malawi government launched a 'sector working group' approach to implementing and managing development initiatives.

The roles and responsibilities assigned to various stakeholders have generated growing pressure on the Ministry of Economic Planning and Development to be more responsive to both internal (e.g. government officials) and external (e.g. civil society organizations) stakeholders on both downward and upward accountability to deliver the aspirations of the people and concrete development results. The penchant for results underscores the need to determine whether planned activities that are being implemented under a development strategy are inclusively improving the welfare of the people.

This pressure has prompted the Ministry of Economic Planning and Development to provide information and data through M&E systems (being established in all public institutions) to senior government officials, development partners and the public on a regular basis. The collection of information and its dissemination has been made possible by having a functioning national M&E system based on the National Monitoring and Evaluation Master Plan. The National Monitoring and Evaluation Master Plan describes the type of data to be collected for a specific time period for a particular national strategy per sector.

Previously, the challenge was that a great deal of data related to development initiatives was being collected by different stakeholders (e.g. research institutes, universities, government institutions, civil society organizations) and was hardly shared with the public. The establishment of the 'National Monitoring and Evaluation Master Plan' ensured that collected data would be harmonized, streamlined, disseminated, discussed and utilized for decision-making within the national monitoring framework.

The national M&E system has therefore become an important public-sector instrument in managing and monitoring delivery of development results and other services.

THE FRAMEWORK ARCHITECTURE AND INSTITUTIONAL ARRANGEMENT

Sector working groups

On 20 November 2008, the Government of Malawi launched and institutionalized sector working groups in its development approach. The overall aim of sector working groups is to provide a platform for negotiation, policy dialogue and agreement on plans, strategies and undertakings among stakeholders at the sectoral level. There are a total of 16 sector working groups. Each group is composed of representatives from ministries, civil society organizations, development partners and the private sector. Groups are chaired by a controlling officer and co-chaired by a development partner representative. Sector working groups also provide an operational framework for Malawi's development assistance strategy, which aims to ensure that external resources mobilized by the government are effectively utilized to implement the 'Malawi Growth and Development Strategy'. The goals of the development assistance strategy promote the five norms embedded in the Paris Declaration on Aid Effectiveness (2005) and the Accra Agenda for Action (2008): ownership, alignment, harmonization, managing for results and mutual accountability. The attraction of this approach is that although these documents were crafted to increase aid effectiveness, the principles are the same as those required to enhance the effective utilization of national resources.

To complete the equation of accountability, the Ministry of Economic Planning and Development put in place a programme to promote the use of evidence by controlling officers in managing development initiatives by establishing results-based management and M&E systems in all public institutions. The results-based M&E system has a component that emphasizes utilization of evidence-based evaluation results. This component aims to support the effective allocation of resources via an increased use of evidence to inform decision-making at programme, strategy and policy levels. In addition, it aims to enhance the likelihood that strategies and programmes will be effective in achieving their intended objectives, targets and results.

Utilizing evidence-based or evaluation results has the following objectives:

- Increasing the capacity of public policymakers and intermediaries to access, appraise and use available evidence;
- Increasing the capacity of policymakers and senior government officials to commission useful, relevant and high-quality research to inform their decisions;
- Establishing and strengthening working partnerships between government departments, research-institutions, universities, development partners and civil society organizations; and
- Informing capacity-building efforts that address skills gaps in application and understanding of evaluation results.

Technical working groups and committees

Technical working groups are key components in the development process, facilitating maintenance of technical standards and bringing best practices into the sectors. Membership is drawn from specific thematic areas, sector working groups and development partners to pursue actionable recommendations or decision points submitted through the secretariat. They are involved in reviewing technical studies and progress reports from implementing institutions and making recommendations to the sector working groups. The technical working groups are also responsible for selecting projects/programmes for implementation within the M&E framework.

OVERALL APPROACH

The approach focuses on the demand side through active engagement with senior government officials as research-users in which an environment is created conducive to understanding evidence-based assessments of development initiatives. The starting point is the Office of the President and Cabinet where a Monitoring and Evaluation Board was established. The board is chaired by the Chief Secretary; the Ministry of Economic Planning and Development is the secretariat.

Based on the Paris Declaration and the Accra Agenda for Action, the board, through the Secretariat, demands that each ministry provides output-based annual budgets that have clearly articulated indicators, annual targets and baseline data. The Board then demands monthly and quarterly progress reports from the secretariat, highlighting contributions made to achieving annual targets as set out in the national development strategies.

The M&E Division of the Ministry of Economic Planning and Development is responsible for coordinating all outcome and impact monitoring activities across all sectors in the country, and other *ad hoc* surveys conducted within and outside the national monitoring system. The evaluation/review results are synthesized into a management report and

submitted to the board. The secretariat is responsible for disseminating statistics and information through various channels (e.g. press releases, media briefings).

To ensure the functionality of the approach, line ministries have been provided with technical and financial support to build strong M&E sections that collect input and output indicator data and produce annual, quarterly and monthly sector review reports.

The planning process of the Malawi Growth and Development Strategy and the economic recovery plan has been participatory, with wide consultations resulting in clearly defined indicators and targets for all sectors. The district-planning manual is a framework that links national strategies with the subnational strategies. The district-planning framework serves as the basis for National Strategy Implementation monitoring at the district level. This entails close integration of national and district-level monitoring systems.

M&E committees have been established at the district (subnational) level as a mirror of committees at the national level. These committees facilitate the discussion and dissemination of district-specific evaluation results and draw participation from a range of community and other institutions at the district level. As beneficiaries, communities take an active role in the actual assessment of local development activities. Comprehensive Community-based Scorecards serve as project and programme assessment tools. The scorecards have been provided to select communities and are being used as a benchmark for assessing the quality of public service delivery and as a feedback system in the public policy arena.

Civil society organizations are also part of the institutional framework at the level of data collection, analysis and feedback. The M&E Division of the Ministry of Economic Planning and Development has made efforts to collaborate with civil society organizations so as to provide independence and credibility to evaluation results. Civil society organizations play a crucial role in implementing community-based monitoring activities.

At the national level, a technical working committee considers monthly, quarterly and annual technical reports and recommendations from the secretariat, which include progress reports from implementing institutions. The technical working committee makes recommendations to the Principal Secretaries Development Results Monitoring Committee. The technical working committee draws on the expertise of its members (primarily research centres) on policy and impact analysis. This ensures the independence and credibility of evaluation results. Reports from the Principal Secretaries are summarized with key recommendations and submitted to the Office of the President and Cabinet for decision-making through the board.

There is also an annual Development Stakeholders Forum, which draws participants from civil society organizations, the donor community, media, academia and the private sector. It acts as a national accountability forum where issues of underperformance (based on the annual review of the Malawi Growth and Development Strategy) are raised and the government is called upon to act. The forum also facilitates information dissemination, discussion and policy recommendations.

CHALLENGES, OPPORTUNITIES AND CONCLUSION

Despite the aspirations to design and deliver coherent and harmonized M&E systems in Malawi, a number of challenges have worked to hamstring such systems' operationalization and functionality. These include weak stakeholder capacities to collect data at the district, council and sectoral levels, non-compliance to scheduled reporting timelines, and weak skills in data analysis, data quality, and storage and dissemination of development results to facilitate usage and accountability. The demand for evaluation results and usage is obvious, but the supply of credible research results is limited. For example, most of the demand is coming in the context of mid-term and end-term impact evaluations of policy and project interventions by government, donors and civil society. The skills and personnel to undertake such key development inputs are trifling or negligible.

On the supply side, Malawi does not have the type of research institutions that are active and prominent in other countries (e.g. independent think tanks, research institutions, evaluation networks or associations). Nevertheless, the establishment of the 'National Monitoring and Evaluation Master Plan' provides an opportunity for evidence-based evaluation utilization. In addition, some evaluation institutions, such as 3ie and the University of Johannesburg, have shown interest and commitment to collaborating and working with the Ministry of Economic Planning and Development and local universities to harness the creation and use of evidence-based evaluation. This activity signifies South-South capacity-building initiatives.

REFERENCES

- Government of Malawi. 2006. "National Monitoring and Evaluation Master Plan". Lilongwe, Malawi.
- Government of Malawi. 2005. "Institutionalizing Sector Working Groups to Strengthen the Implementation of the Malawi Growth and Development". Lilongwe, Malawi.

24. SOUTH AFRICA

REFLECTIONS ON THE SOUTH AFRICAN EXPERIENCE WITH EVALUATION AND THE USE OF EVALUATIVE EVIDENCE TO ORIENT PUBLIC POLICY FORMULATION

IAN GOLDMAN, Head of Evaluation and Research
STANLEY NTAKUMBA, Chief, Director, Policy and Capacity Development
CHRISTEL JACOB, Director, Evaluation and Research Unit

Department of Performance Monitoring and Evaluation
South African Presidency

BACKGROUND

South Africa has a semi-federal system with three spheres of government (national, provincial and local). Therefore, M&E has to be applied at all three spheres.

EVALUATION PRIOR TO DEVELOPMENT OF THE NATIONAL EVALUATION SYSTEM

Prior to 1994, rigorous M&E activities started in South Africa in relation to donor support for non-profit organizations. The New Public Management Approach that gained popularity in the 1990s and the results-based management paradigms of the 2000s brought in demand for a greater hierarchical alignment between activities and different levels of outcomes. A 2007 study by the Public Service Commission noted that M&E was generally conducted in an “isolated or vertical manner” and not integrated into a comprehensive system (Government of South Africa 2007b).

During the 2000s, there was a growing interest in M&E and pressure mounted to introduce a more coherent approach to government-wide monitoring and evaluation. In 2005, the cabinet approved a plan for the development of a government-wide M&E system. This system was envisaged as a 'system of systems', in which each department would have a functional monitoring system, out of which the necessary information could be extracted. The policy framework to guide the overarching government-wide M&E system was published in 2007 (Government of South Africa 2007a).

The government that came to power following the 2009 elections faced a number of service delivery challenges, resulting in a greater willingness to be frank about the quality of public services, corruption and other governance problems. There was a political consensus to improve government performance, including through a greater focus on M&E. In 2009, a Ministry of Performance M&E was created in the Presidency; a Department of Performance M&E was created in 2010.

DEVELOPMENT OF THE SOUTH AFRICAN NATIONAL EVALUATION SYSTEM

South Africa's M&E work in the 2000s focused on monitoring, although some departments did undertake evaluations. The findings of the Department of Performance M&E's first round of implementation of the Management Performance Assessment Tool indicated that only 13 percent of the 103 national and provincial departments were conducting evaluations.

In 2011, a study tour to Columbia, Mexico and the United States led to development of a national evaluation policy framework, adopted by the Cabinet in November 2011. The strategic approach that has been taken focuses on important policies, programmes and plans; those selected are embedded in a national evaluation plan approved by the Cabinet. The focus has been on utilization, the public availability of all non-confidential evaluations, and the existence of improvement plans (which are then monitored). This approach emphasizes learning over punitive measures, building evaluation into the culture of departments rather than promoting resistance and malicious compliance.

Several types of evaluations are envisaged, including diagnostic, design, implementation, impact, economic and evaluation synthesis. These types may be combined for specific evaluations. This means that evaluations are not only undertaken at the end of an intervention, but can be conducted at any stage in the life cycle—before (diagnostic), during (implementation) or at the end of a phase (impact). An economic evaluation can be performed at any stage.

Evaluations are implemented as a partnership between the department(s) concerned and the Department of Performance M&E (which partially funds the evaluations). An Evaluation and Research Unit (supported by a cross-government Evaluation Technical Working Group) has been established in the Department of Performance M&E to drive the system and provide technical support. Guidelines were developed for standards for evaluation, competencies for programme managers, M&E staff and evaluators, and training courses started in September 2012. Five courses were developed and are being rolled out.

In June 2012, the cabinet approved the first national evaluation plan, with eight evaluations. The second national evaluation plan, with 15 evaluations planned for 2013 to 2014,

was approved in June 2012. In total, 24 evaluations have been commissioned and three have been completed at the time of this paper.

USING EVALUATIVE EVIDENCE TO ORIENT PUBLIC POLICY FORMULATION

The Department of Performance M&E recently commissioned an audit of evaluations conducted for the government between 2006 and 2011. The evaluations reflect a range of uses:

- Revising policy (e.g. Diagnostic Review of Early Childhood, Impact Evaluation of the Reception Year of Schooling);
- Revising plans and priorities (e.g. Schools that Work, Report on the State of the Environment, Overview of Health Care, Impact Evaluation of the Reception Year of Schooling);
- Changing funding regimes (e.g. Child Support Grant);
- Changing programmes (e.g. Mid-term Review of the Expanded Public Works Programme); and
- Increasing the knowledge base (e.g. Profile of Social Security Beneficiaries).

THE DEMAND FOR EVALUATIVE EVIDENCE

What evidence of demand is there?

In 2012, the Department of Performance M&E undertook a survey in order to assess the state of M&E in the national and in provincial governments (Government of South Africa 2013). In terms of culture-based barriers, more than half of the respondents (54 percent) indicated that problems are not treated as opportunities for learning and improvement. Other noteworthy responses indicated that senior management often failed to champion M&E (45 percent), M&E is regarded as the job of the M&E Unit rather than of all managers (44 percent), there is not a strong M&E culture (40 percent), M&E is seen as policing and controlling (39 percent) and M&E units have little influence. These all point to the challenge in using M&E as a strategic function to inform policy and decision-making.

In 57 percent of cases, M&E information had limited or no influence on decision-making. Nearly half of the respondents (46 percent) regarded integration with policy development as either non-existent or very limited. Just under half of the departments (48 percent) reported that integration of M&E with budgeting is limited. This lack of integration implies a poor environment for the demand and use of M&E evidence, since it is likely to be viewed as a stand-alone activity detached from other key management processes.

Encouraging demand and use

There are several elements of the national evaluation system that are explicitly designed to ensure that evaluations are demanded and findings are implemented. These include:

- Departments are requested to submit evaluations, rather than being told they will be subject to evaluations. This means they are more likely to want them and to want the results;
- The Cabinet approves the National Evaluation Plan, which means that there is at least some central high-level interest and awareness of the evaluations;
- Evaluation Steering Committees are chaired by the department requesting the evaluation, while the Department of Performance M&E provides the secretariat and is therefore able to ensure that the national evaluation system is complied with. Evaluation Steering Committees have significant power, including approval of terms of reference and reports;
- Provisions for a management response so that departments have the opportunity to respond to evaluation findings; and
- All evaluations must be followed by an improvement plan, which is monitored for two years.

Despite these elements, there is reticence among some managers, partly because the results of evaluations are made public and they are wary of being exposed (as indicated earlier, in most departments the identification of problems is not seen as an opportunity for learning). Other measures being undertaken to stimulate demand include:

- Making presentations at senior management fora;
- Developing a course for senior managers in Evidence-based Policymaking and Implementation; and
- Making parliamentary portfolio committees aware of how they can use M&E findings to support their oversight functions. This year saw the first portfolio committee (Mineral Resources) request a department to submit evaluations in the call for evaluations in 2014 to 2015. A capacity development programme to support the parliament's use of information from the Department of Performance M&E's is being planned.

While the Department of Performance M&E has been concentrating on the 15 evaluations per year in the National Evaluation Plan, it has also been stimulating the demand for evaluations more widely. It has piloted the development of provincial evaluation plans with the Western Cape and Gauteng Provinces, and three departments have now developed departmental evaluation plans. This aims to stimulate a wider use of evaluation than could be covered under the national evaluation plan, and also to stimulate departments and provinces to think of what they should cover themselves (as opposed to those with major national interest and covered in the national evaluation plan).

The Department of Performance M&E is also stimulating improved accountability by making evaluations publicly accessible. The audit of evaluations mentioned previously is being made available through an evaluation repository on the Department of Performance M&E website, and all evaluations undertaken through the national evaluation plan will be made public, once they have been submitted to the cabinet.

LESSONS FOR A UTILIZATION-FOCUSED EVALUATION SYSTEM

Supply of evaluation in South Africa has historically been weak, and there has also been limited demand (though with patches of excellence). In general, there has been insufficient evidence use across government, and a tendency for political judgement rather than political decisions informed by strong evidence. Since 2010 there has been an increasing supply of evidence catalysed by the Department of Performance M&E (initially monitoring evidence, with evaluation evidence starting to emerge). The Department of Performance M&E is also about to examine the roles it should play in promoting research evidence. The Cabinet's positive response to the Department of Performance M&E's systems for evaluation and its systems of management performance assessment and front-line service delivery monitoring point to the cabinet's receptivity for good evidence. There are issues, however, about the consequences of problems identified, which creates the incentives for addressing the challenges identified.

The design of the evaluation system is focused on utilization, seeking to build from a demand-driven system. As the 23 evaluations underway start to report from June 2013, and improvement plans are developed and monitored, it will be interesting to see how challenging findings are taken by departments and how far the findings are taken up in practice.

The next year, to mid-2014, will test the evaluation system and show how well the ambitious system that has been established is achieving what it intends to, where evaluations are informing significant changes to policy, decision-making and implementation. Key questions will be:

- How well is the Department of Performance M&E able to play its independent role, especially when evaluations are challenging?
- Can the 'learning as opposed to punitive' focus of the system be strengthened?
- Should the wider public be more strongly involved in the evaluation system, and if so, how?

REFERENCES

- Engela, Ronette, and Tania Ajam. 2010. "Implementing a Government-wide Monitoring and Evaluation System in South Africa". ECD Working Paper Series No. 21. Washington: World Bank.
- Goldman, Ian, Ronette Engela, Ismail Akhalwaya, Nolwazi Gasa, Bernadette Leon, Hassen Mohamed and Sean Phillips. 2012. "Establishing a National M&E System in South Africa". Nuts and Bolts of M&E 21. Washington: World Bank.
- Government of South Africa. 2013. "A Survey of 96 National and Provincial Departments on the State and Use of M&E in the Public Service". Department of Performance Monitoring and Evaluation. Pretoria.
- Government of South Africa. 2007a. "Policy Framework for a Government-wide M&E System". Presidency. Pretoria.
- Government of South Africa. 2007b. "Report on the Audit of Reporting Requirements and Departmental Monitoring and Evaluation Systems within Central and Provincial Government". Public Service Commission. Pretoria.

25. EGYPT

INSTITUTIONALIZING AND STREAMLINING DEVELOPMENT MONITORING AND EVALUATION IN PRE- AND POST-REVOLUTIONARY EGYPT

DOHA ABDELHAMID

EvalMENA Board Member

International Organization for Cooperation in Evaluation

BACKGROUND

M&E in Egypt has taken stunning steps since the beginning of the new millennium. Although far from independent or widely used, diagnostic and readiness assessments were taken as tools to enhance credibility.

In early 2000, the Government of Egypt recognized that budgetary processes seemed to create obstacles that inhibited the changes the government desired to induce performance. The Ministry of Finance therefore requested the World Bank to review the budgetary process in light of the identified obstacles. Looking through an Egyptian lens, the cause for development M&E was taken up by a champion minister who managed to assemble a like-minded ministerial team. Upon concluding a diagnostic study by the World Bank, it was recognized that Egypt possessed leadership, incentives and drivers, mandates and clear authorities, a well-defined strategy for results-based budgeting, pockets of innovation, links to resource decisions, a workable strategy to implement and a set of donor-sponsored activities. The diagnosing team met and interviewed government officials, academics, donors and others. This was an early opportunity to interview and advocate for the use of M&E that was just beginning to bud in the Ministry of Finance. Development M&E was a completely new concept to Egyptians—and still is.

STRATEGY: THE FIRST READINESS ASSESSMENT, CHAMPIONS SELECTION, CAPACITY BUILDING, IMPLEMENTATION AND ADVOCACY

In September 2001, upon completion of the readiness assessment, the Ministry of Finance announced the launch of the national capacity-building programme in performance-based budgeting development M&E. The minister communicated to the public the various approaches to results and budgeting orientation that had been adopted worldwide, and relayed his vision and provisional implementation strategy that he hoped to realize in collaboration with other pilot line ministries.

He divulged the results of the World Bank's diagnostic mission and the road ahead to implement a development M&E/performance-based budgeting (PBB) capacity-building effort in five pilot ministries at the initial stage of project implementation. Training efforts and awareness raising to ministries commenced at this stage. The national strategy was announced in a widely covered international economic conference held in Egypt. The Government of Egypt's programme was launched as a pilot, starting in 2003. The pilot programmes chosen were closely linked to public service delivery in key development sectors such as budget and resource allocation, education, health and taxation.

Advocacy work proceeded at different levels, including communications with parliamentary members on programme progress, periodic meetings among champion ministers to review progress and facilitate overcoming obstacles, forming sub-ministerial committees to coordinate operations at individual line ministries, assigning national experts to design capacity-building programmes at the entry, intermediate and advanced levels of development M&E and conducting meta-reviews by visiting international experts collaborating with the Ministry of Finance.

Press conferences were held to pave the way towards creating a culture of performance. Progress on building government capacities in performance-based budgeting M&E was used to create other products, bulletins and publications citing and documenting government reforms. Other reports were generated by development partners (e.g. UNDP and the World Bank) operating in Egypt and supporting the development M&E reform programme. The partners acknowledged the unmatched Egyptian success story led by the Government of Egypt.

PROGRAMME PROGRESS AND RESULTS

Within project life (one and a half year), 1,500 government officials were trained on the 10 steps to establishing sustainable M&E systems at various levels of sophistication and different economic sectors. Technical coaching resulted in preparing multi-year performance-based budgets for the pilots. As an incentive towards this paradigm shift, the Minister of Finance pledged to retain ministerial budgetary appropriations at least equal to what line ministers had requested earlier. This was coupled with constant communication with the non-champion ministers, parliament's powerful planning and budget committees, the media and the public. Civil society was experiencing serious bottlenecks and was versed in neither development M&E nor its importance.

Upon building national capacities in the government through the pilots, the Minister of Finance, together with the champion ministers, decided to sustain the effort to the future. In the meantime, the Ministry of Finance received requests from all non-pilot ministries to subscribe to the programme as a result of its demonstrated success. The Minister of Finance issued a decree to establish and fund sustainable M&E units in the state administrative apparatus on the night of the government shuffle that overthrew the champion ministers, wasting the opportunity to upscale and continue the programme in pilots and other ministries.

Overnight, political priorities changed, and the Egyptian success story turned to failure. That said, the pilots and pockets of innovation proceeded with the work advocated for and based on their faith and buy-in for the cause of development M&E.

THE ROLE OF DEVELOPMENT M&E IN CIVIL SOCIETY

The first Egyptian non-government network evolved in 2005 upon the conclusion of the International Development Evaluation Association's first biennale conference. The network connects around 40 professional development M&E experts and is in the process of formalization as the Egyptian Development Evaluation Association (EgyDEval). The network runs discussion groups and advocacy workshops for national evaluators and the media. A second non-government network was formalized with the support of United Nations Children's Fund (UNICEF) in early 2013 and subscribes around 200 research and evaluation experts. It has run a number of awareness creation workshops, launched a diploma in public research and evaluation with one of the state universities, and has invested in 'Arabizing' leading world-class publications in development M&E.

THE SECOND READINESS ASSESSMENT STUDY

An independent readiness assessment was conducted in the interim period between the two revolutionary waves in Egypt (January 2011 and June 2013). The assessment was conducted by two scholars, one of which is the author of this paper. After the first national experience with national capacity building in the Egyptian government prior to termination in July 2004, a mapping exercise of available innovations were cited. These included civil society organizations, government institutions, university curricula, donor-sponsored capacity-building programmes and others. Sixty-two development M&E professionals were interviewed through face-to-face meetings and/or surveyed electronically. They spanned development partners; donors; university scholars; political parties; national, regional and international development M&E networks; government institutions; research centres and think tanks; parliaments and legislative bodies; media; and independent consultants.

The self-designed, mixed-method readiness assessment analytic tool is composed of four main sections: background and information on the respondent, development M&E status quo in the country and its institutional level, the impact of the revolution on country and institutional development M&E, and a set of probing queries on a vision for the future.

CONCLUSIONS

The diagnostic tool derived a set of informative conclusions on the status of development M&E in Egypt in the post-revolutionary era. They are summarized as follows:

- Almost 90 percent of those surveyed suggested that there is a lack of interest and understanding in Egypt of development M&E, and its importance in the context of development projects;
- Two thirds of respondents view development M&E as a demand-driven activity within their organizations. In a number of cases, respondents cited development M&E reporting being commissioned in response to donor demands;
- Respondents were asked specifically at what point in development programmes' project flow evaluations should take place. The majority suggested that evaluations should be conducted at each point in the project implementation stage, but there is a greater majority who see development M&E as a necessity on the completion of projects;
- The majority of organizations perform development M&E at the project level and to a lesser extent at the programme, policy and sector level.;
- Most of those surveyed have significant experience (more than five years) in development M&E;
- Respondents are split on matters related to development M&E data accessibility. Slightly more than half felt that the data required to conduct a satisfactory evaluation of a development project is not easily accessible or available. The disparity in responses could be due to differing resources across organizations;
- Half of the organizations implement development M&E at the working team level. However, the majority have less than five staffers involved in evaluation;
- Most employees working in development M&E hold at least an undergraduate qualification, with almost half possessing an advanced university degree. This, together with the years of experience mentioned above, indicates that the development M&E employees surveyed are well-educated and trained in the development M&E field, which lends extra credibility to the responses' robustness;
- Most respondents prepared M&E reports on quarterly and annual basis. Many respondents confirmed that development M&E reports are prepared on *ad hoc* basis, based on project demand;
- Development M&E reports are disseminated through websites and to internal stakeholders;
- Development M&E reports are mostly prepared at project end and mainly on donors request rather than being embedded within project design. The purpose is mainly for securing future funding requirements, either for the same project or others;

- Reports are, if and when, used to address operational issues in the course of project implementation. In other instances, reports are used in budget negotiations for future project results improvements. Sometimes, development M&E reports are used in higher policy-level discussions and especially in support of policy amendments;
- Over two thirds of respondents offered capacity-building opportunities on development M&E to their organizations and others. The majority of capacity-building efforts are delivered through on-the-job-learning;
- Almost two thirds of respondents believed that development M&E has no effect on policy formulation or implementation. In addition, development M&E reporting is insufficiently appreciated as a tool for planning and budgeting;
- Within organizations, development M&E is seen as a useful tool when it comes to designing and streamlining projects and also in providing insights and policy recommendations for future projects. However, this does not often get translated into a tangible benefit for projects when it comes to decision-making at the management and government levels;
- Practitioners compiling development M&E reports listed many challenges to completing them. The primary issue, facing 43 percent of organizations, is a lack of data access and, when accessible, its inaccuracy. When data is accessible, organizational staffers often lack the necessary analytical skills. Both of these can serve to undermine the effectiveness and credibility of development M&E reports by giving conclusions that are not based on reliable data or analysis;
- Another key challenge is a perceived lack of awareness of or appreciation for the value of development M&E in Egypt. Even donors who encouraged evaluations were more focused on evaluation of inputs than results. Media was complacent to exposing the negative results of government performance;
- Respondents affirmed the challenge of lacking development M&E capacity-building opportunities and its considerable high cost;
- The perceived effect of the 25 January 2011 revolution on the development industry is negative. Eighty percent of respondents contend that the revolution either had a negative or no effect on the realization of development objectives in Egypt. Many respondents believe that Egypt has taken a step back from where it was in the pre-revolution context when it comes to the realization of development objectives; and
- Specific to development M&E, there is a mixed opinion, with just over half of responses indicating that the 2011 revolution had some influence on the development M&E reporting in Egypt. On the negative side, the revolution resulted in greater degree of economic and political instability, which not only makes development M&E reporting more difficult to conduct because of funding constraints, but also undermines its significance in the context of a country in crisis. On the other hand, the increased role of youth dictates higher demands for development projects and their

results. Some respondents also viewed the sweeping constitutional reforms that followed the revolution as an opportunity for development M&E in Egypt, as they hope that some provision can be made for mandatory evaluations going forward.

SOLUTIONS

- Fuel demand for the institutionalizing development M&E in state administrative bodies in order to fulfil citizenry needs and revolutionary motos;
- Create an enabling environment through socio-economic pressures and lobby for effective development projects that yield results through the adoption of development M&E techniques and tools and ensure mutual accountability;
- Facilitate donor harmonization of efforts in support of creating and effecting a culture of performance within the government;
- Actively engage civil society organizations through government partnerships to build capacity, conduct development M&E evaluations and encourage their dissemination as learning tools;
- Develop consulted, comprehensive national capacity-building programmes in cooperation with non-governmental organizations, independent think tanks, training institutes and universities;
- Embed policy, programme and project evaluation in the constitution, and ensure the competent authority is equipped with the capacity required to do the job; and
- Improve utilization, standardized quality measures, enforcement mechanisms, independence, incentives, codes of ethics, transparency and public accountability.

REFERENCES

- Hassanein, Medhat, and Doha Abdelhamid. 2013. "Results-Oriented Budgeting in Egypt". *Interdisciplinary Journal of Research in Business*. 2(9):13-25.
- El Baradei, Laila, Doha Abdelhamid and Nermine Wally. 2013. "Institutionalizing and Streamlining Development Monitoring and Evaluation in Post-Revolutionary Egypt". Presentation to IDEAS Global Assembly 2013, Evaluation and Inequality: Moving Beyond the Discussion of Poverty, 6-9 May 2013, Bridgetown, Barbados.

26. MOROCCO

GENDER EVALUATION OF PUBLIC POLICIES FROM A HUMAN RIGHTS PERSPECTIVE

CHAFIKI MOHAMED

Director of Studies and Financial Forecasts
Ministry of Economy and Finance

Public policy corresponds to a choice of values implemented in order to achieve an objective defined by the political authority. Policy is therefore defined via the goals that it sets, the means allocated to achieve these goals, the expected results and the nature of the intervention's impact on the target population.

Therefore, evaluation of public policies is based on a comparison between their achievements and the goals they were initially assigned, taking into consideration the logic behind the intervention (normative dimension). The aim is to develop public action or to define new, more efficient and more relevant, policies (instrumental dimension). The evaluation of a policy or a public action consists, therefore, in evaluating its effectiveness by comparing its results to the goals it set and the means that were used, based on the concept of performance indicators. It therefore becomes an essential decision-making tool.

This is the framework for the gender evaluation of public policies from a human rights perspective; it starts with the construction of concepts and tools. Its aim is to strengthen the mechanisms that make it possible to ensure that citizens have equal access to civil, political, social, economic, cultural and environmental rights.

RELEVANCE OF THE GENDER EVALUATION OF PUBLIC POLICIES FROM A HUMAN RIGHTS PERSPECTIVE

An analysis of public policies based on human rights is in line with the founding principles endorsed by international consensus around new visions of human development. The human rights perspective places the analysis of inequalities that hinder development processes at the forefront of its concerns, with the aim of realigning development trajectories on a more inclusive, fairer basis that is concerned with achieving expected results in terms of enjoying rights.

Plans, policies and development mechanisms are therefore drawn upon to carry out an overhaul of their logic and their objectives, while respecting the system of rights and obligations established by international law. The ultimate goal of such an overhaul is to promote the sustainability of development initiatives while encouraging the autonomy of individuals and the diversification of choices available to them in order to allow them to live decent and dignified lives.

The gender approach, which is based on a search for optimum effectiveness and performance in public policies and a better targeted impact with regards to the differentiated needs of the various components of the population, is perfectly in line with human rights-based approaches. Indeed, the two approaches share a number of elements. They are built partly around a framework of analysis applicable to all development activities that takes into consideration the different needs and roles of men and women in a given society. They are also built on a normative framework based on rights and obligations.

In addition, both approaches are interested in the impact of the implemented policies and programmes on the well-being of particular groups. The two approaches are based on a range of concepts and information where M&E tools play an important role. The idea that the development process should include a dimension of long-term social and economic transformation because it is strategic supposes that, for a pertinent evaluation of progress, relevant indicators have to be found with regards to sustainable development objectives based on the principles of equal opportunities and equal access to the fruits of this progress.

PRINCIPLES GOVERNING THE GENDER EVALUATION OF PUBLIC POLICIES FROM A HUMAN RIGHTS PERSPECTIVE⁸⁵

Application of a human rights-based approach on public policies and programmes is based on translating universal standards of human rights into measurable indicators. The obligation of accountability leads to the translation of the normative content of a right into quantitative indicators and is perfectly in line with the principle of results and performance-based budget management.

85 Report by the United Nations High Commissioner for Human Rights, April 2011.

CHOOSING THE INDICATORS⁸⁶

The United Nations High Commissioner for Human Rights has developed a framework for monitoring the achievement of human rights. The framework uses of a range of indicators that are structural, procedural and results-based. These parameters establish relationships among a state's commitment and acceptance of the obligations stemming from international human rights standards (structural indicators), the efforts it made to achieve these obligations, the measures and programmes it put in place (indicators of method) and the results obtained (results indicators).

Structural indicators relate to the ratification and adoption of legal instruments and the existence of the core institutional mechanisms considered necessary to facilitate the achievement of the human right in question. Structural indicators must highlight the nature of national legislation applicable to the right in question, and check that the legislation takes international standards into account.

Indicators of method provide information on the way in which the state endeavours to achieve their human rights obligations through specific public programmes, measures and interventions. These efforts demonstrate state's acceptance of standards relating to human rights and their intent to obtain the results associated with the achievement of a given right.

Results indicators make it possible to measure to extent to which a right has been achieved and to appreciate the extent to which it can be exercised. It should be noted that results indicators can come from a number of underlying processes that can be highlighted by a number of method indicators linked to a number of rights (e.g. indicators relating to life expectancy or mortality can be linked to the vaccination of a population and education or greater awareness of public health, as well as to the availability of an adequate food supply and individuals' ability to access it).

MOROCCAN EXPERIENCE OF THE GENDER EVALUATION OF PUBLIC POLICIES FROM A HUMAN RIGHTS PERSPECTIVE

Morocco's gradual experimentation over the last 10 years with results-focused gender sensitive budgets (carried out by the Finance and Economy Ministry in partnership with UN Women) is an inherent part of an overall approach that tends to integrate internationally recognized human rights principles into the legal, institutional and governance sectors. In addition, this approach constitutes an important advantage for evaluating public policies from a gender perspective, with regards to expected outcomes in the populations that have already been targeted and whose needs are clearly differentiated.

At the operational level, the pragmatic and progressive approach that has been followed since 2002 with the view to integrating gender dimensions into budget programmes has allowed Morocco to develop a range of analytical instruments based on dedicated budgetary guidelines, underpinned by a range of gender-sensitive synthetic indicators that strengthen the M&E mechanisms of gender-related public policies.

86 Report on the use of indicators for the promotion and monitoring of the implementation of human rights, United Nations High Commissioner for Human Rights, May 2008.

The gender budget rapport, an excellent instrument for effecting gender-sensitive budgets at the national level, has accompanied the presentation of the Finance Law since 2005 and, in parallel, has enjoyed a pragmatic approach that is continually progress-oriented. Consequently, it has gone through a number of development phases that, in 2012, led to the adoption of the analytical approach based on the gender evaluation of public policies through a human rights perspective. This approach coincides in its method, design and philosophy with the great institutional progress that Morocco has seen through its 2011 adoption of a new Constitution, which entrenched the principles of the respect for the human rights of equality, fairness and citizenship. This is coupled with localized management that should get stronger within the framework of the advanced regionalization project and where the demands for responsibility and accountability will be fully respected.

This is also the spirit of the new Constitution, which represents major progress in developing the rule of law in Morocco. The Constitution is qualified by a charter of citizens' freedoms and fundamental rights similar to those in the Universal Declaration of Human Rights. In the preamble, which forms an integral part of the Constitution, the Kingdom of Morocco reaffirms its commitment to human rights as they are universally recognized. The Constitution recognizes the superiority of international conventions, duly ratified by Morocco over domestic law, which is within the framework of the measures of the Constitution and the laws of the kingdom, within the respect of its immutable national identity and from the moment the conventions are published. The rights described by the new Constitution include civil, political, economic, social and cultural rights. Title II of the Constitution is entirely dedicated to fundamental rights and freedoms. It includes 21 articles reinforcing civil, economic, social, cultural, environmental and development rights.

The results obtained through a human rights-based analysis in the evaluation of public policies illustrate the relevance of this approach. The process has made it possible to identify the advances made in a number of areas and to pinpoint the challenges that hinder equal access to rights. The analysis carried out covered the programmes of 27 ministerial departments spread across three centres in line with three categories of rights: fair access to civil and political rights, fair access to social rights and fair distribution of economic rights. Ministerial departments included:

- **Fair access to civil and political rights:** Justice, Social Development, General Directorate of Local Municipalities, Civil Service, Economy and Finances, Foreign Affairs, and Cooperation and Communication.
- **Fair access to social rights:** National Human Development Initiative, Water, Energy, Environment, Habitat, Infrastructure and Transport, Health, Education and Literacy, Professional Training and Youth.
- **Fair distribution of economic rights:** Employment, Agriculture, Fisheries, Trade & Industry, NICTs, External Trade, Tourism, Crafts and Social Economy.

The Moroccan experience in terms of the gender evaluation of public policies through the perspective of human rights has made it possible to identify certain requirements to

ensuring the optimal application of this approach, including: a good appropriation of the normative framework governing the respect of human rights; a strong interest in analysing the coherence and the convergence of public policies with regards to the indivisibility of human rights; and a rich and regularly updated information system.

It is in this sense that the gender approach carried out from a human rights perspective is such a crucial breakthrough in terms of progress. It should make it possible to restructure sustainable development trajectories along new paradigms that are more conscious of the need to respect human dignity and more favourable towards inclusion and equality. This is particularly relevant in a changing world that is looking for new development models that are oriented towards the guaranteed access to different categories of human rights. For this reason, it proves to be a precious tool, not only for the democratic evaluation of public policies but also for the renovation of their design and the optimization of their impact.

27. MOROCCO

EVALUATION OF THE NATIONAL INITIATIVE FOR HUMAN DEVELOPMENT: INDEPENDENCE AND USE

MOHAMED MOUIME

Director, Information Systems Division
National Observatory for Human Development

PRESENTATION OF THE NATIONAL OBSERVATORY FOR HUMAN DEVELOPMENT

The decree establishing the National Observatory for Human Development was published in the Official Bulletin of 11 September 2008. (Decree No. 2-08-394 of 23 October 2008). However, the National Observatory actually began its work in December 2006, following His Majesty the King's launch of the National Initiative for Human Development in May 2005. Its implementation, evaluation in particular, is defined in the guidelines of the Royal speech: "The implementation of the National Initiative for Human Development will, moreover, be an opportunity to emerge in our country, a true social engineering through innovation in the types of intervention, efficient ways and maximum impact, supported by qualified human resources and mechanisms for the vigilant and objective observation of the phenomena of poverty and social exclusion."

By its very nature cross-tracking implemented public policies in this area, the Observatory reports directly to the prime minister. It must, therefore, contribute to evaluation, funding and producing advice and recommendations on the progress and the constraints to human development in the Kingdom.

To accomplish its general mandate, the Observatory has, under the authority of a president nominated by His Majesty the King, a slight administration and a council of

24 members chosen *intuitu personae*, among the high responsibility for public administration, civil society actors, academics and private-sector operators. As part of its core business, the council submits annually to the appreciation of His Majesty the King an annual report on human development.

CONTENT OF THE NATIONAL INITIATIVE FOR HUMAN DEVELOPMENT

The National Initiative for Human Development for 2005–2010 was composed of four key programmes:

1. The programme to fight against poverty in rural areas, which extends to the poorest 403 rural communities, representing 3.8 million people;
2. The programme to fight against social exclusion in urban areas, which encompasses 264 of the most disadvantaged districts, representing 2.5 million people;
3. The programme to fight against precariousness, which comprises eligible projects that support 50,000 people living in extreme precarity throughout the country; and
4. The transversal programme that is flexible to answer the needs not initially planned, which comprises eligible projects with high impacts on human development, such as income-generating activities and the strengthening of social engineering and stakeholder capacities.

Two evaluations have been done since the National Observatory for Human Development's official establishment in 2006. The first, a mid-term evaluation, was conducted in 2008. The second evaluation was conducted in 2011 and focused on the socio-economic impacts of the initiative.

EVALUATION OF THE NATIONAL INITIATIVE FOR HUMAN DEVELOPMENT: WHAT INDEPENDENCE?

The mid-term evaluation focused on:

- **The population's perception of this initiative:** The referential of this initiative has served in this study as a starting point to examine perceptions. This referential includes a development philosophy and a mechanism for the mobilization and participation of the population, enabling analysis of how this referential is perceived by the population and by stakeholders.
- **The development of a protocol for conducting participatory assessments as part of the Initiative:** To address this area, we chose an approach of analytical restitution of field measurements in two stages. First, we analysed the legal and institutional framework of the initiative (texts and circular, the structures established). Second, we analysed how processes actually took place, targeting implementation, and possibly its sustainability, in order to document, *in situ* and on a daily basis, the ways of doing things.

- **Convergence of the initiative's programmes with those conducted by different departments:** This evaluation consisted of a comprehensive diagnosis of the convergence of the programmes of the initiative and those conducted by the public sector and in the proposition of recommendations for greater convergence at the institutional and territorial levels in programmatic and operational aspects.
- **The role of associations in the Initiative:** With the advent of the Initiative, the vital role of civil society is now explicitly recognized in human development programmes. To this end, the study objectives were to observe and assess the situation of human development activities carried out by civil society and assess their strengths and weaknesses in order to strengthen them and support their dynamics.
- **The analysis of physical and financial achievements:** Analysing the data of nearly 20,000 projects and activities carried out under the initiative.

The first four qualitative studies were performed by private consultants and academic researchers. The results analysis was conducted by a private consultancy. The results of these studies were synthesized by an international consultant hired for this purpose by the Observatory.

THE IMPACT ASSESSMENT

After conducting a feasibility study of the impact assessment in 2007, the Observatory has decided to conduct a study (with the assistance of an international expert) to evaluate the impacts on the socio-economic conditions of rural and urban populations targeted by the initiative.

The methodology proposed by the expert involves, in particular, comparing selected outcome indicators in targeted areas and not targeted areas, between a start date and an end date. That requires indicators on both dates for each area.

As there had been no survey at the start date of the initiative (i.e. there is no baseline data), the Observatory found itself compelled to conduct a two-passage survey as quickly as possible. The first passage was made in 2008 and the second, marking the end of Phase I of the initiative, was made in 2011. The total sample size of the survey was 3,570 households visited in 2008 and subsequently revisited in 2011.

A questionnaire was designed for households and their members, as well as a 'rural town' and an 'urban area' designation. The questionnaires were designed to meet the needs of the evaluation, after consultation with national and international experts and all relevant departments (e.g. National Coordination Initiative, ministries of education, employment, health, social development, housing).

In rural areas, 403 rural municipalities with a 2004 poverty rate of at least 30 percent were selected for the fight against poverty initiative. The rural sample was composed of 124 rural communes with similar poverty levels (between 27 percent and 32 percent), half of which were among the targeted communes of the initiative. The impact evaluation of the initiative for rural areas was made on the basis of regression discontinuity design.

For the urban sample, the methodology benefited from the existence of a panel survey in two periods, making it possible to measure the various outcome variables twice for the same households. Assuming that the evolution of different outcome variables before the implementation of the programme was the same for the treated areas and non-target areas, the change of the outcome variables between the two types of areas between 2008 and 2011 will measure the impact of treatment by the initiative.

The criteria for selecting the districts included inadequate housing, the lack of employment opportunities and limited access to infrastructure and basic social services. Focusing on cities with a population of more than 100,000, 264 areas were selected for the initiative programme on the fight against social exclusion in urban areas.

The urban sample consisted of 114 neighbourhoods with characteristics close to the available indicators and criteria used; 75 are among the areas targeted by the initiative. (For more information on methodological aspects, see the full report, available on National Observatory for Human Development website.⁸⁷)

The two phases of the survey were carried out by a private consultancy firm under the supervision of the international expert and a university professor at the National Institute of Statistics. In addition, both experts conducted the survey processing and estimation of the impacts.

EVALUATION OF THE NATIONAL INITIATIVE FOR HUMAN DEVELOPMENT: WHAT DOES IT DO?

On the mid-term evaluation, the National Observatory for Human Development sent in for the first time a report of the evaluation to His Majesty the King and Head of Government. Two presentations of the evaluation's main conclusions were then made, the first one to the head of government in the presence of ministers, and the second to the media and representatives of the departments and agencies concerned.

After that, the Observatory transmitted the report of the mid-term evaluation to all government departments and local authorities, and published it on the Observatory website.

On 31 July 2009, His Majesty the King, in his speech to the nation on the occasion of the tenth anniversary of the celebration of the throne, highlighted the need to submit this initiative to evaluation, and to ensure the recommendations of the National Observatory are followed.

To translate these recommendations into action, the government established three committees made up of representatives from the government, non-governmental organizations, local authorities and elected officials. The first commission is responsible for the convergence projects. The second commission should produce guidelines in order to ensure the sustainability of projects. The third commission was mandated to develop M&E tools.

87 See ondh.ma.

On the evaluation of the impact of the initiative, the Observatory sent the report incorporating the findings of this evaluation to His Majesty the King and Head of Government on 18 May 2013.

The Observatory intends to follow the same approach that was developed for sharing and using the findings of this evaluation.

REFERENCES

- Heckman, J. J., and E. J. Vytlačil. 2005. "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica* 7:669-738.
- Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142(2):615-635.

28. CANADA

ALPHABET SOUP: LINKING THE LETTERS TO PROMOTE EVALUATION USE

MARTHA MCGUIRE

Board Member, International Organization for Cooperation in Evaluation
EvalPartners Management Group, Cathexis Consulting

JAMES MELANSON

Director, Development Evaluation
Foreign Affairs, Trade and Development Canada

INTRODUCTION

Looking at the various networks and organizations that support development evaluation can be a bit overwhelming. Within Canada, we are fortunate to have the Canadian Evaluation Society, a well-established voluntary organization of professional evaluators (VOPE). On the government side, there is the Centre for Excellence in Evaluation (CEE) of the Federal Treasury Board Secretariat, the Canadian Association of International Development Professionals (CAIDP), and Development within the Foreign Affairs, Trade and Development Canada (formerly the Canadian International Development Agency [CIDA]). At the international level, there is the International Organization for Cooperation in Evaluation (IOCE), the International Evaluation Partnership Initiative (EvalPartners) and the International Development Evaluation Association (IDEAS). On the education side, there is the Consortium of Universities for Evaluation Education (CUEE) within Canada. Have we missed any within the Canadian context? I am sure we have. Then there is the world.

This paper explores the complexity of the evaluation world first through a single country lens, then through the benefits of intra-country linkages and how these linkages and their inter-country connections contribute to creating enabling environments and promoting evaluation use.

THE CANADIAN EVALUATION SOCIETY

The Canadian Evaluation Society, established in 1981, advances evaluation, theory, knowledge and practice through leadership, advocacy and professional development. It is a bilingual organization (English and French) with 11 regional chapters and more than 1,900 members. The Canadian Evaluation Society offers:

- An annual national conference;
- *Canadian Journal of Program Evaluation*;
- Canadian Evaluation Society Fellowship;
- Connection to a national and international community of evaluators;
- Professional Designation Program;
- Professional development opportunities;
- Reduced rates for membership in the American Evaluation Association and the Australasian Evaluation Society;
- Canadian Evaluation Society Ethics and Evaluation Standards (adopted those developed by the Joint Committee for Evaluation of Education Programs); and
- Discounted association-based insurance plans and Professional Liability Insurance.⁸⁸

The Canadian Evaluation Society maintains formal relationships with the International Organization for Cooperation in Evaluation/EvalPartners, CUEE, the American Evaluation Association and Canadian Evaluation Society chapters. However, many of its members belong to other organizations within Canada—and globally—including IDEAS, CAIDP and other VOPEs.

FOREIGN AFFAIRS, TRADE AND DEVELOPMENT CANADA (FORMERLY CIDA)

The Government of Canada is committed to sustainable international efforts that help people who live in poverty in the developing world. Guided by its Aid Effectiveness Agenda, Canada is making its international assistance efficient, focused and accountable.

Canada's development assistance focuses on three priority themes:

- Increasing food security;
- Securing the future of children and youth; and
- Stimulating sustainable economic growth.

88 Available at evaluationcanada.ca/txt/ces_brochure_e.pdf.

In addition to these priority themes, Canada leads global commitments such as improving maternal, newborn and child health. This initiative supports efforts to reduce the number of preventable deaths and improve the health of the world's most vulnerable mothers and children.

Canada also integrates three cross-cutting themes into all of its programmes and policies:

- Supporting environmental sustainability;
- Promoting equality between women and men; and
- Strengthening governance institutions and practices.

Canada reports on its programmes and projects and shares the results widely, focusing on maximizing value, delivering results, increasing ownership of development outcomes, increasing transparency and accountability, and building partnerships to accelerate development results.⁸⁹

AN OVERVIEW OF EVALUATION NETWORKS AND ORGANIZATIONS

With the initiation of EvalPartners in 2012, there is increasing emphasis on partnerships and working together towards the common goals of:

- Increasing individual evaluation capacity;
- Increasing institutional capacity of VOPEs; and
- Creating enabling environments within which evaluation can occur.

CONCLUSIONS

While new linkages are being formed, the lack of linkages within countries and across VOPEs needs further development. Key questions for discussion include:

- What are the benefits of intra-country linkages among the various evaluation organizations?
- How can the intra-country linkages contribute to improved inter-country linkages?
- Does there need to be a balance between intra-country and inter-country linkages? Between formal and informal relationships?
- How can that balance be attained?

89 Available at acdi-cida.gc.ca/acdi-cida/ACDI-CIDA.nsf/eng/ANN-5682956-GNR.

REFERENCES

- Canadian Evaluation Society. Brochure. Available at evaluationcanada.ca/txt/ces_brochure_e.pdf. Last accessed 5 February 2014.
- Foreign Affairs, Trade and Development Canada. 2013. "About International Development". Available at acdi-cida.gc.ca/aboutcida. Last accessed 5 February 2014.
- Rugh, J., and M. Segone, 2013. *Voluntary Organizations for Professional Evaluation (VOPEs): Learning from Africa, Americas, Asia, Australasia, Europe and Middle East*. United Nations Children's Fund (UNICEF), EvalPartners and International Organization for Cooperation in Evaluation.

29. ALBANIA

CHALLENGES OF A NEW EVALUATION NATIONAL SOCIETY

FATION LULI

Albanian Society of Programme Evaluation

MARIE GERVAIS

Quebec Society of Programme Evaluation

BRIEF OVERVIEW OF THE NATIONAL EVALUATION EFFORT IN ALBANIA

The post-Communist era has led to the establishment of democracy in Albania and created an environment favourable to the development of a culture of healthy governance, managing for results and accountability. In addition, the stability of democracy observed over the past years has paved the way for significant changes, calling for greater effectiveness of the government and for more transparency regarding expenditures and outcomes. From this perspective, a major effort was made to renew the Albanian public administration. The context was thus favourable for the development of the evaluation.

However, it must be recognized that evaluation functions remain underdeveloped in Albania, and the degree that evaluation has been institutionalized within the structures of the country is weak to non-existent. It is an environment where everything has to be built. There is neither supply nor demand for evaluation, which is evident by the lack of technical and financial resources. Programme evaluation is generally justified by audit activities, monitoring or procedures to improve the quality (e.g. seeking certification from the International Organization for Standardization [ISO]), thus creating confusion about the true nature of evaluation.

From this perspective, and considering possible strategies to develop an evaluation culture and build evaluation capacity in Albania, the establishment of a national society was chosen as the priority strategy to put forward.

TABLE 1: ASPE PATHS OF ACTION TO STRENGTHEN INDIVIDUAL, INSTITUTIONAL AND NATIONAL EVALUATION CAPACITIES

<p>ACTIONS TO STRENGTHEN INDIVIDUAL CAPACITIES</p> <p>University education Develop a course in the universities of Shkoder and Elbasan Develop training sessions with the Training Institute of Public Administration of Albania (ITAP) Explore the medium-term prospect for development of a diploma in evaluation</p> <p>Employment training Develop a training module in evaluation for government ministries, municipalities, communes and civil society Provide training to Albanian trainers.</p> <p>Professionalization Train a skills- and profile-evaluator Draft standards of practice, ethical principles and certification requirements Translate evaluation documents into Albanian Develop an Albanian section of the Better Evaluation website.</p>
<p>ACTIONS TO STRENGTHEN ASPE INSTITUTIONAL CAPACITIES</p> <p>Create thematic committees Training Promotion Charter of evaluation</p> <p>Communicate with members Animation of the website Development of a distribution</p> <p>Generate knowledge of the member profiles</p> <p>Build relations with the partners</p> <p>Identify opportunities to support capacity building</p>
<p>ACTIONS TO DEVELOP AN ENABLING ENVIRONMENT FOR EVALUATION</p> <p>Develop advocacy to institutionalize evaluation of public policies Create information seminars for the parliamentary deputies for the development of a legal framework in evaluation Develop a national evaluation policy</p>

USE OF EVALUATION

The first steps of the Albanian Society of Programme Evaluation

The processes leading to formalizing the Albanian Society of Programme Evaluation (ASPE) began in April 2011. A first informal network was formed in July 2011; ASPE was legally established and officially recognized by the Albanian government in October 2011. The reasons for establishing the society were to raise awareness, create supply of and demand for evaluation, and build a community of evaluation stakeholders. The Quebec–Canada learning case

was chosen after a brief analysis of different models of national evaluation societies by the ASPE Board of Directors.

Mission and strategic plan

The ASPE mission includes contributing to the development of programme evaluation in Albania and promoting the use of evaluation in public and private organizations. In accord with this mission, ASPE developed its strategic plan 2012–2015 in four components: promoting evaluation, professional international cooperation, professional development and teaching, and funding.

Strengthening individual, institutional and national evaluation capacities

Exchanges conducted with Albanian stakeholders made it possible to identify trends concerning the future paths of action for ASPE to take in three areas: strengthening individual capacities, strengthening ASPE institutional capacities and developing an enabling environment for evaluation (see Table 1).

Much can be done with little

It is possible to create a national evaluation society without external funding, based primarily on the work of volunteers and partnerships. However, this has its limits and can jeopardize the short-term survival of such a society. The issue of funding is at the heart of the development capacity of voluntary organizations of professional evaluators.

Recognize the importance of the human factor

Maintaining the active engagement of volunteers over time as well as their ability to work in a team are essential ingredients to ensuring the progress of a project to create a national association in evaluation.

Timely networking

Experience has shown that it is more productive to focus efforts on networking projects and issues that allow the quick development of collaborations than on going for early but general support without specific targets, which quickly generates a decrease in interest.

Innovation

To have quickly sought the mentorship of a more experienced national society such as the Societe Quebecoise d'Evaluation de Programmes (SQEP, Quebec Society of Programme Evaluation) gave ASPE founding members confidence in the path taken. This synergy enabled them to test some decisions and strategies before their implementation in addition to benefiting from the experiences of others.

NEXT STEPS

ASPE still faces many challenges that must be strategically addressed in order to prevent burnout or unproductive dispersion of resources. ASPE should now deliver on its strategic plan and the EVALbania initiative. Thus, rapid ASPE growth must translate into stable and

continuous actions on the ground, which will confirm its legitimacy and strengthen its credibility as an Albanian actor working towards a healthy democracy and good governance.

Although it has demonstrated that it can create and maintain in its early stages of existence a national evaluation society without direct political support and with very limited resources, ASPE must now assure the interest, support and funding to achieve its ambitions. Further, it must stimulate interest in evaluation and promote the commitment of the public sector in partnerships, at different levels of government, in academia and internationally. ASPE will quickly seek the necessary funds to start the first outreach and capacity-building activities.

ASPE is aware of the challenges of leading an emerging community of evaluators. How to interest and engage evaluators, ensure their professional development and support their professional identity, in addition to enhancing quality of the practice in such a context? Efforts to develop the services offered to its members, more meetings with key stakeholders involved in the development of evaluation in Albania and to support the professionalization of evaluation will also take place. On this point, ASPE intends to devote some time to reflecting on skills, standards of practice and ethical principles in Albania.

Finally, ASPE is aware of the great scale of efforts to be deployed in Albanian society without an established evaluation culture. This makes ASPE's role as promoter of evaluation to institutions and individuals is even more important, as it carries significant potential for influence.

CONCLUSIONS

ASPE has progressed rapidly in less than two years. It has formalized, obtained legal status and determined its statutes. Equipped with a strategic plan, ASPE has signed a collaboration agreement with SQEP, developed the structuring initiative EVALbania with support of EvalPartners and SQEP, and began international outreach.

A movement in favour of developing a culture of evaluation is clearly growing in Albania, and national structures, including the main stakeholders and civil society, are interested in actively contributing to it. The desire to develop a national evaluation society and then advance the evaluation process within different institutions and national structures is challenging, particularly when the country lacks an evaluation culture or national evaluation capacities. How does an organization operate in an environment where everything in evaluation must be built and promoted? That's why the following goals are important:

- Develop a vision of evaluation and development strategies that correspond to the specific characteristics of the country;
- Proceed with realistic and achievable steps that allow small successes that continue to encourage and stimulate;
- Get in touch and communicate with the various stakeholders interested in evaluation; and
- Adjust regularly, given the responsiveness of the environment and resources actually available.

30. UNITED STATES OF AMERICA

THE DIVERSITY OF THE EVALUATION ENTERPRISE INFLUENCES ITS USE IN THE UNITED STATES

STEPHANIE SHIPMAN

Assistant Director, U.S. Government Accountability Office

JODY FITZPATRICK

Professor, University of Colorado

HISTORY AND ORGANIZATION OF EVALUATION IN THE UNITED STATES

Public programme evaluation began in the United States in the 1960s with the advent of President Lyndon Johnson's new programmes to address the causes and consequences of poverty. This legislation required programmes to be "evaluated," a new term then. Those early efforts were generally conducted by university-based researchers in the fields of education, psychology and sociology, hired by public agencies to fulfil the requirement to assess their success. Many early evaluators also worked closely with the programmes in order to provide evaluation information for programme improvement as they tested out new policies and practices.

In its organization, evaluation in the United States today reflects its beginnings; evaluation continues to primarily be the province of the social sciences. The field is not centralized, but quite diffuse, performed by university faculty members, evaluation contract organizations, self-employed evaluators and public agency personnel. This diversity also reflects the variety of parties involved in programme decision-making and the decentralized structure of political power in the country. In many policy areas, the 50 states and localities have primary responsibility for public policy and programming, while the federal government provides limited financial and technical assistance. In addition, private charities also fund and provide health and social services. Thus, federal, state and local public agencies, as well as private

charities or foundations all may request or conduct evaluations to address accountability and programme improvement concerns.

In the federal government, individual agencies typically contract with evaluation professionals to assess: (1) the implementation or outcomes of federal programmes and policies; (2) the quality of agency or programme management; or (3) the effectiveness of specific practices employed within a federal programme. Federal agencies are not generally legally required to conduct evaluations of their programmes, although many individual programmes do have legislated evaluation requirements. A recent survey of federal managers showed that only 37 percent reported that an evaluation had been conducted on a programme, policy or activity they worked on in the last five years (U.S. Government Accountability Office 2013). However, since the Government Performance and Results Act of 1993, all federal agencies are required to establish annual performance goals and report on their progress in achieving them (performance monitoring). In the legislative branch, the U.S. Government Accountability Office also conducts audits and evaluations of federal agencies, programmes and policies to support the Congress' oversight responsibilities.

The federal government's interest in evaluation has waxed and waned over the years. Although progress has been made since reforms were enacted in the 1990s to improve financial and performance management in the federal government, more work remains to increase the use of programme evaluation in government. Although a recent survey showed only a little over one-third of programmes had had evaluations in the last five years, of those that did, 80 percent reported they contributed to a moderate or greater extent to improving programme management or performance. Thus it would appear that agencies' lack of evaluations may be the greatest barrier to their ability to inform programme management and policymaking.

EFFORTS TO FURTHER EVALUATION USE

The literature has identified distinct challenges to the use of evaluation and similar forms of evidence by politicians and agency managers: (1) the credibility of the evaluation study; (2) agency evaluation capacity; (3) stakeholder involvement in the evaluation; and, of course, (4) the policy context of decision-making.

For example, studies must be timely, relevant and credible to decision makers. Credibility is influenced by the choice of study design, the competence of its execution and by the author's institutional affiliation and reputation for quality and objectivity. Ambiguity in study findings and lack of a clear, actionable message also hinder use. Of course, agency managers must also be familiar with research methods and be comfortable relying on that form of evidence to form and justify their decisions. However, findings that conflict with their experience, or strongly-held beliefs, may be brushed aside. Because so many choices must be made in defining the evaluation question, selecting a design and analysing results, these choices should be informed (but not dictated) by the questions and concerns of the anticipated decision makers. Finally, evaluators operating in a national government context are not naïve; budgetary, practical and political realities often constrain or preclude policy-makers' willingness or ability to act on evaluation findings.

The Office of Management and Budget within the Office of the President has encouraged agencies, both formally and informally, to expand their evaluation efforts and use evidence and evaluation in budgetary, management and policy decisions to improve government effectiveness. However, progress has been uneven. As identified in the Government Accountability Office study discussed above, selected agency evaluation officials recommended three basic strategies to facilitate use of their studies: (1) demonstrate leadership support of evaluation for accountability and programme improvement; (2) build a strong body of evidence by attending to rigour in whatever methods are used; and (3) engage programme stakeholders throughout the evaluation process (U.S. Government Accountability Office 2013).

Agency leaders demonstrate support for evaluation fundamentally through forming and funding designated units with the necessary analytic expertise, and turning to them for advice on management and policy decisions. These units may conduct or oversee studies and can develop evaluation tools and guidance to ensure that consistent, high-quality work is done across their agency. It is rare, of course, that a single study will change policy. Instead, evidence accumulated over time, across locations, using various methods, provides stronger conclusions than does a single study, and thus, has the power to change assumptions and understanding of programme performance.

Because of the quick pace of policy discussions, developing a portfolio of studies over time better prepares an evaluator to answer whatever questions are posed. Experienced agency evaluators emphasize the importance of building good relationships and trust with programme managers to gain an understanding of their needs and build credibility for—and interest in—their studies. These evaluators consult programme managers while developing their work agendas and designing individual studies. They train programme staff in measurement and evaluation methods, and provide them with ongoing technical assistance in developing performance-monitoring systems and interpreting study results, which facilitates use.

The Government Accountability Office and the American Evaluation Association aim to complement these efforts to further programme evaluation and its use in the US government through promoting evaluation as a key management function and helping organizations and individuals develop evaluation capacity. With 7,000 members and two academic journals, the American Evaluation Association is the primary organization of evaluation professionals in the United States and supports evaluation capacity development through a variety of national and international activities.

PROMOTING EVALUATION AS A KEY MANAGEMENT FUNCTION

The Government Accountability Office encourages agencies to conduct evaluation by holding them accountable for reporting to the public and using data to improve programme performance. The Government Performance and Results Act requires federal agencies to develop strategic and annual performance plans, performance goals for all their major programmes and to report annually on their progress in meeting those goals. Programme evaluations are not required, but their findings must be included in the annual performance reports. These planning and reporting activities are intended to provide both congressional

and executive decision makers with more objective information with which to make tough choices to reduce government costs and better link resources to results. In an era of shrinking federal resources, Congress expects agencies to provide evidence of effectiveness in achieving meaningful outcomes for taxpayers and holds them accountable for making the most efficient and effective use of public resources. Primarily in response to congressional requests, the Government Accountability Office has reviewed the quality, coverage and relevance of agency evaluations.

To inform congressional deliberations, the Government Accountability Office is often called on to summarize the available research evidence on existing programmes, practices or policy proposals. This, of course, requires assessing the quality of those studies and may result in a critique of the quality and relevance of the agency's research programme. The Government Accountability Office is often asked to assess the quality of agency performance data, their reliability, and whether they represent programme results as well as the quality and efficiency of programme activities. In subject areas such as education, where the federal government's primary role is to provide leadership rather than direct services, the Government Accountability Office has surveyed local programme administrators to assess their perspectives on the relevance, credibility and availability of a federal agency's published research and evaluation.

The American Evaluation Association works to establish the credibility of evaluation through its Evaluation Policy Task Force, a permanent task force within the American Evaluation Association that is intended to influence not only evaluation policy but also to enhance policymakers' and managers' understanding of evaluation practice. The Task Force promotes government evaluation policies and legislation through consultations with congressional and executive branch parties and international evaluation organizations. It developed *An Evaluation Roadmap for a More Effective Government*, which focuses on how evaluation can be organized and used within the federal government and promotes evaluation as a key management function (American Evaluation Association 2010). This paper stresses the importance of establishing policies and procedures to safeguard the objectivity and independence of the evaluation process from political interference; ensuring evaluation credibility through commitment to the evaluator's objectivity and the use of the most rigorous, appropriate methods available; and addressing programme stakeholders' concerns and information needs in order to ensure evaluation is useful for programme management and policymaking. The 'roadmap' was revised slightly in 2013 and continues to be a major document for the Association to use in consulting with federal officials and elected decision makers.

ASSISTING EVALUATION CAPACITY DEVELOPMENT

To support progress towards a more results-oriented federal government, the Government Accountability Office promotes high quality, useful agency evaluations by identifying effective practices in the areas of performance M&E and by publishing guidance on evaluation methods and practices. The Government Accountability Office reports have highlighted good practices in the areas of measuring impacts on long-term goals, forming effective collaborations with programme partners, building agency evaluation capacity,

setting policy-relevant research agendas and facilitating evaluation use. In addition, the Government Accountability Office published its own guide to designing evaluations in order to share its lessons with other agencies (U.S. Government Accountability Office 2012). That guide stresses the importance of developing background on the programme and policy issues and thoroughly exploring alternative design options in advance to ensure the study will meet decision makers' needs most effectively and efficiently.

The American Evaluation Association helps develop organization and individual competencies by publishing two academic journals, promoting *Guiding Principles for Evaluators* and *The Program Evaluation Standards*, and by supporting a variety of skill development opportunities. Around 3,000 people attend the annual conference, which provides opportunities to share knowledge, experience and innovative evaluation approaches, as well as numerous formal workshops for skill development. The organization also supports online learning opportunities throughout the year and is beginning to explore certification of educational programmes for evaluators and credentialing of evaluators as another means to further the credibility of the profession, its practitioners and their work.

The American Evaluation Association's *Guiding Principles for Evaluators*, a code of ethics for evaluators, was developed in 1995 and is reviewed and revised every five years by the independent American Evaluation Association Advisory Group. These principles are organized into five categories: systematic inquiry, competence, integrity/honesty, respect for people and responsibilities for general and public welfare. Educating evaluators, major stakeholders and the public in the *Guiding Principles* and their application is one important way for the American Evaluation Association to maintain the credibility of the evaluation profession with its clients and other stakeholders. The *Guiding Principles* are published in every issue of the *American Journal of Evaluation*, brochures on the Principles are provided to evaluators to distribute to clients and training and discussion sections are held on the Principles at annual conferences.

The *Program Evaluation Standards*, developed and revised periodically by the Joint Committee on Standards, is another method the American Evaluation Association uses to maintain the quality and credibility of evaluations. The American Evaluation Association representatives and those of other professional associations concerned with evaluation and assessment developed the standards and continue to be involved in their revisions and dissemination. The standards are not an ethical guide for the behaviour of practitioners, but instead are standards for developing or judging an individual evaluation. The Joint Committee has identified five standards for a good evaluation plan: utility, feasibility, propriety, accuracy and accountability. Each standard is defined and developed into sub-categories to be used in evaluation, and their application is illustrated through case studies (Yarbrough et al. 2011). The standards serve as another means for educating evaluators and users in what is expected in evaluation and, in so doing, helps maintain the credibility of evaluation.

Both the American Evaluation Association and the Government Accountability Office participate in national and international organizations and networks to share and develop information and resources. Informally, the Government Accountability Office staff support

a knowledge-sharing network of federal evaluation officials and participate in national and international audit- and evaluation-related professional organizations and conferences. For example, the Government Accountability Office participates in the International Organisation of Supreme Audit Institutions, which has a Working Group on Program Evaluation that seeks to extend programme evaluation to other national audit organizations around the world. The American Evaluation Association has long collaborated with international evaluation organizations in the discussion of evaluation policies in the public and private non-profit sectors.

REFERENCES

- American Evaluation Association. 2013. *Guiding Principles for Evaluators*. Fairhaven, Mass.
- American Evaluation Association. 2010. *An Evaluation Roadmap for a More Effective Government*. Fairhaven, Mass.
- U.S. Government Accountability Office. 2013. "Program Evaluation: Strategies to Facilitate Agencies' Use of Evaluation in Program Management and Policymaking". GAO-13-570. Available at gao.gov/products/gao-13-570. Last accessed 5 February 2014.
- U.S. Government Accountability Office. 2012. "Designing Evaluations: 2012 Revision". GAO-12-208G. Available at gao.gov/products/GAO-12-208G. Last accessed 5 February 2014.
- Yarbrough, D.B., L.M. Shula, R.K. Hopson and F.A. Caruthers. 2013. *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users* (3rd ed.). Thousand Oaks: Sage Publications.

31. WORLD BANK

SCIENCE OF DELIVERY: IMPLICATIONS FOR MONITORING AND EVALUATION

ARIANNE WESSAL, Extended Term Consultant

CLAY G. WESCOTT, Senior Consultant

MATTHEW COURTNEY TREUTH, JR, Intern

Independent Evaluation Group, The World Bank Group

INTRODUCTION

The developing world has seen rapid improvements in recent years. The number of people in poverty has been cut by over half since 1990. Over the same period, 2.1 billion people gained access to improved drinking water, and the child mortality rate has dropped by 41 percent (United Nations 2013). However, rapid improvements are creating expectations for more equitable and just patterns of progress. The tensions we see in Brazil, Egypt, India and Turkey arise in part from the gap between rising expectations of citizens and their everyday experience (Woolcock 2013).

The different aspects of development are uneven, with more people in the world owning mobile phones (6 billion) than having access to toilets and latrines (4.5 billion) (UN News 2013). In addition, poverty-reduction efforts in some geographic regions have not enjoyed the same level of success as in other regions. Over the last two decades, the number of people living in extreme poverty in sub-Saharan Africa rose from 290 million in 1990 to 414 million in 2010. The failure to achieve poverty reduction goals at the regional level raises the question of why some efforts fail while similar efforts deliver quality results in other regions.

To address these challenges, many developing country governments are trying to understand why the policies put in place to reduce poverty and build prosperity are not leading

to the results they want. One way forward could be a new form of knowledge, the 'science of delivery'. This concept is borrowed from the health care field, where the previous emphasis on understanding the causes and consequences of health issues is shifting to give more attention to organizing, managing and financing health promotion (Catford 2009). Applied to the field of public management, a science of delivery should provide mechanism-based explanations of how and why the implementation capability of countries varies, as well as a guide to action (Woolcock 2013).

This approach differs from the institutional reform model that currently dominates the public management field. In the institutional reform model, 'best practice' solutions are often chosen without significant consideration being given to their external validity. In this model, the focus is on inputs delivered rather than on outputs obtained and projects are often given unrealistic expectations. The result of this approach is that projects frequently fail to achieve their goals, while the specific reasons for this failure remain unknown.

In order to remedy these issues, the science of delivery tailors project components based on local factors such as implementation capacity and political support. As problems arise, consideration is given to concerns at the political, organizational, and project levels before deciding on a solution. Project managers are encouraged to draw on aspects of past successful projects, try new concepts and adapt to changing conditions. The science of delivery approach requires intensive field research, improved data collection at the project level through the use of good monitoring systems and the diffusion of ideas to enable these changes in implementation and management.

The result of using a science of delivery approach is the creation of localized projects that provide both impactful results to the target community as well as useful data and information to the public. This data gives project managers the ability to understand how and why a project was effective rather than just whether it was or not. The science of delivery allows project managers in a region to better understand why their projects fail to achieve their desired impact, as well as give them the ability to draw on lessons learned from successful projects in other regions.

At the same time, there have been recent theoretical advances in many scholarly fields ranging from systems engineering, medicine, economics and public management that are being exploited to help countries organize the emerging evidence on successful delivery to help them improve development results (Kim 2012). These new sources of knowledge help aid managers in adapting their projects to local conditions, ultimately resulting in a higher level of success.

MAIN CONTENTS

The World Bank and other development partners can point to many examples of delivery success, drawing on a treasure trove of evidence obtained using a mix of qualitative and quantitative methods linking successful delivery of interventions with local politics, culture, capacity and other factors that affect delivery outcomes. However, some of this experience is not easily accessible, buried in lengthy reports, files, datasets and as tacit knowledge in the heads of staff and evaluators.

This deeply contextual approach to learning needs to be accessible to practitioners. Where learning is generalizable, there needs to be mechanisms for taking ideas to scale through communities of practice and other forms of diffusion and implementation. Key elements of the science of delivery are to ensure that projects or interventions have adequate M&E mechanisms built in to the project and to ensure these are linked to feedback loops that will ensure continual learning, experimentation, results monitoring and redesign based on experience.

A World Bank project example of this is the Karnataka watershed (*sujala*) project in India, which used real-time M&E to improve targeting and efficiency during delivery, and at project's end documented such huge gains that the project was replicated and scaled up (IEG 2011). Two other examples are Oportunidades (formerly the Program for Education, Health and Nutrition), a health and education conditional cash transfer programme in Mexico, and the Program of Advancement through Health and Education (PATH) in Jamaica. These programmes have built strong monitoring systems at the beginning of the programmes with short-, medium- and long-term outcomes identified.

The programmes undertook regular assessments at each step of implementation and used this in conjunction with monitoring information to make adjustments as the programmes are implemented. In the case of PATH, process evaluations and spot checks were undertaken for activities being implemented. (Rawlings 2009). This enabled the identification of a number of problems, including: stakeholders saw the application process as burdensome and were not clear on programme rules; the system for verifying the eligibility of new beneficiaries was weak; and there was a strong unmet demand for jobs and training. This process led to a decision to revamp the management information system, revise the operations manual, use social workers as focal points to access social services, and create a 'STEPS to Work' programme focused on skills development and employment.

Both programmes demonstrate that implementation of a strong M&E system where information is used for decision-making can yield better development results. Evaluations of PATH showed that it was better at reaching the poor than other Jamaican safety net programmes, while evaluations of Oportunidades showed the programme had a significant positive impact in improving health and education. Both programmes have been lauded for reaching their target populations and yielding better results than other programmes.

Oportunidades is a great example of improved science of delivery through the use of both a strong M&E system and of information learned from past projects that warrants a closer look. The programme began in 1997, providing monetary educational grants to poor rural families for each child less than 22 years of age who was enrolled in school between the third grade of primary and third grade of high school. In addition to education, Oportunidades also has health and nutrition components. Government health institutions provide families with preventative health care. Families also received, in addition to a fixed monthly transfer to improve food consumption, nutritional supplements for young children and their mothers. Where Oportunidades truly shines is in quality at entry.

At implementation, project managers planned to have an independent evaluation done by the International Food Policy Research Institute. They also drew on lessons learned from past projects, recognizing that giving money to female heads of families results in better

financial outcomes. These steps taken during the implementation stages translated into quality results that were reflected in the independent evaluation. The evaluation reported that improvements had been made in increasing school enrolment, nutritional quality and access to medical care. At the time of the evaluation, Oportunidades was said to have increased secondary school enrolment rates by over 20 percent for girls and 10 percent for boys (Parker 2003). This was the first randomized controlled trial of a large programme used in developing country social policy.

The success of Oportunidades caught the eye of the Mexican federal government, although the evaluation methodology has been criticized for its sampling design, inadequate treatment of selective attrition and sample contamination (Faulkner 2012). As of 2003, 46.5 percent of Mexico's federal annual anti-poverty budget was devoted to Oportunidades. This increase in funding allowed Oportunidades to expand to urban areas and to provide high school students with education grants. In summary, steps taken at implementation to improve the science of delivery were crucial in the success and subsequent expansion of Oportunidades. Learning from past projects and having quality external evaluation ultimately led Oportunidades to become one of the most successful conditional cash transfer programmes to date. The close involvement of scholar-practitioners helped to design new conceptual approaches, ensure technical soundness and rigorous monitoring, protect the programme during changes of administration and spread the approach around the world (Lustig 2011).

OTHER EXAMPLES FROM BRAZIL, INDIA AND NIGERIA

In 2008, Brazil began its Second Minas Gerais Development Partnership Project, a sector-wide approach project of over \$1.4 billion aimed at improving the efficiency of public resource use, supporting innovations in public management, and supporting the State Government of Minas Gerais in strengthening its M&E system (World Bank 2008). Funds were disbursed to ten eligible expenditure programmes in five sectors.

To better utilize learning loops in this project, an extensive results monitoring framework was built into the programme at implementation. Individual projects were subject to monthly monitoring and quarterly management meetings were made accessible to the press. The government made yearly implementation data available on the Internet to increase programme transparency. In order to increase its focus on outcomes delivered, the World Bank supported the project by developing a household survey, quality assurance surveys and a series of impact evaluations in the education, health and transport sectors.

These monitoring systems gave managers constant feedback and allowed them to work towards achieving medium-term goals on their way to achieving long-term objectives. The latest Implementation Status and Results Report rated progress towards achieving project development objectives and implementation progress as satisfactory (World Bank 2013a). So far, the programme has succeeded in reducing the amount of time needed to start a business at Minas Facil in Belo Horizonte from 26 to 7 days. The Poverty Reduction Program has already exceeded its initial objective by benefiting over 26,000 rural families.

Projects that provide quality M&E frameworks and learning loops are only one aspect of the science of delivery. The community must utilize this data to understand the specific aspects of projects that contribute to their success. The Social Observatory project in India is a learning organization that works to make effective use of the data that is collected on the project level (World Bank undated). Their learning system consists of the following components. First, they look at real time monitoring to deliver change at the project level. Second, they facilitate long-term learning through quantitative and qualitative impact evaluations. Third, they conduct special case studies to understand key issues for project implementation and design. The result of this research is a better understanding of the impact of a specific intervention on the desired outcomes. Project managers are able to build upon the results found by the Social Observatory to localize their projects and improve the science of delivery.

A very different approach is evident when considering recent road construction in Edo State, Nigeria. Coming out of an intense civil conflict in 2009, the newly elected governor wanted to deliver critically needed roads quickly to gain the confidence of citizens that his government, unlike past regimes, could be effective in delivering public goods. His solution was to depart from standard good practice norms and processes. He set up a centralized team under his direct control charged with contractor selection, budgeting, fiscal management and monitoring. Five contractors won 83 percent of the total value of the contracts because they were the ones that the Governor trusted the most. Unlike standard procurement tenders, this one was based on only sketchy designs and estimates that needed to be fleshed out as the project proceeded.

There were uncertain timelines caused by a highly erratic cash flow to the governorate from central government transfers. The central team developed a project-monitoring dashboard, and used it to work with contractors to change contract budgets and other parameters as designs became more complete and as cash became available. Getting this system to work relied on harnessing the private sector's capacity to enhance the ability of the Ministry of Works to supervise project contractors and to orient project procedures to deliver rapid results. As a result, Edo State's capital spending quadrupled from 2008 to 2012, 85 percent of the roads were completed, and engineering design analysis found that the roads were built to acceptable standards and cost (World Bank 2013b).

CONCLUSIONS

The examples of Oportunidades and the sector-wide approach in Brazil demonstrate how M&E systems can be improved through the use of independent evaluations and household surveys. The Social Observatory project in India stresses the importance of real time data usage and the exploitation of learning loops. The example in Edo State demonstrates that taking context into account is key in improving delivery. For Nigeria, a highly centralized approach allowed for a greater level of project monitoring. By creating better M&E systems, making available user-friendly evidence, linking evidence from monitoring information and evaluation to feedback-loops in learning, and enhancing the diffusion of information, researchers and evaluators can make greater contributions to advancing the science of delivery and, ultimately, lead to well-informed, evidence-based decision-making.

REFERENCES

- Catford, John. 2013. "Advancing the 'Science of Delivery' of Health Promotion: Not Just the 'Science of Discovery'". *Health Promotion International*. 24 (1): 1-5, Oxford University Press, Oxford, England. Available at heapro.oxfordjournals.org/content/24/1/1.full. Last accessed 5 February 2014.
- Faulkner, William. 2012. "Negotiated Validity: Three Perspectives on IFPRI-Progresa's Sampling". *The Evaluator*. London.
- Independent Evaluation Group. 2011. "Project Performance Assessment Report – India: A Cluster Assessment of Forestry and Watershed Development Activities". World Bank Group.
- Kim, Jim Yong. 2012. "Delivering on Development: Harnessing Knowledge to Build Prosperity and End Poverty". Keynote Speech to World Knowledge Forum, Seoul, Korea, Republic of Korea. Available at worldbank.org/en/news/speech/2012/10/08/delivering-development-harnessing-knowledge-build-prosperity-end-poverty?goback=%252Egde_4387230_member_174342256. Last accessed 5 February 2014.
- Lustig, Nora Claudia. 2011. "Scholars who Became Practitioners: The Influence of Research on the Design, Evaluation, and Political Survival of Mexico's Antipoverty Program Progresa/Oportunidades". Center for Global Development Working Paper No. 263. Available at dx.doi.org/10.2139/ssrn.1910074. Last accessed 5 February 2014.
- Parker, Susan. 2003. "The Oportunidades Program in Mexico." Shanghai Poverty Conference.
- Rawlings, Laura. 2009. "Monitoring and Evaluation: the Foundations for Results". Beijing, China: Impact Evaluation Workshop. Slideshow available at siteresources.worldbank.org/INTISPMA/Resources/383704-1184250322738/3986044-1250881992889/M&E_TheFoundationsOfResults_Rawlings_EN.pdf. Last accessed 5 February 2014.
- UN News Centre. 2013. "Deputy UN Chief Calls for Urgent Action To Tackle Global Sanitation Crisis". Available at un.org/apps/news/story.asp?NewsID=44452#.UgD8GZLumJp. Last accessed 5 February 2014.
- United Nations. 2013. "The Millennium Development Goals Report". Available at unstats.un.org/unsd/mdg/Resources/Static/Products/Progress2013/English2013.pdf. Last accessed 5 February 2014.
- Woolcock, Michael. 2013. "The Science of Delivery – Taking Context Seriously: Expanding the Range of Ideas, Methods and Strategies to Enhance Development Effectiveness". Workshop presentation, World Bank Group. Available at siteresources.worldbank.org/INTRES/Resources/469232-1368745341168/TakingContextSeriously.pdf. Last accessed 5 February 2014.
- World Bank. 2008. "Second Minas Gerais Development Partnership: Project Appraisal Document". (Vol.1).
- World Bank. 2013a. "Minas Gerais Partnership II SWAP: Implementation Status and Results Report." Sequence 10.
- World Bank. 2013b. "Capital Spending on Roads in Edo State. Overview of NDSA Research findings". Washington, D.C.: Niger Delta Social Accountability Program, World Bank.
- World Bank. (Undated). "Social Observatory: Rural Livelihoods Projects in India." Available at worldbank.org/content/dam/Worldbank/document/SAR/SO.pdf. Last accessed 5 February 2014.

ANNEXES



ANNEX 1: AGENDA

Monday 30th September

OPENING OF THE CONFERENCE

9:00–10:00 **Paulo de Martino Jannuzzi**, *Secretariat for Evaluation and Information Management, Ministry of Social Development and Fight against Hunger*

Indran A. Naidoo, *Director, Independent Evaluation Office of UNDP*

OPENING ADDRESS

Rebeca Grynspan, *UN Under-Secretary-General and UNDP Associate Administrator*

PLENARY SESSION I

10:20–11:45 DIALOGUE ON THE IMPORTANCE OF EVALYEAR

Moderator: **Marco Segone**, *EvalPartners co-chair*

EVALYEAR – WHAT IS THE EXPECTED GOAL IN TERMS OF NATIONAL EVALUATION CAPACITY DEVELOPMENT BY THE YEAR 2015?

Presenter: **Asela Kalugampitiya** (*Sri Lanka*), *Secretariat of EvalYear on behalf of EvalPartners*

HOW CAN GOVERNMENTS, PARLIAMENTARIANS AND VOPES WORK TOGETHER TO STRENGTHEN THE DEMAND AND USE OF EVALUATION?

Commentators (government, MP and VOPE)

Hon. Kabir Hashim, *Member of Sri Lanka Parliament, Leader of South Asia Parliamentary Forum and member of EvalPartners International Advisory Group*

Velayuthan Sivagnanasothy, *Secretary, Ministry of Traditional Industries and Small Enterprise Development, Sri Lanka*

Mallika Samaranayake, *President of CoE/South Asia/Sri Lanka*

PLENARY SESSION II

11:45–12:30 INNOVATIVE SOLUTIONS TO CHALLENGES LINKED TO USE OF EVALUATION

Moderator: **Juha Uitto**, *Independent Evaluation Office of UNDP*

Keynote address: **Paulo Jannuzzi**, *Secretary for Evaluation and Information Management, Ministry of Social Development and Fight against Hunger*

PLENARY DIALOGUE - USE OF EVALUATION

16:50–18:00 WHAT ARE SOME OF THE INNOVATIVE SOLUTIONS TO DEAL WITH CHALLENGES LINKED TO USE OF EVALUATION?

Moderator: **Vijayalakshmi Vadivelu**, *Independent Evaluation Office of UNDP*

Discussants

Thania de la Garza Navarrete (Mexico), *General Director of Evaluation, Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL)*

Ivett Subero (Dominican Republic), *Director of UN System Monitoring, Multilateral Cooperation Office (DIGECOOP), Ministry of Economy*

Ted Sitima-wina (Malawi), *Permanent Secretary, Ministry of Economic Planning and Development*

Tuesday, 1st October

PLENARY SESSION III

9:30–10:30 INNOVATIVE SOLUTIONS TO CHALLENGES LINKED TO EVALUATION CREDIBILITY

Moderator: **Juha Uitto**, *Independent Evaluation Office of UNDP*

Keynote address: **Maria Bustelo**, *Complutense University of Madrid (UCM)*

PLENARY DIALOGUE - EVALUATION CREDIBILITY

14:00–15:30 HOW TO ENSURE/STRENGTHEN EVALUATION CREDIBILITY?

Moderator: **Vijayalakshmi Vadivelu**, *Independent Evaluation Office of UNDP*

Discussants

Martha McGuire (Canada), *CES Representative to and Treasurer of IOCE Board*

JP van der Merwe (South Africa), *Acting Deputy Director-General: Monitoring and Evaluation*

Clesencio Tizikara (Ghana), *Monitoring and Evaluation Specialist, Forum for Agricultural Research in Africa*

PLENARY SESSION IV

16:00–17:00 INNOVATIONS IN EVALUATION

Moderator: **Jennifer Colville**, *UNDP Bureau for Development Policy*

Presenters

Thomas Winderl, *Innovations in Evaluation*

Natalia Kosheleva, *EvalPartners co-chair and IOCE President, EvalPartners Innovation Challenge*

PLENARY SESSION V

17:00–18:00 HOW EVALUATIONS INFLUENCED BRAZILIAN PROGRAMMES

Moderator: **Paula Montagner**, *MDS/SAGI (Brazil)*

Discussants

Alexandro Rodrigues Pinto, *Acting Director of Evaluation / SAGI/MDS, Programa Bolsa Família (Family Grant Program)*

Cecília Ishikawa Lariú, *Impact Evaluation Coordinator / SAGI/MDS, Cistern Programme*

Luciana Sardinha, *Demand Assessment General Coordinator, Food Acquisition Program*

Wednesday, 2nd October

PLENARY SESSION VI

- 9:30–10:30 INNOVATIVE SOLUTIONS TO CHALLENGES LINKED TO INDEPENDENCE OF EVALUATIONS
 Moderator: **Ximena Fernandez Ordone**, *World Bank IEG*
 Keynote address: **Hans Martin Boehmer**, *Senior Manager, IEGCS/World Bank*

PLENARY DIALOGUE - INDEPENDENCE OF EVALUATION

- 14:00–15:30 HOW TO ENSURE EVALUATION INDEPENDENCE?
 Moderator/commentator: **Juha Uitto**, *Independent Evaluation Office of UNDP*
 Discussants
Hans Martin Boehmer, *Senior Manager, IEGCS/World Bank*
Paulo Jannuzzi, *Ministry of Social Development and Fight against Hunger*
Maria Bustelo, *Complutense University of Madrid (UCM)*

WAY FORWARD AND CLOSING REMARKS

- 16:00–17:30 WHAT ARE THE AGREED COMMITMENTS AND ACTIONS OF THIS NEC CONFERENCE TOWARDS THE PROPOSED EVALYEAR 2015?
 SUGGESTIONS OF THEMES FOR THE NEXT CONFERENCE
 VENUE NOMINATIONS FOR THE NEXT CONFERENCE
 Moderator: **Ana Soares**, *Independent Evaluation Office of UNDP*
 CLOSING REMARKS
Paulo Jannuzzi, *Secretary for Evaluation and Information Management, Ministry of Social Development and Fight Against Hunger*
Indran Naidoo, *Director, Independent Evaluation Office of UNDP*

ANNEX 2: PARTICIPANTS

PARTICIPANT	COUNTRY	INSTITUTION	TITLE
Rangina Kargar	Afghanistan	Government of Afghanistan	Member of Parliament
Hamed Sarwary		Monitoring and Evaluation, Independent Directorate for Local Governance	Director
Aulida Anamali	Albania	Shkodra Municipality	Foreign Relations Specialist
Federico Berardi	Argentina	Cabinet of Ministers, Presidency of the Nation	Undersecretary of Programme Evaluation with External Finances
Norberto Perotti			Undersecretary of Evaluation of National Budget
Pablo Rodriguez-Bilella		EvalPartners Knowledge Management Taskforce, International Development Evaluation Association Board Member	Appointed member of EvalPartners Management Group
Esteban Tapella		Latin American Network of Evaluation, Monitoring and Systematization in Latin America and the Caribbean (ReLAC)	ReLAC representative to the International Organization for Cooperation in Evaluation Board
Thomas Winderl	Austria	winderl.net	Planning, monitoring and evaluation consultant
Patrick Mccaskie	Barbados	Ministry of Finance and Economic Affairs	Director of Research and Planning
Aristide Djidjoho	Benin	Office of Public Policy Evaluation, Office of the Prime Minister	Coordinator
Jigmi Rinzin	Bhutan	Government of Bhutan	Member of the National Council
Sonam Tobgyal		Research and Evaluation Division of the Gross National Happiness Commission	Senior Research Officer
Juan Carlos Lavayen	Bolivia	Ministry of Development Planning	Chief, Transparency Unit
Taiana Fortunato Araújo	Brazil	National Institute of Metrology, Quality and Technology (INMETRO)/ Ministry of Development, Industry and Foreign Trade	Economist
Katia Cristina Bassichetto		Municipal Secretariat of Health	Coordinator of Epidemiology and Information
Rafael de Souza Camelo		Data Analysis System State Foundation(SEADE)	Technical Adviser, Executive Board
Marcelo Trindade Pitta			Professor
Ana Maria Alves Carneiro Silva		University of Campinas	Researcher
Marcia Paterno Joppert		Latin American Network of Evaluation, Monitoring and Systematization in Latin America and the Caribbean (ReLAC)	ReLAC representative to the International Development Evaluation Association Board
Elisabeth Castelo Branco		Banco do Nordeste do Brasil S.A. (BNB)	Coordinator of studies and research
Iraci Soares Ribeiro Maciel			Bank Analyst

PARTICIPANT	COUNTRY	INSTITUTION	TITLE
Welton Luiz Costa Rocha Filho	Brazil	Brazil Cooperation Agency, Ministry of Foreign Affairs	Chancery official
Netinho de Paula		Municipal Secretariat for the Promotion of Racial Equality (SMPIR)	Municipal Secretary
Olimpio Antonio Brasil Cruz		Ministry of Social Development and Fight Against Hunger(MSD), Department of Evaluation	Communications Adviser
Marco Natalino			Coordinator-General for the Evaluation of Impacts and Results
Bruno Cabral França			Office of the Secretary
Cecilia Ishikawa Lariú			Evaluation Department
Paulo Jannuzzi			Secretary of Evaluation and Information Management
Alexandro Rodrigues Pinto			Head, Evaluation Department,
Marcelo Cabral			Director, Management and Monitoring
Luciana Monteiro Sardinha			
Armando Simões			Office of the Secretary
Esther Leblanc			Secretariat of International Relations, São Paulo
Samir Funchal Oliveira		Brazil Cooperation Agency, Ministry of Foreign Affairs	Chancery Officer
Guillherme Pereira		Effectiveness Evaluation Unit, Planning Division (BNDES)	Economist
Marianne Pinotti		Secretariat of People with Disabilities and Reduced Mobility, São Paulo	Municipal Secretary
Selma Maria Hayakawa C. Serpa		Brazilian Court of Accounts	Federal Auditor, Evaluation and Performance Audit Department
Antonio Sergio Spagnuolo	World Centre for Sustainable Development (Rio+ Centre), UNDP Brazil	Communications Officer	
Claudia Valenzuela	Earnest Young	Director, Government and Public Sector	
Igr Saidi Kibeya	Burundi	Republic of Burundi Presidency	Coordinator of the Bureau of Strategic Studies and Development
Theng Pagnathun	Cambodia	General Directorate of Planning, Ministry of Planning	Director General
Marie Gervais	Canada	Quebec Evaluation Society Programme	Appointed member of EvalPartners Management Group
Johann Lucas Jacob		Centre for Research and Evaluation Expertise, National School of Public Administration	Coordinator
Martha Mcguire		Canadian Evaluation Society (CES)	Canadian Evaluation Society representative and Treasurer of International Development Evaluation Association Board
James Melanson		Strategic Policy and Performance Branch, Department of Foreign Affairs, Trade and Development	Director of Development Evaluation

PARTICIPANT	COUNTRY	INSTITUTION	TITLE
Anne Bramble	Caribbean Development Bank	Office of Independent Evaluation, Caribbean Development Bank	Head
Ignacio Irrarazaval	Chile	National Education Council	President
Li Yaohui Li	China	Department of Aid to foreign Countries, Ministry of Commerce	Director of the Association of Southeast Asian Nations (ASEAN)
Victor Manuel Quintero	Colombia	Latin American Network of Evaluation, Monitoring and Systematization in Latin America and the Caribbean (ReLAC)	Vice Coordinator
Florita Azofeifa	Costa Rica	Ministry of National Planning and Economical Politics (MIDEPLAN)	Director of Evaluation and Monitoring Department
Sabrina Storm		German Society for International Cooperation (GIZ), Promoting Evaluation Capacity in Central America Programme (FOCEVAL)	Programme Manager
Rosaura Trigueros Elizondo		Ministry of Public Finance Planning Coordination and Public Debt Control Department	Director
Marco Vinicio Espinal Martínez	Dominican Republic	Monitoring and Evaluation, National Competitiveness Council	Adviser
Rosa Mena		Ministry of Education	Director of Community Participation
Ivett Subero		Multilateral Cooperation Office (DIGECOOM), Ministry of Economy	Director of UN System Monitoring
Doha Abdelhamid	Egypt	The Middle East and North Africa Evaluators Network (EvalMENA)	EvalMENA representative to the International Development Evaluation Association Board
Mahmoud el Said		Centre for Project Evaluation and Macroeconomic Analysis	Ambassador
Samuel Hernandez	El Salvador	Department of Statistics and Censuses	Deputy Director
Susana Lobato		Government of El Salvador	Technical Secretariat of the Presidency
Neha Karkara	EvalPartners	Enabling Environment Taskforce	Advocacy Consultant
Pulkkinen Jyrki	Finland	Development Evaluation, Office of the Under-Secretary of State, Ministry for Foreign Affairs	Director
Riitta Oksanen		Development Evaluation, Office of the Under-Secretary of State, Ministry for Foreign Affairs	Senior Adviser
David Akana	Global Environment Facility	Evaluation Office	Extended Term Consultant
Baljit Wadhwa		Evaluation Office	Senior Evaluation Officer
Charles Amoatey	Ghana	Ghana Institute for Management and Public Administration	Consultant
Bruno Dery		National Development Planning Commission	Deputy Director
Clesensio Tizikara		Forum for Agricultural Research in Africa	Monitoring and Evaluation Specialist
Fidel Arévalo	Guatemala	Ministry of Planning and Programming	Adviser to the Minister

PARTICIPANT	COUNTRY	INSTITUTION	TITLE
Magdalena Ordóñez	Honduras	Executive Committee, Latin American Network of Evaluation, Monitoring and Systematization in Latin America and the Caribbean (ReLAC)	Chair
Julio Raudales		Planning and External Cooperation Coordination (SEPLAN)	Minister
Ratna Anjan Jena	India	Program Evaluation office, Planning Commission, Government of India	Adviser
Santosh Mehrotra		Institute of Applied Manpower, Research Planning Commission	Director General
Arif Haryana	Indonesia	Directorate of Evaluation and Reporting System Performance Development, National Development Planning Board/ BAPPENAS	Director
Barbara Rosenstein	Israel	Israeli Association for Program Evaluation	Chairperson
Shawn Grey	Jamaica	Performance Management and Evaluation	Director
Samson Masese Machuka	Kenya	Ministry of Planning and National Development	Director of the Monitoring and Evaluation Directorate
Jennifer Mutua		African Evaluation Association (AfrEA)	Representative to the International Development Evaluation Association Board
Kubat Jumaliev	Kyrgyzstan	Division on Public Administration and Personnel Work	Expert
Ziad Moussa	Lebanon	Middle East and North Africa Evaluators Network (EvalMENA)	Coordinator and appointed member of EvalPartners Management Group
Jimmy Kawaye	Malawi	Development Effectiveness and Accountability Programme, Ministry of Economic Planning and Development	Programme Coordinator
Winston Nyasulu		Monitoring and Evaluation Section of the Ministry of Finance	Assistant Budget Director
Shahrazat Binti Haji Ahmad	Malaysia	Implementation Coordination Unit, Evaluation Office, the Prime Minister's Office	Director, Outcome Evaluation Division
Mohd Khidir			Deputy Director, Evaluation Division
Azhan Bin Samat			Management Service Division
Cheikh Ould Abdallahi Ould Zeidane	Mauritania	Ministry of Economic Affairs and Development	Director of Strategies and Policies
Thania de la Garza Navarrete	Mexico	National Council for the Evaluation of Social Development Policy (CONEVAL)	General Director of Evaluation
Agustín Escobar			Evaluation Expert
Oyuntsetseg Khurts	Mongolia	Monitoring and Evaluation Department at Cabinet Secretariat	Head of Internal Audit
Baasanjav Radnaabazar		Monitoring, Evaluation and Audit Department, Ministry of Economic Development	Director

PARTICIPANT	COUNTRY	INSTITUTION	TITLE
Mohamed Mouime	Morocco	Information Systems Division, National Observatory of Human Development	Director
Albano Manjate	Mozambique	Monitoring and Evaluation National Directorate of the Ministry of Planning and Development	Deputy National Director
Teertha Raj Dhakal	Nepal	Secretariat Singha Durbar	Joint Secretary, National Planning Commission
Zakari Bonkano	Niger	Ministry of Planning, Spatial Planning and Community Development	Director of Evaluation
Zakari Lawal	Nigeria	, National Planning Commission, The Presidency, Monitoring and Evaluation Department	Director of Monitoring and Evaluation
Qaiser Jamal, Mna	Pakistan	Government of Pakistan	Member of Parliament,
Arlyn Gonzalez	Panama	National Council for the Evaluation of Social Development Policy	Director
Natalia Kosheleva	Russia	International Organization for Cooperation in Evaluation	President and Co-chair, EvalPartners
Lyudmila Vladiko		Government in Krasnoyarsk	
Sergey Vologodsky		Department of Strategic Management, State Programs and Investment Projects, Ministry of Economic Development	Deputy Head of Evaluation of Investment Projects Division
Jennifer Bisgard	South Africa	African Evaluation Association (AfrEA)	Representative to the International Organization for Cooperation in Evaluation Board
Christel Jacob		Department of Performance Monitoring and Evaluation, The Presidency	Director of Evaluation and Research unit
David Mkhwanazi		Performance Monitoring and Evaluation, the Presidency	Public Service Commissioner
Nompumelelo Madonda		Public Service Commission	Deputy Director
Lulama Sizani			Public Service Commissioner
Jakobus Vandermerwe			Acting Deputy Director General, Monitoring and Evaluation
Ahmed Vawda		Evaluation and Research, Department of Performance Monitoring and Evaluation, The Presidency	Outcome Facilitator/ Specialist
Maria Bustelo	Spain	Complutense University of Madrid (UCM)	Associate Professor of Political Science and Public Administration

PARTICIPANT	COUNTRY	INSTITUTION	TITLE
Hashim Kabir	Sri Lanka	Government of Sri Lanka	Member of Parliament
Asela Kalugampitiya		Sri Lanka Evaluation Association (SLEVA), South Asia's Community of Evaluators (CoE), IDEAS	Member and one of the leaders of the South Asia Parliamentarians Forum
Mallika Samaranyake		Community of Evaluators (CoE)	President
Velayuthan Sivagnanasothy		Ministry of Traditional Industries and Small Enterprise Development	Secretary
Beverley Best	St. Lucia	Organization of Eastern Caribbean States (OECS)	Head, Functional Cooperation & Programme Management Unit
Darrel Montrope		Office of the Prime Minister	Cabinet Secretary
Jacqueline Warso	Suriname	Ministry of Foreign Affairs Suriname	UN Desk Officer Division International Development Cooperation
Umed Davlatzod	Tajikistan	Ministry of Economic Development and Trade	Deputy Minister
Theofrid T.L. Kikombele	Tanzania	Finance and Economic Committee at the Cabinet Secretariat, President's Office, State House	Cabinet Assistant Secretary
Ada Ocampo	Thailand	United Nations Children's Fund (UNICEF)	Evaluation Adviser
Dinesh Ragoo	Trinidad & Tobago	National Transformation Unit, Ministry of Planning and Sustainable Development	Programme Director
Albert K. Byamugisha	Uganda	Office of the Prime Minister	Commissioner, Monitoring and Evaluation
Rosetti Nabbumba Nayenga		Government of Uganda	Minister of Finance, Planning and Economic Development and member of the Uganda Evaluation Association
Tessie Tzavaras Catsambas	United States of America	American Evaluation Association (AEA)	Representative to and Secretary of International Organization for Cooperation in Evaluation Board
Jody Fitzpatrick			President
Beverly Parsons			President Elect and Executive Director at InSites
Jim Rugh		EvalPartners	Coordinator
Stephanie Shipman		Centre for Evaluation Methods and Issues, Government Accountability Office	Assistant Director
Leopoldo Font	Uruguay		
Jerónimo Roca		Office of Planning and Budget, President of the Republic	Sub-director of Monitoring and Evaluation
Katongo S.M Chifwepa	Zambia	Policy Implementation, Monitoring and Evaluation, Policy Analysis and Coordination, Cabinet Office	Director

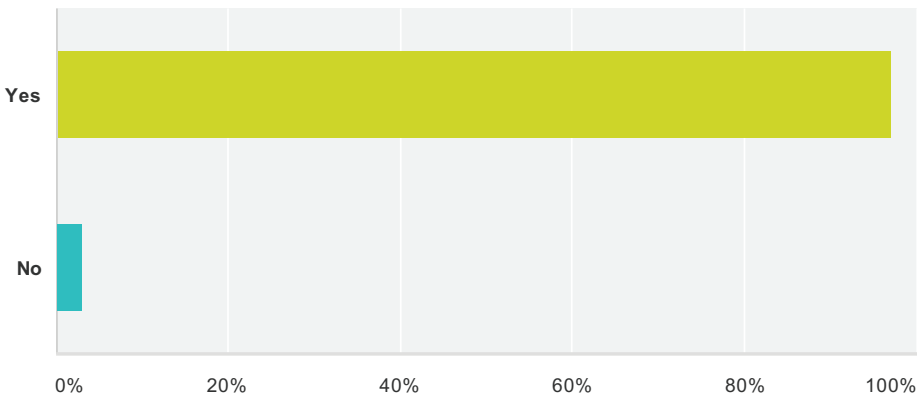
PARTICIPANT	INSTITUTION	OFFICE/DEPARTMENT	TITLE
Jorge Chediek	UNDP	UNDP Brazil	Resident Representative
Maristela Marques Baioni			Assistant Resident Representative
Juliana Wenceslau			Monitoring and Evaluation Officer
Rebeca Grynspan		Executive Office	Under Secretary General and Associate Administrator
Pedro Manuel Moreno			Special Assistant to the Associate Administrator and Policy Specialist
Indran Naidoo		Independent Evaluation Office	Director
Juha Uitto			Deputy Director
Ana Rosa Soares			Evaluation Specialist
Concepcion Cole			Programme Associate
Ximena Rios			Operations Manager
Vijaya Vadivelu			Evaluation Specialist
Xiaojun Grace Wang			Knowledge, Innovation and Capacity Group, Bureau of Development Policy
Romolo Tassone		Consultant	
Diana Oya Sawyer		International Poverty Centre (IPC)	Senior Researcher/ Coordinator
Mariana Hoffman			Communications Officer
Francois-Corneille Kedowide		Regional Service Centre, Addis Ababa	Monitoring and Evaluation Adviser
Inka Matilla		Regional Service Centre, Panama	Monitoring and Evaluation Specialist
Romulo Paes de Sousa	World Centre for Sustainable Development (Rio+ Centre)/UNDP	Coordinator	
Martin Barugahare	UN-HABITAT	Office of the Executive Director	Chief, Monitoring and Evaluation Unit
Marco Segone	UNICEF	Evaluation Office	Senior Evaluation Specialist, Systemic Management; Co-Chair of EvalPartners
Dieudonne Mouafo	United Nations Volunteers	Evaluation Unit	Chief
Laura Gonzalez	UN Women		Regional Evaluation Specialist
Inga Sniukaite		Evaluation Office	Deputy Chief
Hans-Martin Boehmer	World Bank	Independent Evaluation Group	Senior Manager
Ximena Fernandez Ordone			Evaluation Officer and member of Regional Centres for Learning on Evaluation and Results (CLEAR) Secretariat
Cristina Galindez		Independent Evaluation Group, World Bank and Regional Centres for Learning on Evaluation and Results (CLEAR)	Administrative Coordinator
Nidhi Khattri			Lead Evaluation Officer and Head of the Regional Centres for Learning on Evaluation and Results (CLEAR)
Gabriela Pérez Yarahuan			Visiting Professor

ANNEX 3: CONFERENCE ASSESSMENT

Nearly two-thirds of the participants found the conference structure and delivery useful in achieving the objectives of the National Evaluation Capacities conference. The majority of the participants also found the delivery method of the conference (solution exchange forums, plenary presentations and plenary discussions) very useful. A large number of participants were very satisfied with the conference’s organization, design, and implementation. Additional information on the conference assessment is available at nec2013.org/downloads/NEC-Survey-2013.pdf.

Q1 Did you find the engagement useful?

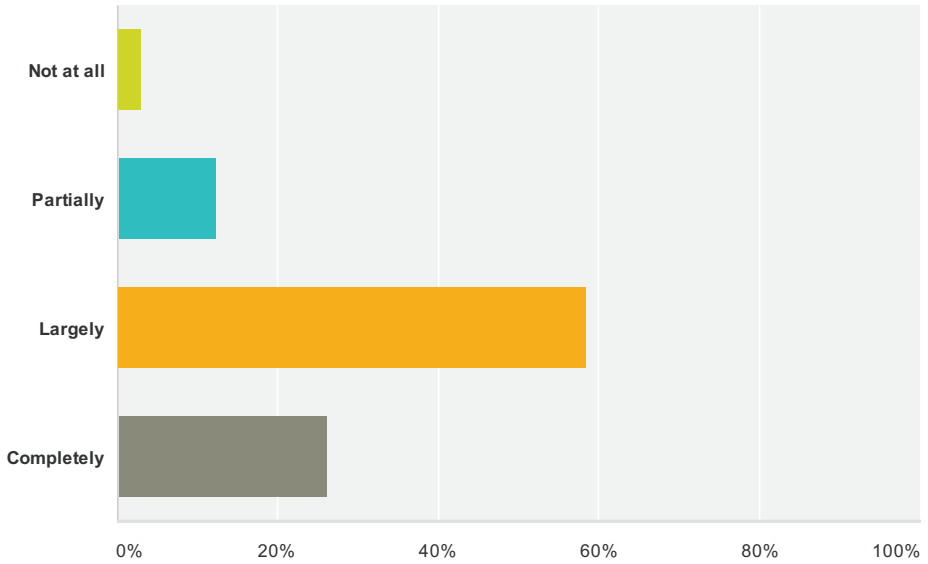
Answered: 65 Skipped: 0



Answer Choices	Responses
Yes	96.92% 63
No	3.08% 2
Total	65

Q2 Was the way the conference was structured and delivered useful in achieving the objectives of National Evaluation Capacities?

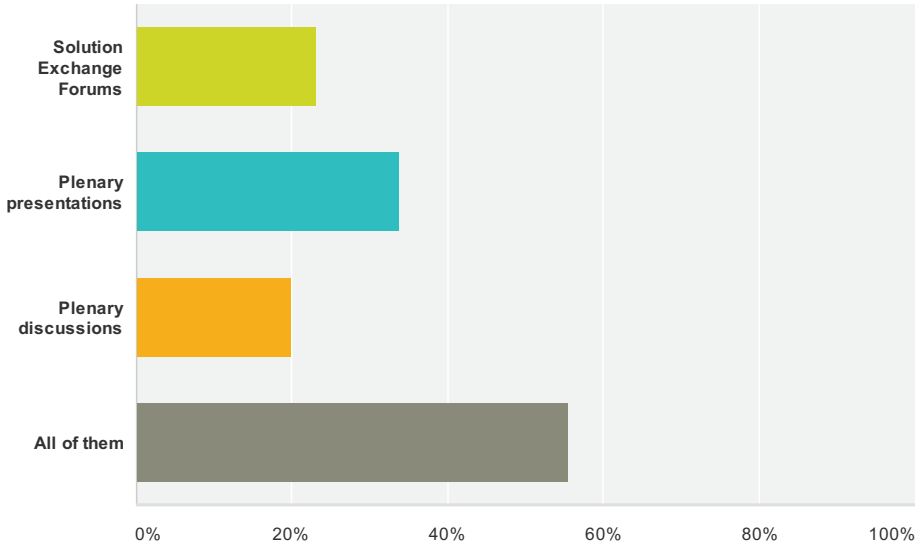
Answered: 65 Skipped: 0



Answer Choices	Responses
Not at all	3.08% 2
Partially	12.31% 8
Largely	58.46% 38
Completely	26.15% 17
Total	65

Q3 The most valuable presentation delivery method was/were (multiple answers possible)

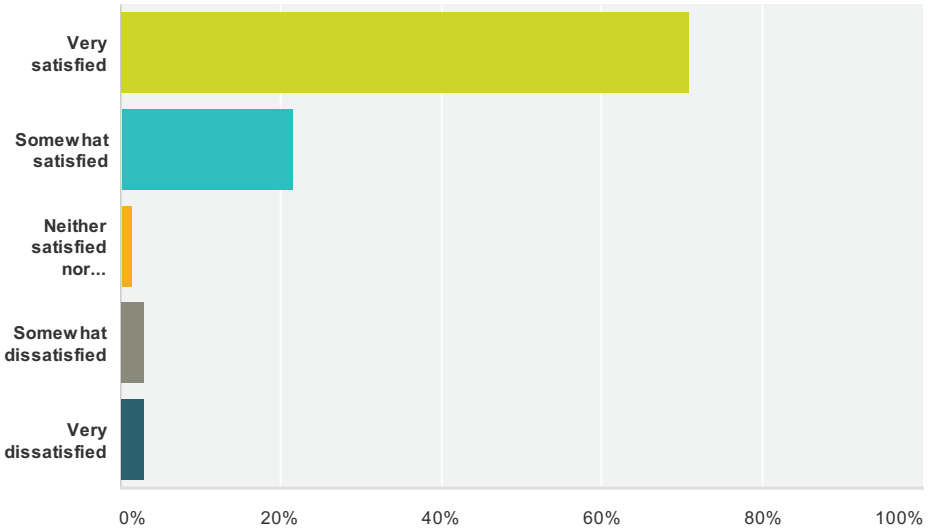
Answered: 65 Skipped: 0



Answer Choices	Responses	Count
Solution Exchange Forums	23.08%	15
Plenary presentations	33.85%	22
Plenary discussions	20%	13
All of them	55.38%	36
Total Respondents: 65		

Q4 Do you think the conference could be improved in terms of organization, design, and implementation? Please rate your overall satisfaction with the conference.

Answered: 65 Skipped: 0



Answer Choices	Responses
Very satisfied	70.77% 46
Somewhat satisfied	21.54% 14
Neither satisfied nor dissatisfied	1.54% 1
Somewhat dissatisfied	3.08% 2
Very dissatisfied	3.08% 2
Total	65

**ATTACHMENT:
THE EVENT IN PHOTOS**











- SANTASH MENDOTA - INDIA
 jacqueline walton
 SERRANIAN
1. V. SIVAGNANASOOTHY - Gov. of Sri Lanka
 2. ENTENRE PERABUWA - SRI LANKA
 3. Malika R. Samaranyake - Sri Lanka - CECS SLEVEN
 4. Teothe Raj Dhakal - NEPAL
 5. BEVERLY BEST - CECS
 6. KIBEYA SAIDI - BURUNDI
 7. ALBERT BYAMUGI SHA - UGANDA
 8. SAMSON MACHUKA - KENYA
 9. MARCO ESPINAL DOMINICAN REPUBLIC
 10. ROSETTI NABBUMBA - UGANDA
 11. WINSTON NYASULU - MALAWI
 12. ARIF HARYANA - INDONESIA









*Empowered lives.
Resilient nations.*



ISBN 978-92-1-126379-4

Independent Evaluation Office
United Nations Development Programme
220 East 42nd Street, New York, NY 10017, USA
Tel. (646) 781 4200, Fax (646) 781 4213
Internet: www.undp.org/evaluation

