WHY DID THE CHICKEN CROSS THE ROAD?

And other stories on development evaluation...

WHY DID THE CHICKEN CROSS THE ROAD?

And other stories on development evaluation...

The IOB Policy and Operations Evaluation Department is an independent arm of the Ministry of Foreign Affairs, the Netherlands. Its objective is to meet the need for independent evaluation in all fields of foreign policy. Its reports are public and are used as feedback to improve policies and their implementation.
Address: IOB, PO Box 20061,
2500 EB The Hague, The Netherlands.

Europe's Forum on International Cooperation (Euforic) is a forum for information exchange and dialogue on international cooperation and development. Drawing strength and inspiration from its members and their partners, its mission is to contribute to more informed, equitable and effective modes of international cooperation and development. Euforic makes information on Europe's international cooperation more accessible, it facilitates communication and dialogue among organisations involved in international cooperation, and it brokers collaboration, learning and knowledge exchange on international cooperation issues.
Address: Euforic, Wycker Grachtstraat 38,
6221 CX Maastricht, The Netherlands.

This book has also been published as number 363 in the KIT Bulletin series. The Bulletin Series of the Royal Tropical Institute (KIT) deals with current themes in international development cooperation. KIT Bulletins offer a multi-disciplinary forum for scientists, policy makers, managers and development advisors in agriculture, natural resource management, health, culture, history and anthropology.
These fields reflect the broad scope of the Royal Tropical Institute's activities.

Royal Tropical Institute (KIT)
KIT Publishers
P.O. Box 95001
1090 HA Amsterdam, The Netherlands
Telephone: +31 (0)20 5688 272
Telefax: +31 (0)20 5688 286
E-mail: publishers@kit.nl
Website: www.kit.nl

Cover: Grafisch Ontwerpbureau Agaatsz
BNO, Meppel, The Netherlands
Graphic design: Nadamo Bos, KIT Publishers
Printing: Meester & de Jonge, Lochem,
The Netherlands

ISBN: 90 6832 8492
ISSN: 0922-7911
NUR: 600/860

Preface

2005 was an important year in which the impact of the Millennium Development Goals (MDGs) to date was assessed. The MDGs were adopted by the worldwide community in 2000 to address poverty reduction, access to health, education, drinking water and proper sanitation, gender equality and other major issues. In September 2005, the leaders of over 170 nations met to review the results so far and to stimulate further progress. Quality assurance is a vital tool to ensure an effective implementation of the MDGs. Quality assurance can take many forms: evaluation is part of many quality assurance processes.

This book contains 11 articles which have appeared on the Euforic website since 2001 as individual issues of the Aid Evaluation Dossier. The articles have now been compiled into this publication because, although publication on the Internet has advantages in terms of ease of access and timeliness, the continuity offered by a printed book cannot be replaced by an electronic medium.

Since 2001, both the discourse and the practice of evaluation have evolved in response to changes taking place in development as a whole. The appearance of a number of global initiatives, particularly the focus on the MDGs, and an increased emphasis on poverty reduction, are some of the challenges to which evaluation practices are responding. Increasing demands for institutional accountability and the need for learning at all levels of development, at the same time as a growing emphasis on multilateral as opposed to bilateral development aid, are calling for new approaches to evaluation.
The 11 articles in this publication have been written by leading experts and thinkers in this field, and provide varying perspectives on this rapidly changing context. In an often innovative way, they identify and discuss the new trends and challenges facing evaluators and their institutions.

This publication and the related Aid Evaluation Dossier are a tribute to the vision of the IOB Policy and Operations Evaluation Department of the Ministry of Foreign Affairs, the Netherlands, in its efforts to stimulate the continuing international debate on evaluation.

Dr Jan Donner
President
Royal Tropical Institute

Acknowledgements

This publication marks the existence of the Aid Evaluation Dossier on the Euforic website. It would not have been possible without the contributions from the 14 authors who have been involved in writing these 11 articles. I would like to thank them all for their contributions but also for their cooperation through out the process of producing this publication. The authors are experts in the challenging field of evaluation, and it has been a pleasure to be able to include their excellent contributions here.

Working on the Aid Evaluation Dossier and this publication has involved collaboration between the IOB Policy and Operations Department of the Ministry of Foreign Affairs of the Netherlands, Euforic and the Royal Tropical Institute (KIT), with special thanks going to Henri Jorritsma and Han van Bommel of IOB. Finally, I would like to give particular mention to Huub Mudde, who played an important part in bringing IOB, Euforic and KIT together to work on this project, and to Jacques van Laar of Euforic.

Kindly note the disclaimer that the articles are the responsibility of the authors and do not represent the opinions of the publishing organizations.

Sarah Cummings
Editor, *Why did the chicken cross the road? And other stories on development evaluation…*

Table of contents

Acronyms

ACP	African, Caribbean and Pacific country signatories of the Cotonou Agreement with the EU
ALNAP	Active Learning Network for Accountability and Performance
CAPE	Centre for Aid and Public Expenditure, UK
DAC	Development Assistance Committee, OECD
DGCD	Foreign Trade and Development Cooperation, Belgium
DFID	Department for International Development, UK
DRDPs	District Rural Development Programmes, Tanzania
ECDPM	European Centre for Development Policy Management, The Netherlands
ECG	Evaluation Cooperation Group (ECG) of the International Financial Institutions
EFELA	Evaluation Feedback for Effective Learning and Accountability workshop
EU	European Union
EUFORIC	Europe's Forum on International Cooperation
GEF	Global Environment Facility
IAWG	Inter Agency Working Group on Evaluation, UN
ICTs	information and communication technologies
IDEAS	International Development Evaluation Association
IDRC	International Development Research Centre, Canada
IFAD	International Fund for Agricultural Development
IMF	International Monetary Fund
INTRAC	International NGO Training and Research Centre, UK
IOB	IOB Policy and Operations Evaluation Department, Ministry of Foreign Affairs, The Netherlands
IT	information technology
KIT	Royal Tropical Institute, The Netherlands
MAPP	Method for Impact Assessment of Poverty Alleviation Projects
M&E	monitoring and evaluation
MDGs	Millennium Development Goals
MSC	Most Significant Change
NGO	non-governmental organization
ODI	Overseas Development Institute, UK
OECD	Organization for Economic Cooperation and Development
OED	Operations Evaluation Department, World Bank
PRA	participatory rural appraisal
PRODERE	*Programa de Desarrollo para Desplazados, Refugiados y Repatriados*, Central America
PRSPs	Poverty Reduction Strategy Papers
SIDA	Swedish International Development Cooperation Agency
SWAP	sector-wide approach
UN	United Nations
UNDP	United Nations Development Programme
USAID	Agency for International Development, USA
WTO	World Trade Organization

Sarah Cummings

An introduction to 'Why did the chicken cross the road?'

The 11 articles in this publication have been written by leading experts and thinkers in the field of evaluation, and provide varying perspectives on this rapidly changing context. In an often innovative way, they identify and discuss the new trends and challenges facing evaluators and their institutions. The 11 articles fall generally into the following themes:
- new approaches to evaluation: a focus on poverty and results;
- collaboration and partnership;
- learning from evaluation;
- indicators and the logical framework; and
- participatory approaches.

This introduction will first describe the background to this publication, including why it has this title and providing a thumbnail sketch of the authors. Next, the 11 articles will be briefly introduced, divided into these five subject areas.

Background to this publication

On 6 November 2001, a Dossier on Aid Evaluation on the Internet was launched as part of a cooperative agreement between the IOB Policy and Operations Evaluation Department at the Netherlands Ministry of Foreign Affairs, Europe's Forum on International Cooperation (Euforic) and the Information and Library Services of the Royal Tropical Institute (KIT). The Dossier comprises a series of regularly updated articles, plus electronic links to online documents and other resources. The aim of the Dossier has been to improve the quality of international development cooperation by promoting knowledge sharing among individuals and agencies involved in evaluation and monitoring of international development. It is part of a larger cooperative agreement between Euforic and IOB to create a gateway on evaluation comprising discovery and dialogue tools that promote access to information resources and, at the same time, foster networking and knowledge sharing among individuals. The Dossier can be found on the Internet at:
http://www.euforic.org/detail_page.phtml?page=resource_briefing_evaluat

This Dossier is part of a new generation of Internet-based communication approaches that the Organization for Economic Cooperation (OECD) and others have lauded as valuable:

> *The new generation of Internet-based communication approaches, including websites and internal intranets, are widely seen as a valuable addition to the evaluation toolkit... In addition, the Internet has proved to be effective not only for disclosure but also as a database. It can also be a useful consultation tool. In the recent World Bank review of the forestry sector, for example, the draft placed on the website attracted comments from over 200 people.* (OECD 2001)

These previously electronic-only documents are now being made available in print form because, although the Internet has advantages in terms of broadcast and dissemination, it is a transient medium. This publication will better preserve these articles for posterity.

Why *did* the chicken cross the road?

The title of the publication is derived from the second article 'Why did the chicken cross the road? Four principles to support collaboration within evaluation' written by Rob van den Berg. He argues:

> *The debate on collaboration within evaluation is reminiscent of the question: 'why did the chicken cross the road?' The simple answer is that it wanted to get to the other side. But the evaluator wants to know:*
> - *whether the chicken took the shortest path across the road, and whether it crossed the road in the quickest possible way;*
> - *whether it can be established whether the chicken indeed and in actual fact reached the other side, and remains there, and is expected to remain there in the future; and*
> - *whether the needs of the chicken have been met by crossing the road.*

The title of this article has been taken as the over-arching title for the whole publication because here we are indeed concerned with the shortest path towards development, whether and how development is taking place, and if the needs of development are being met by the current and future evaluation practices.

The authors

The authors of these articles represent leading thinkers and experts in the field of evaluation of development. They include a past Director of the Operations Evaluation Department (OED) of the World Bank, as well as the former Chair of the Network on Development Evaluation of the Development Assistance Committee (DAC), OECD. In addition to this, authors include eminent researchers as well as individuals holding prominent positions within bilateral and multilateral organizations, and research institutes. For more details of their careers and contact details, kindly consult the 'About the authors' section at the back of the publication.

Introducing the articles

New approaches to evaluation: a focus on poverty and results

Both Henri Jorritsma in his article 'How to measure the impact of poverty evaluation: in search of new methods in evaluation' and Jurrien Toonen and Rob van Poelje in 'Monitoring and evaluation at the crossroads: meeting the challenge of sector-wide approaches and budget support' are looking at the challenges facing evaluation, given the changing development environment and, particularly, an increased focus on poverty.

Henri Jorritsma, based on empirical evidence, indicates that donors often have a weak poverty focus, despite the rhetoric of the MDGs. Against this background, individuals and institutions concerned with evaluation are trying to develop a new focus on impact instead of performance evaluation of projects and programmes. Poverty impact studies by both multilateral and bilateral organizations are facing the same methodological challenges: attempts to gather the perceptions of the target group are impressionistic and provide few concrete data. Perception studies exist but are unrelated to project interventions. However, there are continuing efforts to address this shortfall.

Jurrien Toonen and Rob van Poelje take a different perspective of the same general problematique. Likewise, they argue that approaches to monitoring and evaluation are changing in response to changes taking place in development as a whole. The introduction of sector-wide approaches and budget support are two recent developments that call for innovative monitoring and evaluation systems. In addition, the appearance of a number of global initiatives, here again they mention the MDGs and the Global Funds, together with a renewed focus on poverty reduction, are also playing their part in the development of new evaluation approaches. At the same time, the trend away from bilateral towards multilateral aid mechanisms and partnerships will influence the practice of evaluation in the coming years. These changes represent a considerable challenge to monitoring and evaluation.

Robert Picciotto in 'The Monterrey challenge: re-shaping the development architecture' argues that the elements of a new global partnership were forged at the UN Conference on 'Financing for Development', held in Monterrey, Mexico, during March 2002. The new global partnership calls for the adoption of improved policies and good governance in developing countries with the provision of increased aid and trade opportunities by developed countries. The Monterrey challenge combines ambitious objectives, a focus on results, and an unprecedented partnership between developed and developing countries in pursuit of poverty reduction. The Monterrey Conference also addressed the need to achieve greater coherence in development efforts through the MDGs.

In 'The road to nowhere? Results-based management in international cooperation', Howard White argues that results-based management has become a fact of life for development agencies. He considers that development agencies which are following this approach might be advised to learn from the experience of the US Agency for International Development (USAID) which has already applied results-based management and discovered its limitations. Development agencies have been rightly criticized in the past for paying too little attention to the final impact of their activities so, from this perspective, a results-based approach does appear to offer a means of addressing this criticism. However, results-based management should not be applied blindly when it suffers from severe limitations. Instead, White considers that attempts to link agency performance to development outcomes should rely upon the logical framework or logframe. Although the logframe is not a universal panacea, he argues that, used properly, it can force agencies into a critical examination of their programmes and projects, and the results they achieve.

Collaboration and partnership

The articles by Rob van den Berg 'Why did the chicken cross the road? Four principles to support collaboration within development' and John Roberts 'Managing aid for results: how to select your partner in development?' address the allied subjects of collaboration and partnership.

Rob van den Berg analyzes the challenges facing bilateral organizations and other partners in evaluation, identifying four new principles that could be adopted to support collaborative approaches:

- joint finance should lead to joint evaluation;
- partnership in development should lead to partnership in evaluation;
- ownership of development should lead to ownership of evaluation; and
- if aid becomes integrated into development, we should evaluate development as a whole, not isolated programmes and projects.

He argues that a key issue facing the DAC Network on Development Evaluation is how to meet, exchange information and collaborate with the following constituencies:

- the evaluation community in partner countries;
- the evaluation community in non-governmental organizations (NGOs) and civil society, not only in partner countries but also in member countries of the OECD; and
- any other evaluation communities, like the private sector, human rights organizations, humanitarian assistance institutions, networks or associations of professionals, and so on.

John Roberts deals with the unresolved debate in the donor community on how to select countries for the allocation of aid, particularly how to exclude poor performers. He argues that the imprecision of the notion of 'poor performance', and the variety of the depth and breadth of its manifestations, cast serious doubt on 'sheep-and-goats' selectivity as a sound basis for designing development cooperation programmes. Past performance is no reliable guide to the future. Donors' aid strategies, it now appears, need to be more subtly tailored to local needs, policies, economic and political circumstances, and administrative absorptive capacity, and to be based on perceptive comparative readings of each of these dimensions.

Learning from evaluation

Two articles, 'When will we ever learn? Perspectives on learning from evaluation' by Paul Engel, Charlotte Ørnemark and Arin van Zee and 'The new frontier: evaluation and organizational learning' by Geoff Barnard, are both addressing increasing recognition of the fact that evaluations should be fed into learning, and are not just about accountability.

Geoff Barnard considers how development organizations can get the most out of evaluations, both from an accountability perspective, and from the point of view of institutional learning. This tension between learning and accountability was the focus of a report from the DAC Network on Development Evaluation, then known as the Evaluation Working Group, entitled 'Evaluation feedback for effective learning and accountability' (2001). The report emerged from a workshop organized by the Japanese government in Tokyo on 26-28 September 2000, and highlights the concerns widespread within DAC agencies about the shortcomings of current feedback practices and the need to do much more to ensure that lessons from evaluations are not wasted.

The report identified a number of obstacles that get in the way of learning:
- organizational culture;
- pressure to meet disbursement targets;
- lack of incentives to learn;
- tunnel vision;
- loss of institutional memory;
- insecurity and the pace of change; and
- the unequal nature of the aid relationship.

Also influenced by this workshop report, Paul Engel and colleagues argue that, in a very practical sense, evaluations are now perceived as learning opportunities. It only seems logical, therefore, to try and improve the internalization of evaluation results at different levels. But how can this be done? What dilemmas do we face and what have we learned so far? Improving the internalization of evaluation results implies a shift in perspective with respect to the way in which development processes, and indeed development institutions, are managed. Adaptive management, namely flexibility and responsiveness to experience, must become a key institutional feature. As a consequence, institutional development is a necessary complement to learning by development agencies. Far more emphasis needs to be placed on developing effective conceptual and practical tools.

In an effort to apply some of the lessons on the need for greater learning from evaluation, the European Centre for Development Policy Management (ECDPM), together with the International Development Research Centre (IDRC), Exchange, Bellanet and the East Africa Regional Office of the United Nations

Children's Fund (UNICEF), started the Pelican Initiative, the Platform for Evidence-based Learning & Communications for Social Change, in January 2005.

Indicators and the logical framework

Chris Whitehouse in 'The ants and the cockroach: a challenge to the use of indicators' and Thomas Winderl in 'A pot of chicken soup and why Brits always carry umbrellas: in defence of indicators' provide a comment and reply on the weaknesses and strengths of indicators. Chris Whitehouse provides a critique of the use of indicators, based on the example of a nest of ants at work:

> *One day, the queen ant, after listening to an inspiring documentary on monitoring and evaluation on the radio, decides to improve efficiency. She announces that, henceforth, it is not enough for each ant to just 'chip in' to the tasks at hand. Next time there's a cockroach to be moved, she is going to measure each ant's contribution, and their rights to gnaw at the flesh of the dead cockroach will be awarded in proportion to their contribution to the effort.*

Based on the example of the ants and of, among other things, hydrologists taking Thai cookery courses, he criticizes the validity and efficacy of what he calls the 'measuring stick approach' to development enshrined in indicators. The key problems with indicators fall into three main categories: the time and resources spent on identifying and measuring indicators; the skewing effects that inclusion of indicators may have on programme design and implementation; and, perhaps most seriously of all, the logical fallacy to which many development workers fall victim: confusing concomitance with causality.

Thomas Winderl, in his reply to Chris Whitehouse, argues that indicators and a results-orientation are part of contemporary development speech which, having emerged in the 1980s, are now here to stay. Although it is possible to criticize spending time and money on indicators which is out of proportion to the project itself, the development reality is that there is not enough time and funds invested in designing and monitoring clear and relevant indicators. Decades after the invention of logical frameworks and indicators, the inadequate and skewed logical frameworks and bad or plainly wrong indicators still dominate the development business. Both Thomas Winderl and Chris Whitehouse agree, however, that bad indicators are worse then no indicators at all.

Participatory approaches

In the final article in this publication Sophie da Câmara Santa Clara Gomes considers 'Participatory approaches to monitoring and evaluation'. She argues that once the principle of participation in evaluation is agreed, there remain two key challenges in terms of operationalization. The first is how to actually include the recipient. At what level can the beneficiaries-recipients be involved in the actual monitoring and evaluation exercise? And through which mechanisms? Second, how can we measure the social changes induced by the development intervention? How do we measure intangibles?

Some concluding comments

These articles provide a diverse variety of perspectives on evaluation, linking new perspectives, such as learning with key events, such as the Monterrey Conference in 2002, and key approaches such as the MDGs. They cover new approaches to evaluation, particularly the increasing attempts, not always successfully, to focus on poverty and results. The growing importance of collaboration and partnership between donor partners and between donors and recipient countries is also emphasized. Indicators and the logical framework approach stand fairly central to the process of evaluation.

Although these articles were written independently of each other, it does appear to be possible to reach some general conclusions. Firstly, the MDGs seem to be forcing a discipline on evaluation, mediating an approach that is focused on poverty and results. Although these new foci bring their own problems, this general shift must be, on the whole, a positive trend. There is now a general recognition that it is important for organizations to learn from evaluation, not just to improve the implementation of development initiatives but also to improve the evaluation practices themselves. Indicators and the logical framework, properly applied, seem to show promise of being able to limit some of the shortcomings of evaluation. As Howard White comments:

> *The logframe is not a universal panacea but, used properly, can force agencies into a critical examination of the nature of their programmes and projects, and the results they achieve.*

References

OECD (2001) *Evaluation Feedback for Effective Learning and Accountability*. Paris: OECD

Henri Jorritsma

1 How to measure the impact of poverty alleviation: in search of new methods in evaluation

> *Today everything is costing. Parents have to pay levy and school fees and if you cannot pay school fees you sell a goat. But in this land we depend on crops and if the crops fail we sell a goat...We have been building schools for nothing. Children are standing outside, teachers are not teaching, they only look after their salaries. People's education has gone down. What is development?*
> An anonymous informant from Mbulu District, Tanzania.

The Millennium Development Goals (MDGs) call for reducing the proportion of people living on less than $1 a day to half the 1990 level by 2015. The Goals also call for halving the proportion of people who suffer from hunger between 1990 and 2015. Achieving these Goals by 2015 will require more focus on development impact in order to effectively measure national progress. The Goals establish yardsticks for measuring results, not just for developing countries but for the rich countries that help to fund development programmes and for the multilateral institutions that help countries implement them. It must not be forgotten, however, that poverty reduction or development is not the sum total of the individual Goals. There is a general understanding that poverty is a complex multidimensional concept and that poverty reduction strategies should be based on a thorough analysis of its root causes. In other words, progress on such Goals as basic education or primary health care can, of course, be evaluated, but that does not automatically give much insight into the improvements of the poverty situation as such.

A study on decentralization, commissioned by the Organization for Economic Cooperation and Development/Development Assistance Committee (OECD/DAC) Network on Aid Evaluation (OECD 2004), pointed towards the weak poverty focus of many donor programmes, despite the rhetoric of the MDGs. It found little evidence that donor interventions contributed significantly towards poverty reduction, and identified lack of commitment on the part of governments of recipient countries. Against this background, individuals and institutions concerned with evaluation are trying to develop a new focus on impact instead of performance evaluation of projects and programmes.

Poverty and evaluation

The IOB Policy and Operations Evaluation Department is a unit in the Dutch Ministry of Foreign Affairs, established in 1977. It is independent in terms of programming, terms of reference, evaluation designs and methods, and reports directly to the Dutch Parliament. IOB's overall objective is to meet the need for independent evaluation in all fields of foreign policy. Specifically, the aim is to fit the results of the evaluations into the knowledge cycle of the Ministry of Foreign Affairs. The reports of the evaluations are used as feedback to improve both policies and their implementation. With the results of previous exercises at their disposal, policymakers can prepare new interventions more purposefully and more effectively.

Since the early 1970s, the reduction of poverty has been the major objective of Dutch development cooperation policy. Improving the living conditions of poor groups has chiefly been addressed through projects, either through improvements in community services such as education, health care and drinking water supply, or through support for activities that generate income. In evaluating these projects, the emphasis was usually placed on the degree to which the project goals had been achieved and the efficiency of implementation. Generally, evaluations started with an analysis of government policy and goals of projects and programmes, rather than with an analysis of the poverty problem in the field. Evaluations also paid far less attention to processes that led to this situation and the underlying causes of poverty; the perception of the population regarding the poverty situation, its underlying causes and the relevance of aid activities; and the extent to which aid activities aimed at removing the causes of poverty as perceived by the population.

In addition, programme development goals are generally defined in sectoral or thematic terms. As a consequence, evaluations apply logical frameworks, also known as logframes, which measure outputs and outcomes in sectoral or thematic terms. Observations on poverty impact seldom go beyond the level of impressionistic description. In most evaluations, target group perceptions on poverty are used in an illustrative manner. Field observations focus on project or programme execution, measuring efficiency and effectiveness. Due to time constraints, discussions with target groups remain largely anecdotal.

Given that reduction of poverty is the principal objective of Dutch development cooperation, the degree to which development activities contribute towards that objective is the most crucial question facing evaluations. IOB is thus currently faced with three methodological challenges. Firstly, how to go beyond the outcome level in evaluation and measure the impact on poverty level? Outcome level in IOB evaluations relates to project/programme goals, while impact assessment refers to overall development goals and, ultimately, socio-economic development. The second major methodological dilemma is how to find out what is going on at the grassroots. It is perfectly possible to analyze the quality of national policies and find out to what extent they are pro-poor, so to speak. It is also quite feasible to analyze the donor interventions at that level. However, that does not say much about what is going on at the grassroots level. Changes in the poverty situation of local poor people are not easily attributed to policy changes at the macro-level. There is a whole series of known or unknown interfering factors that have to be identified (and isolated) before it is possible to formulate plausible linkages between the micro- and macro-level. The third challenge is how to systematically integrate target group perceptions in evaluation.

IOB experience so far

Recent evaluation of two District Rural Development Programmes (DRDPs) in Mbulu and Songea Districts of Tanzania took target group perceptions of poverty as an entry point (IOB 2004). New methods and techniques were adopted in confronting perceptions, trends and policy in the two districts. The approach chosen focused on the way in which the people perceive their own poverty situation (for which the activities are intended), as well as long-term social and economic trends. In Mbulu, a video registration was made of the people's perceptions, but due to the costs involved this was not feasible in Songea. In both districts, the people's views were recorded by a research team from the University of Dar es Salaam. The evaluation concluded that the two DRDPs have not yet achieved their main long-term objective, structural poverty alleviation, nor their short-term objective of strengthening local governance. It argued that if the DRDPs continue in the present way, their contribution to poverty reduction will be negligible, also in future. Despite the development of new methods and techniques for this evaluation, problems remained: inconsistencies in data collection; no systematic match between perceptions and 'hard' statistical data; and little attention to micro-macro linkages.

Earlier experiments in country evaluations from Mali (IOB 2000), Bolivia (IOB 1998a) and Egypt (IOB 1998b), evaluation of the 'Women and development' approach (IOB 1998c) and evaluation of SNV Netherlands (1998d) identified similar shortcomings in terms of their poverty focus: they were anecdotal and impressionistic; representativeness was questionable; and gathering group perceptions was time consuming.

Can we learn from other experiences?

Poverty impact studies by the Operations Evaluations Department (OED) of the World Bank and the British Department for International Development (DFID) face the same methodological challenges: attempts to gather the perceptions of the target group are impressionistic and provide few concrete data. Perception studies exist but are unrelated to project interventions. In 'Voices of the poor', the World Bank collected the voices of more than 60,000 poor women and men from 60 countries, in an unprecedented effort to understand poverty from the perspective of the poor themselves (World Bank 2002). In his book 'Whose reality counts: putting the first last', Robert Chambers (1997) argues that central issues in development have been overlooked, and that many past errors have flowed from domination by those with power. Development professionals need new approaches and methods for interacting, learning and knowing. Through analyzing experience of past mistakes and myths, and of the continuing methodological revolution of participatory rural appraisal (PRA), Chambers points towards solutions, arguing that personal, professional and institutional change is essential if the poor are to receive greater recognition. Changing the face of evaluations should make it easier to reach these objectives.

In search of new approaches

IOB is currently undertaking an evaluation of the sector-wide approach (SWAP), assessing changes in Dutch policy and in the poverty orientation in policies of recipient countries. It will consider how eligible countries were selected. How were aid sectors in the country selected? Did SWAPs actually get off the ground? Did SWAPs enhance the focus on poverty alleviation of recipient governments in the Poverty Reduction Strategy Papers (PRSPs)? Complementary evaluation of poverty impact in selected countries (Burkina Faso, Zambia, Bangladesh, and Bolivia), currently in a preparatory phase, will be looking systematically into target group perceptions. At the same time, it will try to link changes in the poverty situation with changes in the development architecture through the introduction of SWAPs. Literature and document research on the selected countries, focusing on overall development trends (economic, social, institutional, governance) and taking a timeframe of 20 years, will provide the background. Fieldwork will be undertaken in two selected regions in the countries where Dutch-financed interventions took place, focusing on 'groundtruthing' the information gathered during the desk study, and exposing findings to target group perceptions. These evaluation findings will be linked to the question of attribution. Finally, there will be country-specific reports plus a synthesis for the four countries.

For the planned evaluation of poverty impact in selected countries, IOB is currently preparing a workshop which will discuss the methodological challenges highlighted here. The workshop will need to consider how the concept of poverty should be dealt with. Should poverty be defined in a holistic way, narrowed down to target group perceptions or predefined and/or restricted by the evaluator? What methods should be applied for perception studies: PRA, the Method for Impact Assessment of Poverty Alleviation Projects (MAPP), life history, transect, or something else? How do we measure impact of interventions on poverty without sufficient baseline data? Can this be done with counterfactual analysis or before-after comparison? Only when these challenges have been addressed will this, and other evaluations, be able to go beyond the outcome level and measure the impact on poverty.

References

Chambers, R. (1997) *Whose reality counts? Putting the first last*. London: Intermediate Technology, 297 pp

IOB (1998a) The Netherlands Programme Aid to Bolivia. *IOB Evaluations* No. 278, 232 pp. http://www.euforic.org/iob/en/bolivia98.htm

IOB (1998b) *Egypt. Evaluation of the Netherlands support to water management and drainage, 1975-1996*. The Hague: IOB

IOB (1998c) Women and development. Policy and implementation in Netherlands development cooperation 1985-1996. *IOB Evaluations* No. 271, 276 pp. http://www.euforic.org/iob/publ/reports/271e.pdf
IOB (1998d) 1998 *Evaluation of SNV in Benin, Nepal and Bolivia*. The Hague: IOB
IOB (2000) Review of the cooperation between Mali and the Netherlands 1994-1998. *IOB Key Sheet* from Report 285, 2 pp. http://www.euforic.org/iob/publ/keysheets/285e-keysheet.pdf
IOB (2004) Poverty, policies and perceptions in Tanzania: an evaluation of Dutch aid to two District Rural Development Programmes. *IOB Evaluations* No. 296, 188 pp. http://www.euforic.org/iob/docs/200409071439567376.pdf
OECD (2004) Lessons learned on donor support to decentralization and local governance. DAC Evaluation Series, 87 pp. Paris http://www.oecd.org/dataoecd/46/60/30395116.pdf
World Bank (2002) Voices of the Poor. Washington: World Bank

Rob van den Berg

2 Why did the chicken cross the road? Four principles to support collaboration within evaluation

On 25-26 March 2003, the Development Assistance Committee (DAC) Network on Development Evaluation, then known as the Working Party on Aid Evaluation, held a workshop on 'Partners in development evaluation.' This workshop was born of a growing sense of unease within the Network that it was discussing pressing evaluation issues which should really be discussed with a wider audience. This article is a slightly revised version of a keynote address to this workshop. The starting point of the keynote address was the premise that evaluators have by nature some difficulty in developing a common language: they are able to turn simple questions into difficult exercises. A children's joke is, for example, the question of why the chicken crossed the road. My challenge is that evaluators need to address these simple questions together.

Why *did* the chicken cross the road?

The debate on collaboration within evaluation is reminiscent of the question: 'why did the chicken cross the road?' The simple answer is that it wanted to get to the other side. But the evaluator wants to know:
- whether the chicken took the shortest path across the road, and whether it crossed the road in the quickest possible way;
- whether it can be established whether the chicken in actual fact reached the other side, and remains there, and is expected to remain there in the future; and
- whether the needs of the chicken have been met by crossing the road.

To answer these questions, the evaluator will enter into a participatory process with the chicken, establish an intervention logic, logframe or theory of 'how chickens cross roads' and 'intended short-term and long-term results of the chicken crossing the road' which will lead to the gathering and analysis of relevant data. A full discussion of chickens and roads, and all the ramifications and possible benefits of future cooperation, is needed. This article analyzes the challenges facing bilateral organizations and other partners in evaluation, identifying four new principles that could be adopted to support collaborative approaches.

Background

The DAC Network on Development Evaluation comprises mainly representatives of evaluation departments of bilateral donors. It generally focuses on bilateral concerns, partly because other concerns are difficult to address in isolation. There are at least two other constituencies which have great relevance to the debate surrounding approaches to evaluation in international cooperation: the Inter Agency Working Group on Evaluation (IAWG) which brings together all the evaluation units and offices of the United Nations (UN) agencies, and the Evaluation Cooperation Group (ECG) of the International Financial Institutions. These groups address the evaluation concerns of their respective constituencies. The most important group of actors in development, our partner countries, is not represented in these groups. Of course, wherever possible the Working Party, the IAWG and the ECG involve partner countries and/or other partners in their activities – and most importantly in their evaluations. Furthermore, the three groups exchange observers and collaborate on many issues. However, none of the three groups has a mechanism for consulting beyond its own members. A key issue facing the DAC Working Party is how to meet, exchange information and collaborate with the following constituencies:

- the evaluation community in partner countries;
- the evaluation community in non-governmental organizations (NGOs) and civil society, not only in partner countries but also in member countries of the Organization for Economic Cooperation and Development (OECD);
- any other evaluation communities, like the private sector, human rights organizations, humanitarian assistance institutions, networks or associations of professionals, and so on.

Principle 1: Joint finance should lead to joint evaluation

Concern about possible lack of communication channels between different evaluation communities was sparked by the challenges that Network members were facing in evaluating bilateral aid. First of all, many donors have initiated a shift in funding from projects to programmes, from programmes to sector-wide support, and from separate and clearly identifiable funding to co-funding, joint funding or even budget support. Although we know that many donors do not put all their money where their mouth is, this shift still poses a huge evaluation challenge. For many of the bilateral evaluation units, it entails a shift from project or activity level evaluations to higher level evaluations.

This trend towards joint funding also poses the challenge of attribution. It is no longer possible to track the contribution of one donor into a joint basket. A principle that we could adopt as a community in this regard, both to prevent evaluation bombardment and to recognize the problems of attribution, is that: where donors finance together, they should evaluate together (Principle 1).

Principle 2: Partnership in development should lead to partnership in evaluation

In the past decade, the donor community has tried to improve its credentials as a partner in development. Partnership is one of the four principles of the Comprehensive Development Framework. As the call for a truly global partnership, it is the last (and perhaps most difficult) of the Millennium Development Goals (MDGs). In all our bilateral programmes, we observe efforts to improve partnership. Again, the reality may not always match the rhetoric. For evaluators, however, increased partnership should have implications for joint learning and for mutual accountability. This brings us to a second principle that the evaluation community could adopt: increasing partnership should include partnership in evaluation (Principle 2).

Principle 3: Ownership of development should lead to ownership of evaluation

The third challenge is the drive towards ownership. This is also one of the four principles of the Comprehensive Development Framework, and it has been carried over into the Poverty Reduction Strategy Papers (PRSPs). It is a common refrain in evaluation findings: a lack of success, a lack of sustainability in projects, programmes and policies, all because of a lack of ownership. Apart from being thoroughly backed up by the evidence, the emphasis on ownership is based on a recognition that people do not 'get' developed, but develop themselves. Similarly, countries do not 'get' developed, but develop themselves. The role of aid is nothing more and nothing less than to support or to act as a catalyst for indigenous development. Again, the reality in international relations may not always match these lofty principles. However, the changing role of the donor, with greater emphasis being placed on the part played by the recipient, has implications for the evaluator. This leads us to a third principle that the evaluation community could adopt: increasing ownership should also lead to ownership of evaluations (Principle 3).

These principles are up for discussion. The challenges do not affect the ultimate purpose of evaluation: to learn from the past and to account for what has happened. The changes we need to discuss are in the composition of the teams, in the shift in emphasis in questions, in scope and focus, and in reporting. The purpose of this workshop is not to adopt new principles for aid evaluation – we do not have the mandate to do that – but perhaps we can formulate a way forward in development evaluation.

Principle 4: If aid becomes integrated into development, we should evaluate development

As a fourth and perhaps final principle in such a declaration, we need to challenge the central role that aid plays in our work. The principle of ownership states quite clearly that development is an endogenous phenomenon in which aid can only play a catalytic or supportive role. The MDGs pose the challenge of evaluating what has been done to reach these goals, locally, nationally and internationally. Should we focus on evaluating development rather than just the small contribution made by aid to development? And surely our partners are more interested in an evaluation of the overall development of a sector, rather than just the contribution of the donors to that sector?

The last MDG recognizes that endogenous development is only possible in an enabling international environment. This presents us with another challenge: ensuring that other areas of international policy, like trade, migration or issues of peace and security actually enable rather than obstruct development. Many donors and partner countries are already working on improving the international enabling environment. How will the evaluation community react to the challenge of the eighth MDG? Should we, as an evaluation community, go from evaluating aid to evaluating development (Principle 4)?

Existing forms of collaboration

The purpose of the workshop is not only to learn from each other and to discuss common concerns and issues of interest. It also aims to explore how we can meet the challenges of evaluation. Will we do so as separate communities, or will we be able to identify avenues for future cooperation? Fortunately, we can build on existing forms of collaboration. These include:

- The Active Learning Network for Accountability and Performance (ALNAP) in humanitarian assistance includes non-governmental participants. It has built up a significant track record in supporting, strengthening and learning from evaluations of humanitarian assistance. Is this a model which can be adopted elsewhere?
- Both the evaluation of the Comprehensive Development Framework, led by the Operations Evaluation Department (OED) of the World Bank and the Joint Evaluation of External Support to Basic Education, led by the IOB Policy and Operations Evaluation Department of the Ministry of Foreign Affairs, the Netherlands, have concluded new and innovative partnerships with partner countries and an interesting mix of donor organizations, providing examples of joint international evaluations.
- The Working Party works on best practices, guidelines and lessons to be learned through a series of evaluations on related topics. Increasingly, partner countries, NGOs and the research community have become involved in this kind of work, often through involvement in seminars in which draft reports were discussed. Will it be possible to broaden the scope of best practices and synthesis work of the Working Party to make it relevant to other evaluation communities?
- The three well-established evaluation communities tend to work through so-called task forces set up to tackle specific jobs. These task forces tend to include external partners, depending on the subject. This is an existing vehicle for collaboration across communities which has not yet been used explicitly for that purpose.
- There are currently some mechanisms for information exchange on the Internet, such as the Development Gateway and evaluation inventories, which could enable and promote an ongoing debate within the evaluation communities on issues of common interest. Should we look for additional opportunities? Is there a willingness to facilitate discussions?

The greatest challenge is undoubtedly the mobilization of our colleagues and partners in the South. The number, scope, level and quality of joint evaluations are increasing, yet there is room for improvement. Training, creating an enabling environment, supporting institutions and individuals – things have got better in recent years, but can improve further. Funding may be an important obstacle. Many of us have funds available to undertake evaluations, but not to mobilize capacity in the South. Other channels of support need to be explored. Ideas need to be discussed. And the International Development Evaluation Association (IDEAS) may play a role in this.

Jurrien Toonen and Rob van Poelje

3 Monitoring and evaluation at the crossroads: meeting the challenge of sector-wide approaches and budget support

Approaches to monitoring and evaluation (M&E) are changing in response to changes taking place in development as a whole. The introduction of sector-wide approaches (SWAPS) and budget support are two recent developments that call for innovative M&E systems. In addition, the appearance of a number of global initiatives, such as the Millennium Development Goals (MDGs) and the Global Funds, together with a renewed focus on poverty reduction, are also playing their part in the development of new evaluation approaches. At the same time, the trend away from bilateral towards multilateral aid mechanisms and partnerships will influence the practice of evaluation in the coming years. These changes represent a considerable challenge to M&E.

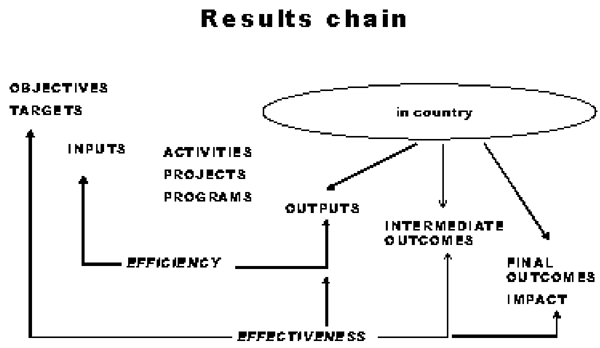
The project approach and the logframe

In the recent past, donor support was always managed by a *project approach*, limited to well defined geographical areas or to certain levels, specific programmes or well defined services. Both donors and national governments increasingly felt that the project approach was inherently flawed. Three main shortfalls were identified. Firstly, the coherence of the sector was hindered by creating 'islands of excellence', with their concomitant 'islands of failure'. In addition to this, donor-driven elements in projects were felt to incur the risk that national authorities would not feel ownership of the projects, resulting in sustainability constraints. And, thirdly, M&E exercises of projects were a substantial task facing already overburdened authorities.

In the project approach era, evaluation generally focused on relevance, efficiency of implementation, efficacy and sustainability, and impact on the ultimate beneficiaries. The related approach of the results chain, enshrined in the logical framework or logframe, placed more emphasis on attaining the expected results and objectives related to inputs of human, financial and material resources (Figure 1).

New challenges facing evaluation

Given the shortfalls of the project approach to development, several donors and national governments agreed that development interventions should be targeted at a sector as a whole, such as agriculture, transport, education or health, the so-called sector-wide approach. This approach implied important change in the way aid is organized and in the relationship between donors and



Source: Scott 2005

Figure 1: The results chain

recipient governments. For the first time, government leadership of the development process was fully recognized with a resultant reliance on government procedures (Foster 2000) and the establishment of joint management systems by governments and donors (Cassels 1997).

Recognition of the shortcomings of the project approach also led to the introduction of budget support, a form of programme aid supporting a plan or policy programme developed by the recipient government, typically using a Poverty Reduction Strategy Paper (PRSP). This budget support is channelled directly into the financial management, accountability and procurement system of the recipient country. These features, like that of the sector-wide approach, are intended to strengthen national ownership of policy and policymaking, to develop government and administrative capacity, and to improve aid effectiveness.

Both sector-wide approaches and budget support are based on the premise that donors are withdrawing from the management of projects in exchange for a place at the policy table. In this context, focused M&E may then be the only instrument left for donors to influence the direction of a programme. Thus, if a sector plan, owned by national authorities, is at the heart of the sector-wide approach, a well-functioning M&E system must be its backbone. Donors and recipient countries will need to negotiate the form this system should take and what type of information is essential for each stakeholder in their new roles.

In this new context, a number of operational challenges are facing M&E: the question of accountability and attribution; methodological challenges; the nature of partnerships and how to evaluate them; the need for effective monitoring processes; and the implications of the new focus on poverty.

Accountability and attribution

New approaches to evaluation will place more emphasis on accountability. Until recently, accountability of development aid was mainly related to financial and administrative issues and value for money. In the new search for accountability, donors are pushing for results-based management and results-based budgeting, management strategies focusing on performance, and achievement of outputs, outcomes and impacts. This approach stems from a criticism of governments and donors who have tended to focus too much on resource inputs *(what they spend)*, activities *(what they do)*, and outputs *(what they produce)* – and less on being responsive to people's needs and the wider impacts on society.

With a sector-wide approach, results should also become discernable at a higher level of outcome or impact. This will lead to the determination of process indicators like 'blockades in the legal or institutional framework have been reduced'. Or in broad sector indicators like 'the vaccination coverage' which will not only depend on external support.

Attribution remains a problem given multi-donor support for these new approaches. In this context, it is difficult to determine to which donor eventual effects may be attributed. In reality, many donors (certainly the Nordic so-called like-minded donors) will agree that results should be analyzed for them jointly. However, it is not inconceivable that individual donors may need to demonstrate their own individual results in order to convince the taxpayer or their government masters.

At the same time, there is a growing demand from the taxpayer and donor governments for evidence on the impact of development aid. For this reason, impact evaluation, in simple terms focusing on the socio-economic change brought about by development interventions, is coming increasingly onto the M&E agenda. In line with this, there are also attempts to appreciate the impact of development aid on the ultimate beneficiaries in quantitative terms. Members of parliament in donor countries are not easy convinced by results like 'a sound sector policy was developed'. They want to know if more people had access to health services or to schooling. It is not clear on what issues they will hold development interventions accountable, and the risk exists that this will mean a return to management on details.

Methodological challenges

The new approaches to development aid set a number of methodological challenges for evaluation. These include:

- If a sector is supported as a whole, where could controls be found?
- If control groups are to be studied, case and control groups will need to be large, or the intervention needs to be 'narrow' if a statistical significant difference between cases and controls is to be measured.
- The effects of development interventions are often difficult to define sharply, making it difficult to define cases and controls, and facilitating the appearance of selection bias, confounding factors, etc.
- Should we look at the sector as a whole or should the sector be regarded as a bundle of sector projects? This choice is important for developing the logic of the indicators.

There are possibilities to overcome the methodological difficulties described above. Firstly, evaluation should lose the ambition to cover comprehensively the total of a sector. Instead, a reasoned selection of variables in the sector should be studied in terms of cause-effect chains in order to define key tracer indicators, also known as proxy or intermediate indicators. The difficulty in finding controls may be solved by studying interventions that were introduced in a step-wise approach – the areas that were affected by the intervention in a later phase could then serve as control areas.

Partnerships

While withdrawing from project support and aiming for a sense of ownership, the development of partnerships in multi-stakeholder environments has become increasingly common. Partnership is a general term, widely used to describe a range of inter-organizational relationships and collaborations. Two key qualities of a partnership are unity and direction: the unity between individuals and organizations; and the direction and goals of the partnership, shared among its members.

Working in partnership should produce better results than each of the partners would achieve alone; there should be an added value. At the same time, attribution will become difficult to establish. Key issues in the quality of the partnership include:
- shared values and objectives;
- the existence of basic trust between the partners;
- good communication;
- network attributes; and
- stability in the partnership.

Monitoring processes

These new challenges make it especially important that an appropriate institutional, legal and organizational framework is in place. For this, the processes of institutional development need to be monitored. Key elements of process monitoring include:
- identification of key processes for project (or partner) success;
- measuring of results against project objectives;
- continuous testing of key processes;
- selection of activities and processes to be monitored is iterative;
- both quantitative and qualitative indicators are needed, but main focus is on qualitative indicators;
- a two-way process is required where information flows back and forth between field staff and management;
- effectiveness of communication between stakeholders at different levels is a key indicator; and
- self-evaluation and correction are built in.

Some new tools have been developed which illustrate this focus on process. Firstly, Outcome Mapping focuses on changes in behaviours, relationships, actions, and/or activities of the people and organizations with whom a development programme works directly. These changes are called 'outcomes'. Secondly, the Most Significant Change (MSC) technique is a story-based technique. Its primary purpose is to facilitate programme improvement by focusing on the directions and changes as valued by the various stakeholders and thus moving away from less valued, predefined directions. MSC represents a radical departure from the conventional monitoring against quantitative indicators that is commonly seen in development.

Poverty reduction

National strategy plans have been developed to fight poverty. Debt relief has also been granted to a number of countries on the condition that the earnings will be spent on programmes of poverty reduction. Such programmes all state objectives that need to be monitored and evaluated. Global initiatives also demand national statistics to establish whether the MDGs will be attained.

The M&E of poverty is less straightforward than it seems. The question is how to assess the poverty focus in these programmes and to measure the effects on the poor. The effects of development interventions on the poor cannot be monitored because disaggregated data does not exist. Most data (like for MDGs) are aggregated national data, therefore hiding differences between geographic areas, etc. There are population based studies that give an indication (like Demographic Health Surveys) but these are extensive, extremely expensive studies with a 5 year time-frame that do not facilitate targeting of the poor. New ways should be found to meet these challenges.

The poverty focus and the MDGs also raise the question of political accountability. Who is accountable to whom? Such mechanisms for accountability have not been developed at national or at international level.

Towards a new monitoring and evaluation approach

In addition to the challenges coming from sector-wide approaches and budget support, a number of more general tensions still exist: the link between learning and accountability; and the need to provide a bridge between monitoring and evaluation.

Linking learning and accountability

It is necessary to develop M&E systems which provide for accountability but also facilitate learning. From whose perspective should this learning take place? From the perspective of stakeholders, organizations or project partners? Whose learning is important and what are the conditions?

Many so-called evaluations are, in fact, little more than a strengths, weaknesses, opportunities, threats (SWOT) analysis, auto-evaluations, peer monitoring or lesson learning. These may perhaps not meet objectivity standards in terms of accountability but may have more impact on the functioning of projects and programmes because of the learning component. These exercises are often undertaken by those who are responsible for implementation of the recommendations which is a great advantage.

Bridging monitoring with evaluation

A bridge needs to be built between monitoring and evaluation. An example of where this has been done is the M&E system for health reform in Egypt. Here, a number of indicators were monitored in a fixed time-frame. The related evaluation was based on fixed tracer indicators plus a number of variable indicators that could be decided upon according to the outcome of the monitoring system. In this way, M&E was designed as one single system.

Some changes in evaluation practice have already appeared as a result of the introduction of sector-wide approaches and budget support. There has already been a change from extensive sector-specific information (like vaccination coverage for each village) towards information for management (which area should receive more resources to upgrade the vaccination coverage). Joint evaluations do take place. Donors are, however, tending to put new M&E systems in place next to the existing ones, and are organizing evaluations of their new interventions as well as participating in the joint missions. This means that the burden of work in M&E for national authorities has not decreased but, on the contrary, has increased. This is probably because some donors still seem to be more interested in M&E for accountability rather than in strengthening national management capacities. This is in spite of the fact that these capacities are recognized to be an important weakness affecting the performance of development.

References

Cassels, A. (1997) *A guide to SWAps for health and development: concepts, issues, and working arrangements*. Geneva: World Health Organization

Foster, M. (2000) New approaches to development co-operation: What can we learn from experience with implementing sector wide approaches? *ODI Working Paper* No. 140, 43 pp. http://www.odi.org.uk/publications/wp140.pdf

Scott, A. (2005) *The MEFF methodology: a review of DFID's multilateral effectiveness framework*. Unpublished paper. http://www.mande.co.uk/docs/Methodologypaper%20final%2016%20March.doc

Chris Whitehouse

4 The ants and the cockroach: a challenge to the use of indicators

Imagine this scene: thousands of ants are dragging a dead cockroach towards their nest. It is an amazing feat of teamwork. But now, study any individual ant, and you will find that it may be doing any number of things; it may be a key player, pulling the dead cockroach in the right direction towards the nest. It may unwittingly be one of a minority pulling it in the wrong direction. It may be running around doing nothing, but just feeling very excited to be near the action. It may, through its very presence, be encouraging the other ants in their work. Conversely, its comments might be irritating the other ants, and lowering their motivation. It may be helping to clear the route, flattening the path and thereby easing the movement of the cockroach… or it may, in its excitement, be making the path more difficult.

So how to measure this ant's contribution to the task at hand? Ask it, and it will say it contributed actively, as it indeed shares the colony's dream of having the cockroach at the door of their nest. It can provide us with supporting information. For example, it can tell us for how many minutes it (itself) was pulling at the cockroach (but in which direction, it might not be able to tell us!). It can advise us on how many pieces of debris it moved (but off the path, or onto the path?) It can tell us how many suggestions it made (but did they help or hinder its colleagues?)

Or we can be more 'objective', and ask the other ants. But again, there may be factors that influence their reply: they might not like this particular ant; perhaps it had previously been alleged to have contributed to the collapse of the west wing of the ants' nest. They may not have noticed this ant. They may be too busy just now, and give a quick but inaccurate reply. Another option is to go to a higher level of assessment. We could measure the average speed of progress of the cockroach, but this would be affected by such factors as the weather, the gradient of the slope, the time of day, the nutritional levels of the ants, etc. Or we can just assess whether the objective was met: did the cockroach arrive at the ants' nest door? The broad measures may seem objective but they will be vulnerable to external factors. The more finely tuned measures (specific to the one ant) are less vulnerable to external factors, but more vulnerable to subjectivity.

A new idea: the queen ant introduces indicators

One day, the queen ant, after listening to an inspiring documentary on monitoring and evaluation on the radio, decides to improve efficiency. She announces that, henceforth, it is not enough for each ant to just ‘chip in’ to the tasks at hand. Next time there’s a cockroach to be moved, she is going to measure each ant’s contribution, and their rights to gnaw at the flesh of the dead cockroach will be awarded in proportion to their contribution to the effort. Furthermore, she instructs 10% of the ants to carry out the monitoring on her behalf. Each ant will have to report to his or her monitor ant, as well as to be observed by the monitor ant during his or her work. The monitor ants will collate the findings and report to her every quarter.

Of course, this initiative made a lot of sense. The queen had noticed that some of the ants were not pulling their weight, were free-riding the system. Indeed, she had begun to feel that those who ate the most were the very ones who did the least work. This new initiative was very timely, also, as the number of ants in her nest was rising, while at the same time the supply of dead animals for food was getting scarce. No, it was really high time to make a change.

That evening, there was considerable unease in the nest. The ants were all tired, of course, after the day’s work, and the prospect of being watched all the time was, to say the least, unsettling. Plus, as one of the more mathematically minded ants pointed out, they would now only have 90% of the workforce actually pulling the cockroach; the other 10% would just be patrolling up and down, taking notes. Not only that, they would have less time to do the pulling, as they would, each one, have to report to their monitor ant on what they each had done. Still, the queen was probably right: the lazier ants would either have to start working or leave the colony and, through the supervisory scheme, they would all put in a bit more. So at the end of it all, there would be more food for fewer ants! Yes, the queen was right. By the time they went to sleep, all but the laziest ants were happier than they had ever remembered.

The ants’ story continues: the ants and the dead beetle

Next day was the beginning of a new era for the ants, and they woke up early, looking forward to a new life. And when the report came through the nest’s intercom system that a dead beetle now lay 20 metres east of the nest, they rushed to work. Of course, knowing they had to report their activities to their monitors, they each made sure to greet their monitors; just to be sure they were recorded as having gone to work. They each carried with them a notepad, pencil and stopwatch so that, whenever they were pulling, or clearing debris, or cheering their colleagues on, they could note down how many times they had done each, when and for how long. After all, it would be disastrous if they concentrated so much on their work that they forgot to record it: they could find themselves excluded from the beetle-feast tonight!

The first delay came when the monitors decided to measure the beetle’s weight by finding out the minimum number of ants needed to hold the beetle full off the ground. They then had to record the distance between the beetle and the nest. The best way, they found, was to have all the ants stand head to tail in a long line, and the monitors counted how many ant-lengths lay from the beetle to the nest door: it came to a staggering 198,865,314! It was now mid-morning, and the ants were tired from both of these activities. The whistle blew, and it was time for action! With a sense of exhausted relief, the ants rushed to move the beetle... but the scene was one of unprecedented chaos. Never before had so many ants been so eager to be seen to be doing something useful. The monitors held their front legs to their ears; such was the din of thousands and thousands of well-meaning but contradictory suggestions. And the beetle all but fell apart because so many ants were pulling with all their might in opposite directions.

Eventually, the beetle, now in pieces, lay nearer to the nest door, again the distance was measured; and again the various parts of the beetle were weighed. Then the exhausted ants formed long queues to submit their reports to their monitors who, in turn, had to collate the many thousands of figures which were then submitted to the queen. She was, of course, delighted: it was really working. And she was pleased, too, to find that none of the ants had been found to be lazy.

The only slight disappointment was that the flesh of this beetle was found to taste not quite right and, as they went to sleep, many of the ants complained of feeling a little dizzy. Was it all the excitement and noise? Or was it just exhaustion? Or was there something wrong with the flesh of the beetle? In their enthusiasm that morning, none of the ants had noticed the strange coloured liquid which one of those big humans had pasted onto a leaf near where the beetle had died.

The dangers of indicatorism

The story of the ants illustrates the dangers of indicatorism. The key problems with indicators fall into three main categories: the time and resources spent on identifying and measuring indicators; the skewing effects that inclusion of indicators may have on programme design and implementation; and, perhaps most seriously of all, the fallacy of confusing concomitance with causality.

Problem 1: Time and resources spent on indicators

The story of the ants demonstrates something that many of us already recognize in the real world: that the requirement to report results entails using resources which could otherwise have been used for implementing a project. A balance has to be struck between a perfect monitoring and reporting system, and the desire to spend as much time and money as possible directly on project activities. To ensure efficient monitoring, we need to know (1) that indicators measure what they are supposed to measure; and (2) that the cost of measuring is not out of proportion to the project budget.

As noted earlier, there is a play off between indicators which are easy to measure, directly related to outputs, and which are *de facto* trivial (e.g. number of people trained), versus the more difficult to measure indicators of outcome which are far more significant but which, at the same time, are more vulnerable to external factors (e.g. improvements in government service to the public; or improved standards of living amongst the rural poor).

A review of logical frameworks in project documents shows a spectrum from the banal to the stratospheric:

a. One project might have the activity given as ‘information and communication technology (ICT) training for 20 information technology (IT) managers’; the output is ‘20 IT managers trained’; and the indicator is ‘number of IT managers trained’. Means of verification, an important element in any logical framework, would be ‘attendance record’. Yes, they did the training, the trainees were trained, and we can know it was a success by counting the number of people trained – but was the training course well designed? Was it useful? These questions remain unanswered.
b. At the other extreme, the activity could be the same, and the output (or really, perhaps, the outcome) defined as ‘20 IT managers able and motivated to provide full IT service to all the government offices in their districts’. The indicator could then be something like ‘number of breakdowns of computer systems which remain unrepaired 24 hours after reported breakdown’. Now, this indicator would very clearly reflect the quality of the training but maybe other factors will come into play as well. What will happen to these measures if one of the trainees has to attend a funeral on the other side of the country? What if the computers were using pirated or poorly installed software? What if a lightning storm caused an electricity surge across half the country? Yes, we have a fine indicator, but does it measure only what we want it to? The answer is ‘no’.

I can anticipate the response – ‘oh, you have chosen some particularly weak indicators’ – but I would say that such weaknesses tend to be the rule, rather than the exception. I look forward to the time when I can see a project document whose indicators are measurable, will truly measure what they purport to measure, and which are neither trivially meaningless nor so grand that they are compromised by external factors beyond the control of the project.

Let’s return to the ants again – what went wrong? Firstly, they invested heavily, to the detriment of efficiency, in measuring and monitoring. Indeed, 10% of the labour force was taken out to carry out this monitoring. Normally, time and resources invested in indicators should be kept to a minimum, and must remain proportional to the size of the programme being measured. Did the ants really need to weigh the beetle, and to measure its distance from the nest and, for each one, to record their contributions to the team effort? But another key error of the ants was that they focused on the measurables, and ignored the most important (but least easy to quantify) aspect of their task. They measured weight, distance and total ant-hours spent making suggestions and pulling, but ignored the most important but least quantifiable element – did the beetle taste OK? In fact, the most appropriate indicator for the ant colony may have been something as simple as nutritional levels of ants; or, at the broader level, average life expectancy in the colony. But then, as discussed above, these measures, although more appropriate, would be vulnerable to external factors – in this instance, to beetle demographics, climate change, use of pesticides by humans, etc., over which the ants cannot possibly have any control.

Problem 2: Indicatorism or the skewing of project design and implementation behaviour

Even more worrying than the above is the temptation for designers of projects and programmes to engineer their activities towards 'measurable' achievements. That is to say, focus on indicators can skew development programmes during their design stage. Current programmes may survive, as we implant backdated indicator baselines and target figures which become a useful annex to well-designed programmes. However, many of these good but otherwise 'wishy-washy' programmes may not have been approved had indicatorism been the fashion a decade earlier.

We have seen a shift away from the small independent projects building a bridge here, giving literacy training to a women's group there – and rightly so. It is recognized that we need a more integrated approach, where a bridge, literacy training and many other components should fit into a whole: an integrated programme. The danger with indicatorism is that it will result in a *de facto* re-division of the integrated whole into only the measurable parts. The 'wishy-washiness' that binds an integrated programme together is the very element which will fall away when we focus only on the measurables.

A related worry is that project designers, especially if they know that they, the same people, may be responsible for monitoring the project during implementation, may feel tempted to under-estimate the targets to be reached. Should we aim at 15 out of 20 of the non-governmental organizations (NGOs) achieving growth? No, better aim lower – perhaps we should aim for 5 out of 20. And then, great success, the project is completed and 10 NGOs have seen growth! Wow!

Yet another worry is that the priorities of project management personnel during implementation are likely to be swayed by this emphasis on indicators. Rather than invest time and effort in something which common sense would suggest was necessary or desirable, they may feel tempted to focus on those activities which will most speedily and easily achieve the targets set. Why risk allowing the lower class (and lower educated) people into the training programme? If we are aiming for 75% minimum to pass the final exam, then better to go for those trainees with a higher chance of passing. And certainly, don't get any women trainees as they will be bound to miss some classes due to family commitments.

One is reminded of the absurd situations that arose in the Soviet Union. For example, one of the targets set for a shoe factory was 'number of shoes manufactured in a year'. Since it took time and effort to adjust the machinery to make right shoes after making left shoes, and the target only specified total number of shoes (not pairs of shoes), then it just made sense, didn't it, to churn out only left shoes for the whole year? We can laugh at the Soviet Union now – but would we like to have people laughing at the development community in the same way in years to come?

A focus on indicators, therefore, can have detrimental effects on project design, and on implementation. So, if you are going to have indicators, and are going to take them seriously, enormous care must be taken to ensure that the wording of the indicators is sufficiently tight that the project focus with respect to indicators is exactly matching the project focus as a whole; and that the project focus would remain valid even if indicators were not on the cards.

Can anyone advise on which indicators the queen should use? They would need to be indicators of factors over which the queen and her colony can have control, they should be measurable and (most awkwardly) they should be sufficiently directly relevant to the needs of the colony that introduction of these indicators will not skew the colony's programme activities away from their prime needs.

Problem 3: Scientific validity – concomitance, causality and control

The final challenge to indicatorism is perhaps the most serious. The introduction of indicators at first sight appears to demonstrate a logical, scientifically valid system of auditing, monitoring, proving transparency and accountability. Yet this, as shall be explained below, is definitely *not* the case.

Firstly, in order to see the fatal flaws in the logic of indicators, we have to draw a very important distinction between *concomitance* and *causality*. When an event A is followed by event B, it is tempting to say that A caused B. When the sun sets and darkness follows, especially when it happens time after time, we deduce that the setting of the sun causes the darkness that follows. However, one can also think of instances where A is followed by B, but we know (through our scientific understanding) that A doesn't cause B. For example, the chiming of Big Ben for 6 o'clock in London may be followed every evening by the departure of the 18:01 train to Bristol. Does the chiming cause the train to depart? No – and this can be proved if you sabotage Big Ben to stop it chiming one day, yet still you will see the train depart. Or you could sabotage the engine of the train, and find that even after the clock chimes, the train doesn't go.

Scientists are very aware of this distinction between concomitance and causality. Before any medicine is approved for sale, causality has to be proved: it has to be shown that not only those suffering a headache are cured when they take medicine YYY, but also that their headache is not cured if they don't take medicine YYY. Indeed, modern experimental methods require that double blind tests are carried out. Out of a group of 50 volunteers, 25 would be given medicine YYY, and the other 25 would be given an identical-looking but harmless and inactive placebo, where neither the patient nor even the person administering the treatment know who is getting the real medicine and who is getting the placebo. It would only be through this kind of test that YYY could be demonstrated to work.

In the development world, through indicators, we also hope to test the validity of treatment YYY (e.g. a training programme for civil servants) as a means to arrive at situation ZZZ (i.e. improved service for the public). But what do we do? We provide YYY, and then claim, with zero scientific basis, that situation ZZZ was as a result of the provision of YYY. We fail completely to have a control group – to be able to compare what actually happened to the target group with what would have happened if they hadn't received this programme.

Does this make sense? Let's use a stupid example to show what I mean: I go to a 5-year-old child, who wants to be taller. I say to his parents that I can help. First let's measure his height. Then, let me give him three carrots to eat, once a week, for three years. Then, at the end of three years, let's measure him again. If he's taller, then we know that carrots make you tall. Or, I can go to the government, and offer a course in hydrology for all government hydrologists. Without having any control group (i.e. of hydrologists who are not receiving this training), then how can I show that, simply because they are offering improved hydrology service five years later (outcome indicator), that our hydrology course had any positive influence? Only if you offer a 'placebo' course (e.g. in Asian cookery) to another group, and if the hydrology-trained people fare better than the cookery-trained ones five years later, can you start to show that your course was successful.

It is not enough to show improved scores as against a baseline because progress (whether of the child, getting taller; or of the hydrologist getting wiser) will often happen even without carrots, or without training programmes. We need to have a control group, outside of the support programme, against which to compare any improvements.

It has long been a cause for concern, even before indicators became fashionable, that project reports of income generation programmes, for example, would highlight growth in income for their target group as a success, without seeing what negative impacts there may have been in neighbouring villages. But now we are taking a further, and more dangerous step, of claiming scientifically measurable progress: of giving numbers to 'prove' success. It is tempting to assume that if a certain training programme is followed by an increased efficiency in an office, then the training was a success; even more tempting if the project management shower us with pre-training and post-training measures of office efficiency. But, without a control group against which these figures can be compared, these figures are meaningless. It would be a cause for grave concern if those funding such programmes were so impressed by these 'figures' that they would prioritize further support for this programme, to the detriment of other programmes which might be more valuable, but where the project managements are not bombarding the donors with convincing-looking, but essentially flawed, statistics.

Conclusion

The logical framework is a highly valuable tool for development workers, but a focus on indicators can be detrimental. What is being challenged here is the validity and efficacy of a 'measuring stick approach'. We must take care not to assume that a series of 'before' and 'after' statistics demonstrates the worth of a project, nor should we assume that investment in indicators would have positive impact on the quality of design, or of implementation, of a project. Indeed, indicatorism can have a significant negative impact.

Thomas Winderl

5 A pot of chicken soup and why Brits need umbrellas: in defence of indicators. Reply to the 'The ants and the cockroach' by Chris Whitehouse

Imagine you feel suddenly hungry. While the hydrologists in Chris Whitehouse's article 'The ants and the cockroach' were well versed in Asian cuisine, you would like to stick to something less extravagant. You want to re-heat the succulent chicken soup (no, not for the soul) from last night. While the soup is on the gas stove, you want to monitor progress of the re-heating process although, of course, in cooking no one uses the word 'monitoring'. After all, you hate cold soup, and you don't want it to boil over and burn, nor to gulp down a lukewarm soup. So what do you do? You choose an appropriate indicator.

There are many to choose from. You might want, for example, to visually monitor the soup surface in the pot, and remove the pot once the soup starts to boil. You could use your extended experience in soup re-heating, and use a time frame (e.g. 3 minutes 30 seconds for a half litre pot) as an indication that the soup is ready to eat. If you are a connoisseur, and a rather adventurous one at that, you might stick your finger into the pot every minute or so to feel progress. We could easily think of more possible indicators.

Let's take another day-to-day example, this time not gastronomic but rather meteorological. You intend to take a stroll to the local supermarket, and want to know what clothes to wear, and if you need to take an umbrella (doesn't apply to Brits). What do you do? You would probably look outside the window, see the sun, see the trees moving in the wind, see people wearing t-shirts in the streets, and conclude naturally that it is a sunny, but windy summer day. There is (normally) no need to measure the exact temperature, or the speed of the wind. These casual indications tell us what we want to know: not to wear the fur winter coat, not to take an umbrella (except for the Brits), but to dress lightly, and maybe to take a light jacket against the wind. Even if you normally would not say 'Darling, I'll go and monitor the indicators for the weather', this is exactly what we do every day.

Talk about indicators and result-orientation is a part of contemporary development speech. No self-respecting expert in human development would go without it. Indicators and result-orientation seem to be, as so many trends before, a temporary development fashion, like 'participatory processes', 'structural adjustment', 'grass-root based work', or 'micro-finance'. While result-orientation and indicators emerged in the 1980s, they are here to stay, and with good reason. Indeed, indicators – and monitoring progress in general – come natural to human beings. We have been using them, are using them, and will be using them all the time. If you want, they have been around since the cave men, and might even be a significant part of what makes the *Homo sapiens* a highly successful species. True, in the caves – and even nowadays for that matter – we didn't call them indicators, but rather common sense. A lot of work still needs to be done to de-mystify indicators, and look at them as a standard tool for planning and monitoring progress in any development situation too difficult to appraise immediately.

Argument 1: *More* time and resources

While re-heating yesterday's soup from the fridge, few people would design an elaborate monitoring process with complicated indicators to monitor the heating process. Nobody with a sound mind would call for high-tech instruments, or extensive surveys among soup particles. You would not call in a team of soup-cooking specialists to help you re-heat your soup. You can do it all by yourself. It is relatively easy and not very time-consuming.

Chris Whitehouse is arguing that time and resources invested in indicators must remain proportional to the size of the programme. I couldn't agree more. However, the reality is that there is *not* enough time and funds invested in designing and monitoring clear and relevant indicators. As we know, decades after the 'invention' of logical frameworks and indicators, the inadequate and skewed logical frameworks and bad or plainly wrong indicators still dominate the development business. And here I utterly agree with Chris: bad indicators are worse than no indicators at all.

Let us look at the resources spent on thinking up and monitoring indicators. As a rule of thumb, guidelines for monitoring recommend earmarking about 2-3% of the project budget for this purpose. In development reality, however, it is highly doubtful whether many projects and programmes spend even that much time. In a 3 million USD project, that would mean 90.000 USD dedicated to creating meaningful indicators and monitoring them, which is hardly the case in the contemporary development business.

The same holds true for wasted time. It is true that the selection of a good indicator requires a collective, time-consuming thinking effort. This is not easy. While creatively thinking up good indicators is the time-consuming part, most methods to monitor change using an established indicator are easy, quick and cheap (and they usually don't involve dangerous things like sticking your finger in boiling soup). But if a whole month is spent by the project designers to come up with quality indicators and monitor them in a 5-year project, this would amount to only 1.7% of the total time involved in the project implementation. And it is next to impossible to find programming exercises where even one week is invested in thinking up indicators. It just does not happen.

Here, Chris Whitehouse is indirectly pointing out an important reality. It is not the case that either too much time or too many resources are invested, but that the result is generally of very poor quality. A common mistake is the definition of too many indicators. What do you do if half of them point at a strong improvement, and the other half at a deterioration of the situation? Nothing is won, and we would be better off without them, using our common sense and informed assumptions. Other indicators indicate the wrong thing. They were not thought through, and lack creativity. In these cases – which are without doubt the majority in most development organizations – the monitoring of these indicators brings no additional benefit. Time and resources are lost, which could have been spent more effectively on the project itself.

Chris Whitehouse is absolutely right when he argues that 'a balance has to be maintained between the time and effort spent doing what needs to be done, and that spent reporting.' However, we should interpret his argument as a case for *more* resources and time spent on careful planning and monitoring, rather than less. Given the choice between putting your money into an activity where impact is proven through a (mostly lengthy) narrative using a host of (hopefully not wild, but informed) guesswork, and well thought-through outputs, outcomes and impacts with indicators, a 3-4% share of time and resources is well spent indeed.

Argument 2: Skewed implementation behaviour

Chris Whitehouse's second set of arguments, based on the premise that indicators can skew development programmes during the design stage, is only partially valid. Let me point out some of his misconceptions.

Division of integrated projects

Firstly, Chris' worry about the re-division of integrated projects into its measurable parts is lacking any base. If indicators are taken seriously, they do not only include input, output and outcome indicators, but also long-term goals. And only extremely poor logical frameworks allow for multiple goals. The norm is to have a clearly defined goal at the top level of a logical framework. As we know, if outcomes are not contributing to the overall goal of a project, they should be deleted from the framework. Rather than dividing integrated projects, the logical framework (not so much the indicators), if applied properly and rigorously, force project designers to focus on a coherent set of outputs and outcomes to achieve one goal.

Wishy-washy project design

Secondly, Chris Whitehouse seems to lament that the 'wishy-washiness' of old-style project design will fall away when focusing only on indicators. While it is true that old projects without indicators might sometimes have resulted in valuable development improvements, they did so *despite* – and not because of – the lack of indicators. The approach of 'let's do something and something good will probably come out of it' is not an option any more. This romanticizing image of the positive spill-off of random project activities is clearly, and rightly, a matter of the past, mostly due to the overall poor results of development aid and the increased accountability of donor countries to their clients, the taxpayers.

The process of constructing indicators

Thirdly, an extremely valuable aspect of including indicators in project design is the process itself. It forces the designers to define better what outcome is intended. While it is easy for the United Nations Development Programme (UNDP) to set as a goal the increased capacity of the Royal Civil Service Commission (RCSC) in Bhutan, it is much harder to think up indicators which address what this really means. What is the RCSC actually doing? What aspect of what the RCSC is doing do we want to enhance? Are we happy with 5 people working more effectively and efficiently, or are we targeting 50 people in the RCSC?

Under-estimating the target

Fourthly, 'The ants and the cockroach' argues that there is a strong incentive to under-estimate the target if the designers and those undertaking the project are the same. This is indeed true, if one does not link the framework with the inputs, namely the funds used to achieve the targets. Using Whitehouse's example, it might look acceptable to spend 75.000 USD to help 15 NGOs achieving growth (5000 per NGO). However, if the designers set their target at 5 NGOs, that is 15.000 USD per NGO, the donor organization should decide that this is too much, and refuse the funding of this activity. Agreed, the linkages of inputs and outcomes in logical frameworks are still weak in development practice but this does not indicate a flaw in the concept. On the contrary, the tendency is more and more to ask: how much money do we need to spend to achieve outcome X? Could we do it in a more effective way? What other options do we have at our disposal?

The focus on indicators

Fifth, Chris Whitehouse describes a valid problem, namely the focus on indicators in project implementation. Citing the convincing example of the Soviet shoe factory churning out left shoes in order to achieve the target set, he provides us with a hilarious description of the importance of getting your indicators right. Assuming the intended outcome of the Soviet policy was the provision of sufficient, cheap and high-quality pairs of shoes to its citizens, the chosen indicator was evidently flawed on the output level. It requires careful and creative thinking, and anticipating the fallacies of indicatorism for the implementation phase, to create indicators that capture the intended output, outcome or goal. Rather than being an argument against indicators, it is a good example of the peril of brainless indicators.

Argument 3: Picasso, not Einstein

Chris Whitehouse's third main – and potentially most devastating – argument is that indicators aspire to create a logical, scientifically valid system of monitoring and auditing, providing transparency and accountability. Once again, it is not his line of argument but his underlying assumptions which are flawed. While indicators – together with their twin, the logical framework – aspire to introduce a more structural and logical thinking into the complex realities of projects, it is a misconception that indicators pretend to set up a scientifically valid system. His argument is based on the popular myth that indicators, maybe because they operate with numbers, are science. They are not.

Let me explain, in more detail, two of the fallacies in dealing with indicators: first, they are not – and do not claim to be – scientific. Second, they do not normally measure progress or regress. We cannot claim for most indicators that they are anyway close to scientific measurements. Indeed, the creation of indicators is by no means a scientific action. Although most experts will agree whether a particular indicator is better or worse, there is no systematic way of deciding among indicators which are equally good or bad. In short: far from being science, the development of indicators is art, combined with a large portion of systematic, logical thinking, and an even larger portion of common sense. If you look at the process of how indicators are being thought up, you will see contradictory elements. On the one hand, a certain number of tools are necessary to help in the process: problem trees, logframes, etc. But, on the other hand, the process demands a high degree of creativity, out-of-the-box thinking, or de Bono's 'lateral' thinking. The choice of the right indicators is an art rather than a science.

The second misconception is related to the first one: indicators do not measure progress. One doesn't have to look too closely at the word itself to find out that indicators – well, as the word tells us – 'indicate' a direction. They tell us which direction a change possibly takes, or whether there is hardly any change. If the average time the ant tribe takes to bring in a juicy beetle over the year is half the time it took them last year, it indicates to us, beyond reasonable doubt, that they are doing much better than before. Indeed, the better an indicator, the better it matches the actual direction of a change. A good indicator will represent the actual development as closely as possible. However, this is not always the case.

Let's look at another example: Ms Poor wants to be rich in 10 years. Having defined 5 million USD as her personal threshold of wealth, she can measure her assets easily by looking at her monthly bank account statement. By monitoring this simple indicator (although no one would call it that way, it's just common sense), she knows with certainty where she stands in regard to her ultimate goal. This is one of the few cases where the indicator 'USD in bank account' comes close to measuring her wealth. There are very few indicators which manage to come so close to indicate a change. Ms Poor could have taken another indicator: the value of the cars she is driving. We can assume (yes, that's what indicators do!) that the richer she grows, the more expensive (or bigger? that would be another indicator) her car will be. But it's not as good as the bank account. It could happen that – because she saves all her money – she initially sells her expensive Porsche, and keeps on buying second-hand vehicles. She could look at the dividends her money creates. But stocks go up and down, and – although the general direction might tell us a bit about her financial status quo – we might get a distorted impression as well.

Here is a last example how indicators indicate and not measure. The UNDP supports the Royal Government of Bhutan in its negotiations to join the World Trade Organization (WTO). Finding an indicator is a bit tricky here. A possible solution could be the number of years the WTO grants Bhutan before it has to phase out subsidies for fertilizers. What a wacky indicator, you might think. But think again: subsidies for fertilizers are generally a major point in accession discussions. One could readily assume that the better the Government negotiates with the WTO, the longer the transition period for fertilizer subsidies might be. Together with a few more indicators, this might well indicate to the donor agency how well its money was spent. The point here is: the length of the transitional period for fertilizer subsidies clearly does not measure at all the skills involved in the negotiations. It measures, well, the length of the transitional period. But still the indicator would be – among others – a clue to how well the Government negotiated.

Conclusion

Indicators come naturally to human beings and we are using them all the time, although we usually do not call them indicators but rather common sense. Indicators and a result-orientation are part of contemporary development speech and they are here to stay. Chris Whitehouse argues that too much time and too many resources are invested in creating indicators but this does not reflect the current development reality. Instead, more time and resources need to be spent in constructing well thought-through indicators for evaluating development interventions.

Chris Whitehouse's second set of arguments, based on the premise that indicators skew implementation behaviour, is only partially valid. While indicators and the logical framework aspire to introduce a more logical thinking into the complex realities of projects, it is also a misconception that indicators pretend to set up a scientifically valid system. Indeed, as Chris claims, the development of indicators is art, combined with logical thinking. Finally, the author and Chris Whitehouse are both agreed: bad indicators are worse than no indicators at all.

Paul Engel, Charlotte Ørnemark and Arin van Zee

6 When will we ever learn? Perspectives on learning from evaluation

Development agencies are consistently required to improve their performance, in terms not only of project outcomes, but also of the quality of their programming and their institutional capacity. In a very practical sense, evaluations are now perceived as learning opportunities. It only seems logical, therefore, to try and improve the internalization of evaluation results at different levels. But how can this be done? What dilemmas do we face and what have we learned so far?

Evaluation has always been about learning, about how to be accountable, how to be transparent, how to learn from experience. The issue today, therefore, is not whether it is desirable to learn. Instead, the question is: who should learn, why should they learn and how should they set about it? Quality issues also come into play if learning is to be widely shared. Here, the term evaluation is used in its broadest sense, encompassing the gathering of evidence as well as its processing and valuing. It also includes the more extensive societal interaction with respect to evidence gained through monitoring and evaluation.

Internalizing evidence from evaluations: three perspectives on learning

Perspective 1: Development policy and programming

Probably the most straightforward way of enhancing learning is to look at its internalization in development policy and programming. However, many of the existing feedback mechanisms are still mainly one-directional, drawing on the logic of information dissemination to selected target groups rather than communication around evidence as an iterative learning process. The 2001 workshop on Evaluation Feedback for Effective Learning and Accountability (EFELA), organized by the Organization for Economic Cooperation and Development/Development Assistance Committee (OECD/DAC) Working Party on Evaluation, now known as the OECD/DAC Network on Aid Evaluation, reviewed some of the current experiences (OECD 2001).

Most reports have called for the establishment of what may be called policy innovation networks, mobilizing Southern partners and stakeholders to engage in existing learning and feedback routes. The International Fund for Agricultural Development (IFAD) recognizes the need to 'shift the fulcrum of evaluation feedback to the South' to encourage more direct links between feedback of findings and the planning and monitoring of country programmes.

Many development agencies emphasize results-based planning and management as a way of improving the practical use of evaluation results. The European Commission's Fiche Contradictoire and the UK Department for International Development's (DFID) Public Service Agreements are examples of this. The EFELA review regards the communication of lessons through the mass media as increasingly important. Finally, it also calls on development agencies to do more to share their experiences with each other.

Perspective 2: Within organizations and among partners

A second approach focuses on organizational learning, recognizing that development processes result from actions and interactions by a set of diverse stakeholders. Active participation, capacity building and learning by all these actors are a fundamental rather than an instrumental condition. The locus for change is the facilitation of collective rather than individual learning. As a result, policymakers become one among many, rather than the only intended learner.

An organizational learning approach to evaluation not only fundamentally changes the way social actors relate to each other, it also requires a radical shift in the role of the evaluator. All actors, including the evaluator, have to recognize they are part of a joint learning effort. In such an epistemic community, the evaluator becomes a facilitator in a joint inquiry rather than an expert wielding an objective measuring stick. Yet as Rebecca Sutton (1999) argues, such communities run the risk of clique-building, reducing the diversity of opinion if the discourse is captured by the most vocal actors in the group. Critical self-reflection must be maintained in order to build in 'reality checks' and thus avoid too narrow a discourse among closed circles.

Perspective 3: Society at large

A third perspective focuses on a type of learning that leads to change in society at large. When the sharing and interpretation of evidence extend beyond those directly involved in the evaluation process, conflicts of interest are common and consensus becomes the exception rather than the rule. The question then is whether and how interested parties can exert pressure for change and whose interpretation of the findings is the dominant one. The co-existence of multiple truths requires a more transparent analysis of findings and the creation of sense-making fora for stakeholders to interpret and validate the evidence. Some commentators, including Norman Uphoff and Jerry Combs in their paper 'Some things can't be true but are: rice, rickets and what else? Unlearning conventional wisdoms to remove paradigm blockages' (2001), stress that such broadening of the interpretation of evidence to a wider audience and different interest groups can help to avoid 'paradigm traps' among scientists and policymakers that limit their views on development options.

The perspective on societal learning from evaluations, i.e. learning that goes beyond a small number of directly involved stakeholders, could have far-reaching consequences for our thinking on development cooperation. Traditional evaluation as we know it may gradually fade into the background, to be replaced by multiple forms of evidence-gathering and sharing among diverse groups of stakeholders; the adaptive management of resources; and multiple communication and negotiation processes at various levels. There will be a greater need for good governance to create enabling conditions for such processes and for conflict resolution between stakeholders. As a result, governments and donors will become crucial players who may either enable or prevent society from learning. As both development policymakers and field practitioners alike have urged, this perspective links evaluation with governance issues.

Dilemmas at the evaluation-learning interface

Various dilemmas arise in trying to turn the current evaluation practice into a learning process.

Why strengthen the learning function? What other purposes should evaluation serve?

Evaluation serves a number of different purposes. A growing emphasis on learning from evaluation means a shift in intentions. Traditionally, control has been an important purpose. From this perspective, the aim of evaluation is to enhance transparency and accountability, particularly from the donors' point of view. Another vital purpose is assessment, i.e. deciding whether the agreed objectives have in fact been achieved. With learning in order to improve performance now becoming an increasingly important purpose, evaluations could eventually become geared towards adaptive management. Institutional learning and the development of the institutions responsible for managing development would be prerequisites for this. So would pursuing one purpose imply a trade-off with another? Might these aims be mutually exclusive in an evaluation context? Or would a shift in intentions reflect a growing complexity of evaluation functions rather than a shift from one to the other?

What do we mean by learning?

Learning is a buzzword, often used but not often clearly defined. As a result, one of the challenges defined by the 2001 EFELA workshop (OECD 2001) is the need to unpack the learning concept. After all, there are all sorts of different approaches. According to Ruud van der Veen (2000), different types of learning in adult education are linked to distinct cognitive and motivational processes and fulfil complementary roles. In the their chapter in the book 'Wheelbarrows full of frogs', Christine King and Janice Jiggins (2002) utilize organizational learning theory to distinguish between single-, double- and triple-loop learning, reflecting the degree to which underlying rules, values, norms and behaviour are truly affected. In the same book, Niels Röling (2002) focuses on cognition, the process by which an organism deals with changes in context. He stipulates two fundamental drivers for social learning: the coherence sought among values/emotions/perceptions, on the one hand, and theory/interpretations/

actions on the other; and the need for correspondence between the above elements and the prevailing context.

Whose learning are we talking about?

At the three levels of learning outlined above, the development policy and programming focus mainly emphasizes learning by donors and policymakers. Hence, there is a risk that the participation of local actors is simply instrumental, i.e. more intent upon extractive actor consultation or 'rubber stamping' policies than effective participation in decision-making. The organizational learning approach recognizes the complementary role of stakeholders and a multi-layered institutionalization of findings. It emphasizes the collective understanding of relevant development issues and finding the way forward. Finally, the societal learning approach recognizes the existence of potentially insurmountable differences in perspective among interested parties, yet at the same time acknowledges the need to negotiate sustainable answers to common challenges. In line with its focus on adaptive management and aspects of governance, it focuses on institutional development for improving societal learning and decision-making.

Widespread learning versus the quality of the learning process

In their book on 'Participation: the new tyranny', Bill Cooke and Uma Kothari (2001) argue that unless evaluation instruments and evidence-gathering tools go hand in hand with an outspoken commitment to decentralizing decision-making, engagement with local communities may increase the risk of co-opting them into existing development paradigms, thereby reinforcing rather than counteracting existing inequalities. At the same time, quality could be weakened by a lack of uniformity in the methods used, jeopardizing the scientific rigour of the evaluations. The counter-argument is that quality is enhanced by a greater richness of opinion. Unfortunately, to date, very few studies have paid much attention to assessing learning and its concrete outcomes.

Power relationships and learning

Skewed power relationships, that are inherent to development practice due to the unequal access to resources, affect the process and outcomes of learning even when local ownership and mutual learning are intended. In their analysis of institutional learning with respect to local forest management, Paul Engel and colleagues (2001) point to the need for the empowerment of local stakeholders and conclude that, even then, the national policy context may make or break the process at a local level. This leads us directly to the question of the management of learning: who sets the rules? Who enforces them? Who evaluates the evaluators?

Learning and the role of the evaluators

Where evaluation is used for learning and adaptive management, this implies a fundamental change in the role of the evaluator. From being a remote, research-oriented person trying to systematize the known and unearth the hidden, he or she becomes a process facilitator. His or her role is to help design and organize others' inquiry and learning effectively. Stakeholder analysis and communication skills, as well as the ability to manage group dynamics, become prime assets. Non-partisanship is a must, yet at times even objectivity may be challenged, e.g. when the empowerment of disadvantaged groups is felt to be a prerequisite for their full participation in joint learning. Enhancing learning will eventually mean a fundamental restructuring of the training, methodological baggage, professional skills and outlook of evaluators.

Can learning help to mainstream the internalization of evaluation results?

At first sight, pushing evaluation into a learning mode looks like an easy business. However, it leads more or less automatically to the next level: adaptive management, namely a systematic and continuous effort to adjust individual and organizational behaviour in accordance with the lessons of past experience. Consequently, agencies that are prone to promote learning by others will be drawn irresistibly into asking questions about their own in-house learning capability. Hence, capacity development and the institutionalization of internal learning processes emerge as the logical complements to systematic efforts to improve the internalization of evaluation results.

How can we create the conditions for effective learning?

Different approaches emphasize different enabling conditions for learning. The latter generally include the following:
- the absence of threats to openness and the sharing of opinions;
- curiosity and motivation on the part of the participants;
- the availability of intellectual and/or practical challenges, incentives and institutional support; and
- opportunities for practical follow-up.

In addition, Irene Guijt, Jim Woodhill and colleagues (2002) in their paper on the role of e-networks underscore the need for a systematic and explicit learning process and awareness of the minimum conditions that need to be met to make joint learning effective.

Some lessons

These conclusions may not be new, but they do underline the need to strengthen the effective uptake of evaluation results by means of learning.

Evaluation policy

Improving the internalization of evaluation results implies a shift in perspective with respect to the way in which development processes, and indeed development institutions, are managed. Adaptive management, namely flexibility and responsiveness to experience, must become a key institutional feature. As a consequence, institutional development is a necessary complement to learning by development agencies.

The balance between accountability and transparency on the one hand, and learning and adaptive management on the other, needs further scrutiny. At the same time, it is important to be aware of any power imbalances between donors and aid recipients. We shall need to develop ways and means of safeguarding the credibility of evaluators and development professionals engaged in learning through evaluation.

Evaluation practice

In terms of methodologies, what is needed is a pragmatic, process-oriented approach rather than a dogmatic, content-oriented one. There is no single set of tools that guarantees learning. Indeed, many different ones can be used if the principles governing successful learning are taken into account. These include mutual respect, inclusive thinking, a readiness to understand other people's perspectives, and a willingness to take criticism seriously and to revise institutional cultures, rules and procedures.

Complexity, unpredictability and uncertainty are regular features of multi-actor learning processes. Irene Guijt (2000) argues that methodological issues 'extend far beyond simply which method works best, as these are just a small part of the extensive communication processes that lie at the heart of monitoring and evaluation.'

Evaluation professionals

The need to rethink the professional profile of evaluators is paramount if they are to facilitate learning. Apart from having excellent research and communication skills, the evaluators of the future will also need to possess process facilitation, conflict resolution and negotiating skills. Analytical skills will have to include tools for situation and stakeholder analysis, as well as for understanding conflict and group dynamics.

Evaluation research

Far more emphasis needs to be placed on developing effective conceptual and practical tools. So far, research has hardly touched upon the systematic assessment of the learning effects of different monitoring and evaluation approaches, tools and techniques.

South-North and South-South exchange

The flow of lessons on practices and tools from the South to the North, and between Southern actors, is a field of inquiry in which much can be learned from systematic collaboration among evaluation researchers and practitioners.

Applying these lessons: the Pelican Initiative

In an effort to apply some of the lessons on the need for greater learning from evaluation, ECDPM, together with the International Development Research Centre (IDRC), Exchange, Bellanet and the East Africa Regional Office of the United Nations Children's Fund (UNICEF), started the Pelican Initiative, the Platform for Evidence-based Learning & Communications for Social Change, in January 2005. The Pelican Initiative focuses on the central question:

> *How can we learn more from what we do while at the same time having the biggest possible impact on the social change processes in which we engage?*

This platform (www.dgroups.org/groups/pelican) brings together 127 development practitioners, specialists and policymakers to explore this question, share experiences, and to push the agenda further on three perspectives discussed above:

- evidence and learning for policy change;
- learning in organizations and among partners; and
- society-wide learning among a multitude of stakeholders.

References

Cooke, B. and U. Kothari (2001) *Participation: the new tyranny?* London: Zed Books

Engel, P.G.H., A. Hoeberichts and L. Umans (2001) Accommodating multiple interests in local forest management: a focus on facilitation, actors and practices. *International Journal of Agricultural Resources, Governance and Ecology* Vol. 1

Guijt, I. (2000) Methodological issues in participatory monitoring and evaluation. In: J. Blauert, D. Campilan, J. Gaventa, J. Gonsalves, I. Guijt, D. Johnson and R. Ricafort (eds) *Learning from change: issues and experiences in participatory monitoring and evaluation*. London: Intermediate Technology Publications

Guijt, I., J. Woodhill with J. Berdegué and I. Visser (2002) *Learning through e-networks and related M&E issues*. FIDAMERICA/Grupo Chorlaví

King, C. and J. Jiggins (2002) A systemic model and theory for facilitating social learning. In: C. Leeuwis and R. Pyburn (eds) *Wheelbarrows full of frogs: social learning in rural resource management*. Assen: Van Gorcum

OECD (2001) *Evaluation Feedback for Effective Learning and Accountability*. Paris: OECD

Röling, N.R. (2002) Beyond the aggregation of individual preferences: moving from multiple to distributed cognition in resource dilemmas. In: C. Leeuwis and R. Pyburn (eds) *Wheelbarrows full of frogs: social learning in rural resource management*. Assen: Van Gorcum

Sutton, R. (1999) The policy process: an overview. *ODI Working Paper* No. 118. London: Overseas Development Institute (ODI)

Uphoff, N. and J. Combs (2001) *Some things can't be true but are: rice, rickets and what else? Unlearning conventional wisdoms to remove paradigm blockages*. New York: Cornell International Institute for Food, Agriculture and Development, Cornell University

Van der Veen, R.G.W. (2000) Learning natural resource management. In: I. Guijt, J.A. Berdegué and M. Loevinsohn (eds) *Deepening the basis of rural resource management*. The Hague/Santiago de Chile: ISNAR/RIMISP

John Roberts

7 Managing aid for results: how to select your partner in development

[Development] agencies should be evaluated on the extent to which they have allocated their resources, both of finance and knowledge, to stimulate the policy reform and institutional changes that improve peoples' lives.
Joseph Stiglitz

'Achieving results' is the current buzzword in the development community. But what do we mean by developmental effectiveness? Is it the same as aid effectiveness? This turns on whose version of results we are focusing on – the donors' or developing countries' version – and whether they are the same or different. There is a strong and enduring school of thought that persists in believing that aid will only be effective if provided selectively to countries whose policies and institutions are known to be sound, and that donors should be highly selective in their choice of country partners.

Millennium Development Goals

In a special session of its General Assembly in 2000, the United Nations (UN) committed itself to achieve the Millennium Development Goals (MDGs) for poverty reduction – with their 8 objectives, 18 targets and 48 indicators. The UN has proclaimed targets, as for its several Development Decades, on earlier occasions. The practice had become tired and discredited. But, in its new form, it has acquired a renewed commitment and vigour. The MDGs have become the loadstone and guiding light of the whole donor community, including the UN and other multilateral and bilateral donors. Progress towards them is minutely tracked by the United Nations Development Programme (UNDP) and the World Bank, and statistical capabilities in developing countries are being reinforced to provide more accurate and timely data.

The World Bank's Development Committee met in September 2002 to discuss better measuring, monitoring and managing for development results. At this meeting, it was decided that the Bank will henceforth produce Annual Reports not only on 'Portfolio Performance', but also on development 'Results'. The Development Assistance Committee (DAC) of the Organization for Economic Cooperation and Development (OECD) convened a Partnership Forum in December 2002 on the closely related themes of 'Managing for development results and aid effectiveness'. The conclusion was that DAC member countries should re-dedicate themselves to making their aid developmentally effective, and should not allow their domestic agendas and accountability procedures to stand in the way.

Poverty reduction strategies

The MDGs cannot be mechanistically translated into country-level programmes. They do not necessarily reflect individual developing countries' needs and priorities. Fortunately, a growing number of low income developing countries are drawing up poverty reduction strategies which they agree with their donors and which form the basis of joint assistance strategies. Some 45 countries now have final Poverty Reduction Strategy Papers (PRSPs), and a further 11 have interim PRSPs. These spell out country poverty reduction objectives and targets, and the means of achieving them, at least over the medium term. They go in the broad direction of the MDGs, with emphasis on pro-poor growth and wider access to social services. The major multilateral and bilateral donors have committed themselves to allowing their assistance programmes to be guided by these strategies, even if they harbour doubts about their realism.

The PRSPs provide a template of development effectiveness at the country level. But will aid provided in support of countries' poverty reduction strategies *ipso facto* be effective in helping to achieve development results? Many have doubts about this. They point out that PRSPs are of variable quality, are variously 'owned' by the countries which produce them, and that they are backed by policies and programmes of variable credibility and proven effectiveness. There is a strong and enduring school of thought that persists in believing that aid will only be effective if provided selectively to countries whose policies and institutions are known to be 'sound'.

Aid effectiveness means selectivity?

The selectivity school's thinking has long antecedents, but it has blossomed particularly since the end of the Cold War when the straightjacket of power politics in aid allocation decisions was loosened. Donor agencies were freed to make themselves domestically accountable by demonstrating the effectiveness of their aid through the development results that evaluation studies plausibly attribute to aid actions. Selectivity received further impetus from the growing belief that *ex ante* policy and institutional conditionality, long sought by the International Monetary Fund (IMF) and the World Bank from borrowers of their structural adjustment facility and credits, could not be relied on to create the conditions in which aid would be effective. By the end of the 1990s, there was growing conviction among the staff and in the boards of the international financial institutions that the number of conditions should be reduced to a minimum, and that assistance should be directed first and foremost to countries which had prepared the ground *ex post* for its effective absorption.

It has long been the practice of international development assistance (IDA) to allocate its resources to eligible countries on the basis of performance and policy, as well as need (poverty). This requirement became more formalized in the 1990s with the use of Country Policy and Institutional Assessments (CPIAs) which assess performance on the basis of 20 indicators. CPIA indicators are grouped under four headings: economic management – fiscal, monetary, debt and expenditure sustainability; structural policies – external trade, domestic product and financial markets, and the private sector; social inclusion – poverty focus of public policies, human resource development, and gender; and institutions and public sector management – rule of law, property rights, public expenditure management and accountability. IDA allocations for well-performing countries by these criteria have gradually been increased at the expense of the less well-performing ones. The European Commission and certain bilateral donors use similar allocation criteria, though with less sophistication and less rigour.

The rationale for performance-based allocations was given a fillip by research evidence from Craig Burnside and David Dollar, two World Bank economists, showing that aid was really only effective in promoting faster economic growth if accompanied by sound policies (2004). Sound policies consisted of avoiding macro-economic imbalances and currency overvaluation, the use of market mechanisms, openness to trade, and an enabling environment for private investors. In the absence of these, aid was ineffective.

David Dollar, this time with Paul Collier, has also argued plausibly that economies grow better if soundly managed and if they are not subject to major macro-economic and regulatory uncertainties (1999). In these circumstances, aid is more productive. Dollar and Collier went on to show that aid can be most effective in reducing poverty – given assumptions about the poverty elasticity of economic growth – if allocated to poor countries with sound policies. The numbers of people in poverty could decline dramatically if aid were redirected to 'high impact' countries. Jonathan Beynon found, however, in his work on policy implications for the British Department for International Development (DFID) that aid can be effective in helping the poor, even with unsound policies (Beynon 1999). World Bank conclusions about the poverty-reduction benefits to be gained from reallocating aid according to the Collier-Dollar formula were very sensitive to model specification and selection of data (Beynon 2003).

Controversy has raged among economists about the Burnside-Dollar findings which few other researchers have been able to replicate. Some later studies show that aid has been effective anyway, without reference to policies (Hansen and Tarp 2001; Gomanee et al 2002). Some very recent empirical work casts renewed doubt on the efficacy of aid, while confirming the importance of policy (Rajan and Subramanian 2005). A third group of researchers has found that aid devoted to developing infrastructure and economic services which yield early economic benefits has a strong effect on growth, but aid for the social sectors with benefits more remote in time has had a weaker effect (Clements et al 2004).

Thus, empirical evidence from cross-country, panel data, studies – where the performance of many different countries are averaged – yields ambiguous results on the aid-policy-growth relationship. But it does not follow that no relationship exists at the level of individual countries. Dani Rodrik has argued forcefully that it is now time to revert to country-level analysis to understand the intricacies of how institutions, policies and external financing interact and impact on the real economy. His view of the growth process lays prime emphasis on the quality of public institutions in developing countries, including transparency, the rule of law and expropriation risk (Rodrik et al 2002). The dividing line between policies and institutions is indistinct: good institutions are the product of past reforms.

Selectivity survives, therefore, at least as an ideal, but to implement it properly requires much sensitive institutional and political analysis of patrimonial and clientelist relationships, and their impact on resource appropriation and allocation, entrepreneurship, and the mechanics of a developmental state (Van der Walle 2005). Unfortunately, selection for institutional quality and development commitment, using aid as an incentive, may lead to the exclusion from volume aid allocation of countries which remain the home of many poor people.

The Millennium Project: money and capacity building surmount all obstacles

A quite different approach to aid effectiveness, built on institutional optimism and a degree of policy agnosticism, underlies the Millennium Project, in which Jeffrey Sachs has played the role of inspiration and guide (Sachs 2005). The Millennium Project sets out to accelerate progress towards the MDGs by means of a major increase in aid, sufficient to achieve the 'big push' needed to enable poor countries to break out of their 'poverty traps' where low savings, poor infrastructure, low levels of human development, poor health and poor government combine to prevent per capita income growth and poverty reduction. The Project recognizes that governments in many poor countries have technical and institutional weaknesses, and that some, initially, will be unable to absorb productively radically scaled-up volumes of assistance. For them, the immediate task is accelerated capacity building. After five years or so these countries, too, should be ready to embark on the 'big push' for poverty reduction.

So, for the Millennium Project, all poor countries are selectable, and none, apart from conflict countries, are debarred by history or institutional quality. But there is a technical question of how long it will take some of them to reach the starting block?

To de-emphasize or de-select?

The radical implication drawn from this recent evidence by some economists and right-wing donor country politicians is that poor performers should be sanctioned not by reduced allocations of aid (as with the IDA allocation formula), but by a cessation of significant aid transfers. Some have advocated 'competitions' between developing countries for the available aid in which the winners are those countries with the best past and prospective policy and institutional track records.

In his controversial book on growth, William Easterly (2001) demonstrates the theoretical fallacies and practical ineffectiveness of the traditional donor nostrums of aid-for-investment, aid-for-education, aid-for population, aid-for-debt reduction and aid-for-structural adjustment. He goes on to advocate publicly visible ‘aid contests’ in which each government vies for loans from a common pool on the basis of its track record and its credibility and publicly stated intentions. Easterly’s contention is that countries only grow and achieve poverty reduction if they have policies and institutions that offer incentives to invest in knowledge and physical and human capital accumulation. Competition for aid will be an incentive to policymakers to improve the domestic incentive environment.

The US Government launched its Millennium Challenge Account (MCA) at the time of the Financing for Development summit in Monterrey in March 2002. The MCA is explicitly based on the Easterly principle of competition: it is allocated only to countries exhibiting strong commitment to good governance, to investment in people (education and health) and to sound economic policies fostering open markets, sustainable budgets, enterprise and entrepreneurship. To find the winners in its annual competition, the US Treasury marks eligible countries on the basis of 16 publicly available indicators of performance. The chosen beneficiary countries will have to ‘pass’ at least 8 of the 16 tests. The field of eligibility will initially be restricted to IDA-eligible countries, but will later be enlarged to all lower-middle income countries, thus making access for poor countries more competitive. Marking countries’ exam scripts has proved to be time consuming. By July 2005, of the 17 countries judged to have passed the test only four (Cape Verde, Honduras, Madagascar and Nicaragua) had MCA aid agreements.

Another radical suggestion for dividing low income developing countries into sheep and goats in the interest of greater aid effectiveness emerged in early 2002 from the work of a task force in the World Bank led by Ngozi Okonjo-Iweala and Paul Collier, Director of Research. This group used selected low CPIA scores to identify a group of poorly performing countries that they dubbed Low Income Countries under Stress (LICUS). These were countries where conventional aid transfers (grants, IDA credits) had been and were likely to be ineffective. There was little point in persisting with such transfers until these countries changed their policies and their institutions were reformed. Instead, and pending reform, these countries should be offered intensified capacity building and advisory support.

A paper advocating selectivity along these lines was discussed in the World Bank’s Development Committee in April 2002, but was not agreed by the Bank’s governors. The approach was felt to be altogether too categorical and mechanistic, and to be insensitive to extenuating circumstances and to directions and rates of change, as opposed to levels in policy and institutional performance. In revising their work, the LICUS task force amended its conclusion to recommend that poor performance should be thought of as a continuum, and that countries in the continuum should be treated differently, according to their specific characteristics.

But who are the poor performers?

The debate among donors about what to do with poor performers remains open and unresolved. There is no clear consensus on practice, even if there is general agreement that the twin criteria of need and performance are relevant to aid allocation. The waters are further muddied by the multiplicity of criteria used by analysts from different academic disciplines to characterize developing country performance.

Performance can be judged in terms of outcome, or in terms of process. Outcomes typically considered may be economic, such as GDP growth, reduction in the proportion of people with <1 USD/day per capita incomes, or political/ institutional: the extent of human and political rights, rule of law and access to justice, equity in access to public services, absence of conflict and insecurity for persons and property, corruption, and political instability. The LICUS countries are identified by indicators of process performance: macro-economic and structural policies; poverty orientation; and quality and accountability of public expenditure management etc.

Lists of poor performers based on these criteria overlap. This is to be expected, given the evidence that institutions are important in determining economic growth and income poverty reduction. However, they do not coincide. Some lists are longer than others:

- There are some 47 low income countries whose rate of per capita income growth over the 1990s was less than 1.5%; 44 of these have <1 USD/day rates of poverty over 30%. None of these countries can be said to have performed well in reducing income poverty. In most of them, the headcount rate of poverty has increased.
- DFID has compiled a list of 46 countries, both low- and middle-income, whose CPIA score has fallen into the bottom two quintiles since 1999 and which it considers as potentially ‘fragile’ on grounds of challenges to their sovereignty, physical security, political stability, policy coherence, administrative capacity or commitment to pro-poor policies (DFID 2005).
- The World Bank’s task force on LICUS identifies 17 ‘core’ LICUSs (whose CPIA scores on service delivery, policy management and responsiveness to citizens all fall below 3 on a scale from 1 to 6), and a further 11 others which also come within the frame (because their average CPIA scores are less than 3). These countries’ processes, in the view of the task force, make it unlikely that aid will be very effective in reducing poverty.
- A smaller number of countries are dubbed ‘failed states’ by political scientists. Failed states are ones where governments have become seriously dysfunctional because of large-scale official corruption and predation; the

breakdown of public services and of domestic law, order and security; civil war; or involvement in external conflict and insecurity.
- Nicholas van der Walle (2005) has isolated a group of 26 non-conflict-affected 'stagnant, low-income, states' to which he has devoted a new book on political science analysis, emphasizing the persistence of patrimonial and clientelistic relationships of power and patronage, and their effects on resource allocation and the effectiveness of development expenditure.

Poor performance is multi-faceted, unpredictable and elusive. Lists compiled on the basis of historical outcomes do not coincide with those compiled on the basis of judgements about policy or institutional processes. Whatever the criteria used, the barometer of performance is subject to swift and major gyrations. Some countries in each of these groups have managed to improve poor people's access to basic education, health, water and communication services by refocusing their strategic priorities and improving public sector management. Some slip into one or more states of poor performance, but have managed to emerge and return to their previous dynamism and prosperity, thanks to committed leadership, strengthened by donors' helping hand (Uganda, Ghana, Cambodia and, latterly, Pakistan). Others have exhibited poor performance characteristics for a long time and are unlikely to emerge without a policy or political revolution of some kind (Congo, Somalia). Yet others, having performed poorly for a long time, may now be reverting to better performance (Sierra Leone, Kenya, Nigeria). Recovery from economic collapse, induced by strife or other causes, may be remarkably swift (Roberts 2004).

Selectivity de-selected

The imprecision and judgemental character of the notion of 'poor performance', and the variety of the depth and breadth of its manifestations, cast serious doubt on 'sheep-and-goats' selectivity as a sound basis for designing development cooperation programmes. Past performance is no reliable guide to the future. Wisdom after the event may make agencies regret past allocation mistakes, but it is no sound basis for making future allocations of assured development effectiveness.

Donors' country aid strategies, it now appears, need to be more subtly tailored to local needs, policies, economic and political circumstances, and administrative absorptive capacity, and to be based on perceptive comparative readings of each of these dimensions. Allocation decisions should be the result of inter-country and in-country strategic assessments in which the formulaic application of performance criteria – based on past evidence – is but one element. Other factors, such as political turning points and the strengthening of the leadership's commitment to poverty reduction strategies are equally germane.

Donors' concern about aid effectiveness, which reached its height in the late 1990s, now seems, in the light of more recent events, to be an introspective and self-regarding agenda. Donors were judging themselves by the light of their own, often *ad hoc*, criteria of what would play well with domestic opinion.

The new agenda of development effectiveness takes as its global goals, the results defined in the MDGs; as its country-level goals, objectives enshrined in countries' own poverty reduction strategies; as its principles, local strategic initiative, policy ownership and administrative control; and as its criteria, the achievement of partner-defined development results. Its premise is measured optimism about governance, institutional functioning, policies and managerial effectiveness, and a disinclination to be rigorously selective on grounds of past performance.

However, the Millennium Project and the 2005 exhortation of G8 leaders to increase development assistance by 50 billion USD has opened a debate about whether the swift scaling-up of aid-financed development initiatives will induce worsening standards of governance and a recrudescence of the process characteristics of poor performance. A higher volume of aid may well be globally more effective, but its efficiency, and that of beneficiary countries' own resources may decline. We do not know what will happen. We should however follow the sensible advice of Van der Walle that development partners pursue behaviours and construct incentive mechanisms which limit, as far as possible, adverse effects on efficiency and governance.

References

Beynon, J. (1999) *Re-targeting DFID aid for poverty reduction*. London: DFID

Beynon, J. (2003) Poverty efficient aid allocations: Collier-Dollar revisited. ESAU *Working Paper* No. 2. http://www.odi.org.uk/esau/publications/index.html

Burnside, C. and D. Dollar (2004) Aid, policies, and growth: revisiting the evidence. *World Bank Policy Research Working Paper* No. 3251. http://ssrn.com/abstract=610292

Clements, M., S. Radelet and B. Rikhil (2004) Counting chickens when they hatch: the short-term effects of aid on growth. *Centre for Global Development Working Paper* No. 44

DFID (2005) *Why we need to work more effectively in fragile states*. London: DFID. http://www.dfid.gov.uk/pubs/files/fragilestates-paper.pdf

Dollar, D. and P. Collier (1999) Can the world cut poverty in half? How policy reform and effective aid can meet International Development Goals. *World Bank Policy Research Working Paper* No. 2403. http://ssrn.com/abstract=630775

Easterly, W. (2001) *The elusive quest for growth: economists' adventures and misadventures in the Tropics*. Cambridge: MIT Press

Gomanee, K., S. Girma and O. Morrissey (2002) Aid, investment and growth in Sub-Saharan Africa. *CREDIT Research Paper* No. 02/05. Nottingham: University of Nottingham

Hansen, H. and F. Tarp (2001) Aid and growth regressions. *Journal of Development Economics* 64(2)

Rajan, R. and A. Subramanian (2005) Aid and growth: what does the cross-country evidence really show? *IMF Working Paper* No. 127. Washington DC: IMF

Roberts, J. (2004) Recovery from economic collapse: insight from input-output models and the special case of a collapsed oil producer. *ESAU Working Paper* No. 6

Rodrik, D., A. Subramanian and T. Francesco (2002) Institutions rule: the primacy of institutions over geography and integration in economic development. *CID Working Paper* No. 97. Harvard: Centre for International Development.Sachs J., J. McArthur, G. Schmidt-

Traub, M. Kruk, C. Bahadur, M. Faye and G. McCord (2004) Ending Africa's poverty trap. *Brookings Papers on Economic Activity* No. 1
Van der Walle, Nicholas (2005) 'Overcoming stagnation in aid-dependent states', Centre for Global Development, Washington DC. http://www.cgdev.org/Publications/?PubID=207

Robert Picciotto

8 The Monterrey challenge: re-shaping the development architecture

The elements of a new global partnership were forged at the United Nations (UN) Conference on 'Financing for Development', held in Monterrey, Mexico, during March 2002. The new global partnership calls for the adoption of improved policies and good governance in developing countries with the provision of increased aid and trade opportunities by developed countries. The Monterrey challenge combines ambitious objectives, a focus on results, and an unprecedented partnership between developed and developing countries in pursuit of poverty reduction. The Monterrey Conference also addressed the need to achieve greater coherence in development efforts through the Millennium Development Goals (MDGs). Endorsed by all 189 UN states in 2000, the MDGs symbolize a commitment to achieve results in terms of poverty reduction, education, health and environmental sustainability that can be supported by the rich and poor countries alike.

Implications for performance management

Basic changes in the structures, products, and processes of the aid industry are needed to enhance development effectiveness. Aid is often poorly coordinated and does not conform to the comparative advantage of partners. The efficiency of aid delivery has dropped as the number of donors has multiplied and aid volumes have declined. Harmonization of aid delivery mechanisms, fiduciary processes, and reporting requirements have lagged behind. Pool funding for sector-wide approaches is the exception rather than the rule. The resulting administrative burdens have mortgaged the skills needed to improve public sector effectiveness.

Implications for performance measurement

The new paradigm has four key implications for performance measurement. First, to capture programme results (i.e. outputs, outcomes, and impacts), development indicators should go beyond the measurement of inputs. Second, the primary unit of account for monitoring and evaluation should extend beyond the individual project to the country programme level. Third, monitoring indicators should allow tracking of progress toward the MDGs, as well as the intermediate objectives embedded in country programmes. Fourth, the performance of individual partners should be assessed in terms of their distinctive accountabilities and reciprocal obligations. For development assistance agencies, relevant performance indicators linked to these obligations should be reflected in corporate scorecards in line with results-based management principles.

In order to meet this demanding agenda, the practice of evaluation must overcome substantial challenges. At the project level, inadequate organizational incentives and/or low priority given to monitoring and evaluation have to be overcome. At the country level, emphasis should be given to public expenditure evaluations using logical frameworks, tracking surveys, and participatory methods with results chains used to link country programme outcomes with the MDGs. Evaluation capacity development addressing weaknesses in country monitoring and evaluation systems is central to these tasks. It has lagged behind so that relevant data are rarely collected, interpreted and used for decision-making. Last but not least, monitoring and evaluation has been largely absent at the global level.

Evaluation governance

The conceptual foundations for objective-based performance evaluation do exist. But the credibility of evaluation hinges in large part on the quality of its governance, i.e. on the set-up of professional and independent evaluation units that report to country legislatures, governing bodies or on *ad hoc* construction of independent verification mechanisms. Currently, this prerequisite of evaluation credibility is missing in most governments, companies, voluntary organizations and development agencies.

To achieve development effectiveness, evaluation processes should combine independent and self-evaluation mechanisms. At its best, evaluation is participatory and integrated in business processes, public administration and aid procedures. This helps to ensure that evaluation contributes to accountability, learning and adaptability. Specifically, self-evaluation should be built into all major corporate processes and transactions at the project, country and global levels, while independent evaluation attests to its rigour and quality. Currently, the fragmentation of evaluation products and the diversity of evaluation methods among donors have led to cumbersome reporting requirements and ‘evaluation bombardment’.

For aid evaluation as for aid delivery, the solution to current governance dilemmas lies in harmonization, coordination, and the forging of partnerships. More resources need to be devoted to nurturing of convergence in evaluation methods through networks of evaluators, development of evaluation capacities, professional associations (such as the International Development Evaluation Association), joint or parallel evaluations, and country-based evaluations connected to enhanced processes of public expenditure management.

Country assistance evaluation

The shift from project to country programmes as the unit of account for performance management and evaluation requires the methods designed to overcome three major methodological challenges: aid allocations; aggregation; and attribution. Given that development effectiveness hinges, in large part, on the quality of the enabling policy and institutional environment, performance-based aid allocations are critical. Such allocations must be informed by the results of evaluation. They require the regular compilation of policy and institutional indicators at the country level, optimally allowing for tracking of the MDGs.

A problem of aggregation arises because the quality of a country assistance programme is normally judged through quality assessments of individual projects. For validity, meta-evaluation techniques must be applied and rely on independently validated ratings of lending and non-lending operations – a rarely enforced 'golden rule'. Finally, the problem of attribution (assessing the contribution of various partners, initial conditions and exogenous factors) requires objective assessments of the compliance of development actors with agreed policies and procedures. Focus groups, client surveys, and advisory committees can help produce relevant assessments.

As the development architecture is reshaped, it will have to rely on 'pool funding' by stakeholders that provide independent oversight. Coherent monitoring and evaluation processes will have to be designed to ensure accountability and learning at the country and global levels. It is time to lay the foundations for a country-based evaluation architecture that embraces the UN system, the multilateral development banks and the bilateral aid system, as well as governments, civil society and the private sector so as to better meet the global poverty reduction challenge.

Meeting the challenge

The endorsement of the MDGs by the World Bank and other international development agencies has raised the bar for development performance management and evaluation. Sustained corporate actions are needed throughout the far-flung development evaluation system to:
- strengthen monitoring and evaluation at the project, country, sector, global and corporate levels;
- improve self-evaluation processes for country assistance and sector strategies;
- build evaluation capacity in developing countries; and
- create alliances aimed at evaluation excellence, independence, coordination and harmonization.

Howard White

9 The road to nowhere? Results-based management in international cooperation

Results-based management has become a fact of life for development agencies. They might hope to learn from the experience of the US Agency for International Development (USAID) which has already gone down this road. It is indeed instructive that USAID has come back up the road again saying 'there's nothing down there'. But development agencies have been rightly criticized in the past for paying too little attention to the final impact of their activities, so we would like to support a results-based approach. But we should not do so blindly when it suffers from the severe limitations outlined below. Serious attempts to link agency performance to developmental outcomes must rely upon the logical framework, also known as the logframe. The logframe is not a universal panacea but, used properly, can force agencies into a critical examination of the nature of their programmes and projects, and the results they achieve.

What is results-based management?

Results-based management is the creation of an incentive structure, including the allocation of budgetary resources, reflecting achievements as measured by outcomes. It may seem that all incentives should operate in this way but, in fact, management information systems have often focused on inputs and activities. Hence, for example, educational resources have traditionally been allocated based on student numbers. The philosophy of results-based management has led to changes in countries such as the Netherlands and the UK whereby a part of funding for tertiary education is now allocated according to research performance. In the UK, league tables are published of school exam results, implying that resources will follow pupils to the better schools. More recently, the possibility has been mooted of mortality rate tables for hospitals and even individual doctors.

Monitoring and evaluation for development activities has shared this input-orientation, with the success of a project being judged by the speed at which it gets through its budget. Even end-of-project evaluations have often shied away from the tricky issue of development impact. During the 1990s, governments in developed countries embraced results-based management, an embrace that took in aid agencies.

Results-based management in aid agencies

The longest track record with results-based management amongst aid agencies is that of USAID, following the 1993 Government Performance and Results Act (GPRA) which required all US government agencies to have outcome-based targets against which their performance could be measured. The Agency's experience is particularly instructive.

In 1997, USAID laid out six strategic development goals, such as 'broad-based economic growth and agricultural development encouraged', and for each of these defined a set of outcome indicators at both country and global levels (for example 'average annual growth rates in real per capita income above 1%'). With respect to the growth goal, the Financial Year (FY) 2000 performance report stated that:

> *Nearly 70% of USAID-assisted countries were growing at positive rates in the second half of the 1990s, compared with 45% in the early part of the decade.*
> (USAID 2000)

However, that same performance report noted that:

> *One cannot reasonably attribute overall country progress to USAID programmes.*

Commenting on the USAID performance report from 1999, the General Accounting Office (GAO) had similarly observed that the goals were:

> *so broad and progress affected by many factors other than USAID programmes, [that] the indicators cannot realistically serve as measures of the agency's specific efforts.*
> (GAO 2000)

In response to these criticisms, the FY 2000 performance report announced that the indicators related to the strategic goals would no longer be used to measure USAID's performance. However, they will still be reported as being of interest in their own right, being referred to as Development Performance Benchmarks. Rather, performance will be measured against the strategic objectives of the individual operating units, such as country programmes.

On the other hand, the British Department for International Development (DFID) is just starting to get to grips with these issues. All UK agencies are required to have a Public Service Agreement (PSA) and Service Delivery Agreement (SDA), the former containing outcome targets against which an agency's performance is to be judged and its funding determined. DFID's PSA measures are based on International Development Targets (IDTs), albeit with a much shorter time frame. For health, the performance targets are stated as improvements in child, maternal and reproductive health in the 'top ten' recipients of DFID health care assistance demonstrated by:

- a decrease in the average under-5 mortality rate from 132 per 1,000 live births in 1997 to 103 on the basis of data available in 2004;

- an increase in the proportion of births assisted by skilled attendants from a baseline established in 2000 of 43% to 50% on the basis of data available in 2004; and
- improved access to reproductive health care.

Given the widespread commitment to the IDTs, which is particularly strong in the UK, are such targets likely to lead to improved agency performance? This question is considered by looking at the desirable properties of performance indicators.

Desirable properties of performance measures

Performance measures have a number of desirable properties. Most of these can be dispensed with fairly quickly to focus on three key problems:

Relevance and balance

Performance measures, such as those embodied in the IDTs, are undoubtedly relevant. There are, however, two problems of balance. First, the measures may not span an agency's portfolio. For example, the health targets in DFID's PSA exclude health activities outside of the top ten recipients. And the targets as a whole do not apply to activities in middle-income countries. The second problem of balance is that, for reasons discussed below, it is better to have a set of indicators spanning the range of the logframe from inputs to outcomes.

Measures known, understood and trusted

Performance measures used in development are usually clearly defined and well understood, at least by those who work directly with them. There has been some tension between top-down setting of targets and a more participatory approach. This has also been a factor behind the promotion of the Millennium Development Goals (MDGs) as rivals to the IDTs. Data quality can be more problematic. The IDTs include maternal mortality, for which data are notoriously unreliable. The measure was included in DFID's first PSA but, when it came to reporting, DFID told the Treasury that data were unavailable. Country coverage for most indicators is patchy. But the more serious problem is the timeliness of the data, as discussed below.

Affected and attributable

Changes in performance measures should be affected by the activities of the organization and the extent of the effect measurable (attributable). However, attribution becomes harder to establish as we move through the logframe from inputs to outcomes. As we saw above, USAID has abandoned claiming that changes in economic outcomes can be attributed to its work.

Achievable

Targets should be achievable, but not too easily. Most analyses suggest that the IDTs will not be met, although there is regional variation in performance. But this is not the same as saying that they could not be met if aid and government resources were better focused towards meeting them.

Linked to existing management systems

Organizations will already have in place management information systems. The rise of results-based measures was, in part, a response to the fact that existing systems focused on inputs and internal activities, such as spending and staffing, rather than achieving outcomes. But new results-oriented systems should not be separate or parallel to these existing systems. There are two further problems here: aggregation (can the various measures be added up to summarize performance for regions, sectors and the agency as a whole?) and alignment (do the data tell us about impact on the outcomes of interest?)

Three key problems

The above discussion identifies three problem areas. These problems are less severe (though present) when using results-based management at the project level, becoming more intense as we attempt to measure results at country level or for the whole portfolio.

Data availability

Data become available only with a time lag, this being particularly so for developmental outcome measures like infant mortality and income-poverty. DFID's PSA covering 2001-2004 includes amongst its targets 'a decrease in the average under-5 mortality rate from 132 per 1,000 live births in 1997 to 103 on the basis of data available in 2004'. Implicit here is the recognition that data available in 2004 will probably be for 2001, namely the start of the period of the current PSA. How can results of current strategies be judged on the basis of results achieved before the strategy was put in place?

Attribution

It is very difficult to attribute changes in developmental outcomes to the activities of an individual agency. This problem is encountered where agencies work together, as in sector programmes or providing budget support, or for agency performance at the country level. One agency's support to public sector reform in an African country (providing the funds to pay for retrenchment packages) had as the outcome indicator: maternal mortality. It is difficult to imagine what sensible use could have been made of maternal mortality data (even with no time lag) to judge the performance of this programme.

All donor agencies have some sort of monitoring and evaluation system at the project and programme level which should provide a basis for both feedback at the project level and 'feed-up' to management. But how do these systems tackle the problems of aggregation and alignment?

The World Bank's rating system is an example of a system yielding agency-wide results. These ratings include an overall rating if an activity has been satisfactory. Hence overall portfolio performance can be judged by the percentage of projects deemed satisfactory and 'problem areas' identified. But this information does not allow us to say anything about the contribution of the World Bank to achieving development outcomes as measured by, say, literacy or premature death. The information collected is simply not the right sort to provide that information.

There are two possible responses to this problem of misalignment. The first is to say that it is inevitable. The problem of attribution is not going to be solved for routine monitoring purposes, so no attempt should be made to link agency performance as measured by bottom up systems with agency performance as measured by achieving developmental results. That answer does not seem satisfactory for agencies, like DFID, which have pinned their performance to such results.

The second response to the misalignment problem is to resort to the logframe. The bridge must be made between observing satisfactory activities and presumed impact on development outcomes.

The case for the logframe

The US GRPA suggested that performance measures should span inputs, outputs and outcomes. However, in practice such a logframe based approach has been applied in few GRPA plans. The absence of indicators reflects the fact that there is too little conscious analysis as to how plans and programmes will affect outcomes. The attraction of the logframe is partly that it should force the agency to examine programmes to see if it really will achieve the desired outcomes. Used in this way, the logframe helps tackle the attribution problem.

Attribution becomes harder as we move along the causal chain. It is easy to attribute responsibility for delivering inputs, and usually for carrying out activities, although external factors may play a part. These activities should lead to desired outputs which deliver the target outcomes, again subject to external factors. If the underlying model is correct then indicators should capture whether the organization is doing what it needs to do to achieve the outcomes – which may well be the case even if the targets are not met. The various links in the logframe should, in principle, be scientifically-established, though in practice we rely upon 'plausible association'.

References and further reading

DAC (2000) *Results based management in the development co-operation agencies: a review of experience (Executive Summary)*. Paris: OECD

General Accounting Office (2000) *Observations on the US Agency for International Development's Fiscal Year 1999 Performance Report and Fiscal Years 2000 and 2001 Performance Plans*. Washington DC: GAO

National Audit Office (2002) *Performance management – helping to reduce world poverty*. London: The Stationery Office

USAID (2000) *FY 2000 Performance Review*. Washington D.C: USAID

White, H. (2004) Using development goals and targets for donor agency performance measurement. In: R. Black and H. White (eds) *Targeting development: critical perspectives on the Millennium Development Goals*. London: Routledge.

Geoff Barnard

10 The new frontier: evaluation and organizational learning

If it were so simple, our work would have been completed long ago.
(Tokyo Conference delegate, 2000)

How can development organizations get the most out of evaluations – both from an accountability perspective, and from the point of view of institutional learning? This was the focus of a report from the Network on Development Evaluation, then known as the Working Party on Aid Evaluation, of the Development Assistance Committee (DAC) of the Organization for Economic Cooperation and Development (OECD), entitled 'Evaluation feedback for effective learning and accountability' (2001). The report emerged from a workshop organized by the Japanese government in Tokyo on 26-28 September 2000, and highlights the concerns widespread within DAC agencies about the shortcomings of current feedback practices and the need to do much more to ensure that lessons from evaluations are not wasted.

Accountability has long been a core concern for evaluation units. Organizational learning is the area that most development agencies see as 'the new frontier' where the greatest opportunities and challenges exist. But the report underlines that there are no easy, one size fits all, solutions. Concepts of knowledge management and learning organizations, borrowed from the commercial sector, have made considerable inroads in development agency thinking since the late 1990s. The workshop considered how these concepts can be applied in an evaluation setting. The World Bank, one of the leaders in the application of the knowledge management concepts to development, contrasts three types of learning: learning from 'doing', learning from research and learning from evaluation. The Bank also argues that effective mechanisms need to be in place to allow effective learning from sources outside the institution as well as within it.

The report identified a number of obstacles that get in the way of learning:
- organizational culture;
- pressure to meet disbursement targets;
- lack of incentives to learn;
- tunnel vision;
- loss of institutional memory;
- insecurity and the pace of change; and
- the unequal nature of the aid relationship.

The new generation of Internet-based communication approaches, including websites and internal intranets, are seen as valuable additions to the evaluation toolkit which can facilitate learning. In recognition of this, the IOB Policy and Operations Evaluation Department, the evaluation arm of the Dutch Ministry of Foreign Affairs, has recently further developed its website. From a disclosure perspective, publishing evaluation reports on development agencies' public websites is seen as a cost-effective way of opening up access. It can also be a useful consultation tool. But these tools need to be used intelligently and are not a replacement for conventional communication approaches, particularly face-to-face dialogue.

Institutionalizing lesson learning

Influencing policy and practice is one of the key objectives of aid evaluation. The way it happens in practice is far from being an exact science. Agencies are very different in their culture and the way they work, and have developed a range of internal mechanisms that try to institutionalize the process of evaluations. The British Department for International Development (DFID) has introduced the concepts of Performance Scoring and further developed its Performance Reporting and Information System for Management. In the Swedish International Development Cooperation Agency (SIDA), the central evaluation and audit functions have been combined in an independent department under a joint director who is responsible to the Board. The World Bank continues to develop the concept of results-based management. It is also in the process of revizing its use of Management Action Records, a feedback loop whereby key action points emerging from an evaluation are identified. Despite these and other efforts to systematize the procedures involved in making sure evaluation lessons are learned, there remains a substantial element of chance in the process. In this respect, evaluation is similar to development research which also seeks to influence policy and practice but is equally susceptible to factors outside the immediate control of the researchers themselves.

A review of the evaluation capacities of multilateral organizations, undertaken by AusAid (1998), compared the evaluation and monitoring capacities of five multilateral organizations supported by the Australian aid programme: the World Bank, the Asian Development Bank, the United Nations Development Programme (UNDP), the United Nations High Commissioner for Refugees and the World Food Programme. It concluded that the two Banks, and the UNDP conduct effective evaluation on their own performance, the latter primarily at the central level. It argued that:

> *The learning purpose of evaluation is important, as the lessons learned from experience, both from ongoing and completed operations, should be used continuously to improve the quality of project design and portfolio performance. Evaluation is a means to ensure the continued relevance of operations to emerging problems and potential solutions.*

Improving organizational learning

According to Elliot Berg of the University of Auvergne, France, organizations responsible for international development have been slow learners. He argues:

> *Despite its ambiguities, it is convenient to use the term "organizational learning" as a shorthand for the process by which organizations obtain and use knowledge to adapt old policies, programmes and strategies, or to innovate more broadly.*

Berg argues that formal evaluation should be a major instrument of organizational change. It should uncover failure and induce institutional response. He considers:

> *There is a widespread agreement that in development assistance, evaluation has not performed these functions well.*

One of the main problems pinpointed in Robert Cassen's study 'Does aid work?' (1986) was that aid organizations do not seem to learn from their mistakes. The recent Swedish Foreign Ministry report, 'Organizational learning in development cooperation: how knowledge is generated' and used by Forss, Cracknell and Stromquist, arrives at the same although more nuanced conclusion: that aid agencies learn 'slowly and cautiously'.

Jerker Carlsson (2001) asks if evaluations are used for learning purposes. In their 1992 study of Norwegian aid organizations, Samset, Forss and Hauglin showed that evaluations are only one source of learning for people working in the aid administration. They found that there are other sources that aid agency officials rank much higher. Carlsson argues that evaluations are very donor-focused and that a majority of local stakeholders do not find much value in evaluations at all. He makes four points regarding improving the usefulness of evaluations for learning purposes:
- The intended users must be identified at the beginning of an evaluation process and the evaluation should be planned and designed with utilization in mind.
- Most stakeholders are used as providers of information and are rarely involved in the whole evaluation process.
- Recommendations, produced as part of an evaluation, should make clear who should react to the recommendation, and should be adapted to the situation of a local organization.
- Results should be distributed to a wider audience among the stakeholders.

The relatively new approach of participatory monitoring and evaluation, covered in an article by Sophie de Camâra in this publication, represents a methodology which may be able to overcome some of these shortcomings.

References

Australian Agency for International Development (1998) Review of the evaluation capacities of multilateral organizations. *Evaluation* No. 11. http://www.ausaid.gov.au/publications/pdf/qas11.pdf

Berg, Elliot (2001) Why aren't aid organizations better learners? In: J. Carlsson and L. Wohlgemuth (eds) *Learning in development co-operation*, Stockholm: Almqvist & Wiksell International. http://www.egdi.gov.se/pdf/berg.pdf

Carlsson, Jerker (2001) Learning from evaluations. In: J. Carlsson and L. Wohlgemuth (eds) *Learning in development co-operation*, Stockholm: Almqvist & Wiksell International, pp. 120-129

Cassen, R. (1986) Does aid work? Oxford: Oxford University Press, 336 pp

DAC Working Party on Aid Evaluation (2001) Evaluation feedback for effective learning and accountability. *Evaluation Effectiveness* No. 5, 118 pp. http://www.oecd.org/dataoecd/10/29/2667326.pdf

Forss, K., B. Cracknell and N. Stromquist (1998) Organizational learning in development co-operation: how knowledge is generated and used. *EGDI Working Paper* No. 3

Samset, K., K. Forss and O. Hauglin (1992) *Learning from experience – a study of the feedback from the evaluation system in the Norwegian Aid Administration*. Ministry for Foreign Affairs: Oslo

Sophie da Câmara Santa Clara Gomes

11 Participatory approaches to monitoring and evaluation

The principles to include a monitoring mechanism or to carry out an evaluation are usually commonly agreed. The need for openness and inclusiveness, however, is a slower process and has not yet reached the status of common practice.

The reasons why a monitoring and evaluation (M&E) system is set up are multiple, and evolve in time. Initially, there was accountability, and then the need to make use of lessons learned and to progressively improve and adapt interventions. Participation, ownership and capacity building came later. The uses of M&E are illustrated by a list from the British Department for International Development (DFID):

- improve performance;
- improve impact;
- enhance learning;
- ensure sustainability; and
- to provide accountability.

Having a joint M&E exercise allows legitimate recipients and beneficiaries to have a stake in the process. In the best cases, their contribution could allow a development intervention (programme, project, etc.) to, at least, take their needs into account in a future project cycle. Aside from adapting the practice to the actual context, a joint exercise also brings in democratic management of development aid, as well as a direct process of capacity building.

Democratizing aid management becomes particularly relevant when the development intervention aims to enhance democratic structures, participatory society models, or to help empower local communities. A very large part of development aid nowadays comprises an element of empowerment, ownership, transparency or participation, usually directed at local communities, women or other target groups. Bringing about those changes through a democratic process would, if anything, serve as an example of inclusiveness and transparency.

Operational challenges

Once the principle of why to do it is agreed, there remain two key challenges in terms of operationalization. The first is how to actually include the recipient. At what level can the beneficiaries-recipients be involved in the actual M&E

exercise? And through which mechanisms? Second, how can one measure the social changes induced by the development intervention? How do we measure intangibles?

Enhancing participation

The essence of the first operational challenge revolves around the question of to whom development practitioners consider themselves to be accountable. If the answer is only to their donors, then there is no great need to include the beneficiaries. If, on the other hand, the intervening agency considers itself to also be accountable to its recipients-partners-beneficiaries, then a participatory process is essential.

Generally speaking, there is a general recognition of the need to involve the beneficiaries, and there are specific methods for each stage of the intervention, from needs assessment, planning and programming, to monitoring and finally evaluation. The most inclusive systems bring recipients and partners on board at early stages of project planning, based on the premise that it makes no sense to involve them in the M&E process if the planning itself was not already participatory. Such project-level thinking is today being echoed in grand schemes like the Cotonou Partnership Agreement – signed by the European Union and 77 African, Caribbean, and Pacific (ACP) countries – where there are calls for cooperative resources to be jointly programmed and jointly monitored, and even implemented through joint action approaches.

But is participation always relevant? Certain programmes may not require participation if they are too technical or completely straightforward. Still, in the case of social or human development programmes, in particular in small or wounded communities after a conflict for instance, including the beneficiaries as early as possible in the process of needs assessment and planning can make the difference between failure and success. Generating ownership is certainly the first step towards a sustainable process. This ownership stems from the feeling of being part of the exercise.

Many instruments can be used in this process to collect large numbers of opinions or to assess the needs of heterogeneous communities. They range from participatory rural appraisal (PRA), logical frameworks, stakeholder analysis, focus groups, and interviews. On the whole, while there is a strong trend to better define the role and place of beneficiaries at every stage of the development intervention, these experiences are not yet widespread and commonplace. Some examples of participatory approaches can be found and there is a large literature on lessons learned. These include the *Programa de Desarrollo para Desplazados, Refugiados y Repatriados* (PRODERE) experience in Central America; the literature of the War-Torn Societies Project; publications of the Overseas Development Institute (ODI) on the impact of the development projects of non-governmental organizations (NGOs); and, of course, publications of the International NGO Training and Research Centre (INTRAC).

Measuring intangibles

The second operational challenge is to identify an appropriate set of instruments that could be used to measure the effectiveness of development cooperation in the areas of empowerment, democratization and participation. Which indicators can bear witness to changes as intangible as the social and political changes resulting from a development intervention, like, for instance: empowerment, human rights, awareness raising, and participation?

The trickiest aspect is to assess the impact and effect of the intervention alone, in the midst of all the changes happening in the society, and especially in fast evolving situations such as those found in transition societies or in conflict affected communities. Many instruments have been proposed by the INTRAC experts. The common objective is to grasp qualitative data, and turn it into a positive indicator of change in order to make it measurable. One of the key lessons is a technique to define tailor-made indicators that might provide useful information, that could form a sound basis for decision-making, and can also be verified.

Searching for cases

At the European Centre for Development Policy Management (ECDPM), we have a project called Joint Monitoring of Performance that looks at the new EU Partnership Agreement with the ACP States in which the principle of performance-based aid allocation has been introduced. In this case, performance is to be assessed according to a set of merit criteria. The idea is to reward states that perform well, in compliance with the human rights, democratization and governance objectives of the Convention. Any partner showing signs of good performance (according to the agreed merit criteria) would then be further encouraged via an increased amount of aid.

Although performance evaluation is much more complex than project evaluation, the actual techniques of participatory monitoring are likely to be comparable, and the underlying participatory principles the same. The success of a performance-based partnership will depend, to a large extent, on the way it is put into practice. Performance-based partnership is preferable only insofar as it may help to transform the current, largely inefficient system of donor-imposed conditionalities into a locally owned set of objectives and performance criteria. Participation and ownership are essential elements of such an approach. Stakeholders that were consulted on the issue insisted on the need to organize performance monitoring in a transparent, decentralized and participatory manner.

So far the project has produced two discussion papers and a *Policy Management Brief* (Bossuyt and De Jong 1997). The most important contribution is a set of practical options and suggestions for the possible operationalization of a joint monitoring of EU aid. The *Policy Management Brief* presented the main findings of the African-based research and in-country consultations on the political acceptability and feasibility of performance criteria process, including an inventory of possible measures to allow for a smooth implementation of performance criteria. So far, neither side really has any clear answers and the search is on for cases and ideas that might illuminate the options.

Reference

Bossuyt, J. and K. de Jong (1997) Performance criteria in future ACP-EU cooperation. *Policy Management Brief* No. 8, Maastricht: ECDPM http://www.ecdpm.org/Web_ECDPM/Web/Content/Content.nsf/0/e0db1497b2a860a3c1256c6c0044f9eb

About the authors

Geoff Barnard is Head of Information at the Institute of Development Studies (IDS). He is also co-convenor of the Information Management Working Group of the European Association of Development Research and Training Institutes (EADI).
Address: IDS, University of Sussex, Brighton BN1 9RE, UK.
URL: www.ids.ac.uk
E-mail: g.barnard@ids.ac.uk

Rob van den Berg is currently director of the Global Environment Facility (GEF) Office of Monitoring and Evaluation in Washington DC, USA. He was previously Director of the IOB Policy and Operations Evaluation Department of the Ministry of Foreign Affairs, the Netherlands. He is also a past Chair of the Organization for Economic Cooperation and Development-Development Assistance Committee (OECD-DAC) Evaluation Working Group (2002-2004).
Address: GEF Office of Monitoring and Evaluation, 1818 H Street NW, Washington, DC 20433, USA.
URL: www.thegef.org
E-mail: rvandenberg@thegef.org

Sophie da Câmara Santa Clara Gomes is currently an expert on Central Africa at the Directorate-General for Development Cooperation (DGCD), Belgium. She was previously a Programme Officer at ECDPM.
Address: DGDC, Bureau 310, rue de Brederode 6, 1000 Bruxelles, Belgium.
URL: www.dgcd.be
E-mail: sophie.dacamara@diplobel.fed.be

Sarah Cummings is an Information Manager at the Royal Tropical Institute (KIT). She is also a member of the Advisory Council of Hivos and co-Chief Editor of the 'Knowledge Management for Development Journal.'
Address: Information and Library Services, KIT, PO Box 95001, 1090 AD Amsterdam, The Netherlands.
URL: www.kit.nl
E-mail: s.cummings@kit.nl

Paul Engel is Director of the European Centre for Development Policy Management (ECDPM). He previously worked as an international consultant and taught management, communication and organization for rural innovation at universities in the Netherlands and Chile.
Address: ECDPM, Onze Lieve Vrouweplein 21, 6211 HE Maastricht, The Netherlands.
URL: www.ecdpm.org
E-mail: pe@ecdpm.org

Henri Jorritsma is Deputy Director of IOB Policy and Operations Evaluation Department of the Ministry of Foreign Affairs, The Netherlands.
Address: IOB, Ministry of Foreign Affairs, P.O. Box 20061, 2500 EB Den Haag, The Netherlands.
URL: www.minbuza.nl
E-mail: hej.jorritsma@minbuza.nl

Charlotte Ørnemark is a freelance consultant in Kenya and an ECDPM Programme Associate specialized in development communication and multi-actor learning processes. She was previously a Programme Officer at ECDPM.
Address: ECDPM, Onze Lieve Vrouweplein 21, 6211 HE Maastricht, The Netherlands.
E-mail: ornemark@wananchi.com

Robert Picciotto is currently a Research Fellow at the University of Sussex, a Visiting Professor of King's College London and Director of the Global Policy Project, UK. He was previously Director General of the Operations Evaluation Department (OED) of the World Bank.
Address: Global Policy Project, 21 Palace Gate # 5, London W85LS, UK.
URL: www.globalpolicyproject.org
E-mail: r.picciotto@btinternet.com

Rob van Poelje is a Senior Advisor, Sustainable Economic Development at KIT. He has 18 years of experience working in Burundi, Guinea Conakry, Lesotho, Mali, the Netherlands and Thailand, both in bilateral and multilateral programmes, and as an independent consultant.
Address: Development, Policy and Practice, KIT, PO Box 95001, 1090 AD Amsterdam, The Netherlands.
URL: www.kit.nl
E-mail: r.v.poelje@kit.nl

John Roberts is now heading the Economic and Statistical Analysis Unit, a new research programme funded by the British Department for International Development (DFID), having previously run the Centre for Aid and Public Expenditure (CAPE). John still remains an Overseas Development Institute (ODI) Research Fellow.
Address: ODI, 111 Westminster Bridge Road, London SE1 7JD, UK.
URL: www.odi.org.uk
E-mail: j.roberts@odi.org.uk

Jurrien Toonen is Senior Adviser, Health at KIT. He is a medical doctor and public health specialist focusing on health systems development and health sector reform. He has 20 years' experience in public health, covering various areas of policy development and implementation.
Address: Development, Policy and Practice, KIT, PO Box 95001, 1090 AD Amsterdam, The Netherlands.
URL: www.kit.nl
E-mail: j.toonen@kit.nl

Howard White is a Fellow of IDS, currently on leave of absence to the OED of the World Bank where he is leading a programme on impact evaluations. He has carried out evaluation studies for many development agencies and published on evaluation issues in books and academic journals.
Address: OED, World Bank, 1818 H Street NW, Washington DC 20433, USA.
URL: www.worldbank.org/oed
E-mail: hwhite@worldbank.org

Chris Whitehouse, from the UK, is currently coordinator of the United Nations Volunteers (UNV) programme in Bhutan, and previously held the same post in Nepal. During his eight years with UNV, he took part in evaluations of UNV country programmes and projects in Papua New Guinea, Mongolia and Maldives, over and above regular in-country monitoring of volunteers in Nepal and Bhutan. He has work experience also in Zambia, Kenya, Singapore,Thailand and Turkey. He has a Bachelors in Psychology and Philosophy from the University of Oxford, and a Masters in Development Studies from the University of Bath.
E-mail: whitehousechris@yahoo.com

Thomas R.J. Winderl is a consultant for monitoring and evaluation. He holds a Ph.D. in Political Science from the University of Vienna and has previously worked in Mozambique, Bhutan and Afghanistan.
Address: Dobigat, Ward 4, Lalitpur, Katmandu, Nepal.
E-mail: thomas@winderl.net

Arin van Zee currently works for the Belasting Dienst in the Netherlands. He was previously a Project Assistant at ECDPM.
E-mail: az@dds.nl