Measuring Results

Claire Mcloughlin and Oliver Walton

GSD RC
Governance and Social
Development Resource Centre

Topic Guide on
Measuring Results

Claire Mcloughlin and Oliver Walton

About this guide

How can the impact of governance and social development programmes be assessed with a view to improving their efficiency and effectiveness? What particular challenges are involved in monitoring and evaluating development interventions, and how can these be addressed? How can the 'value for money' of a particular intervention be determined?

Monitoring and evaluation (M&E) is vital to ensuring that lessons are learned in terms of what works, what does not, and why. M&E serves two main functions: 1) it builds accountability by demonstrating good use of public funds; and 2) it supports learning by contributing to knowledge about how and why programmes lead to intended (or unintended) outcomes. There can sometimes be a tension between these functions.

This guide introduces some of the core debates and considerations for development practitioners involved in designing and managing M&E activities. It introduces key tools and approaches, provides case studies of applying different methodological approaches, and presents lessons learned from international experience of M&E in a range of developing country contexts. While the guide focuses on M&E for governance and social development programmes, it has relevance for all programmes.

The guide was originally prepared by Claire Mcloughlin, and was comprehensively updated by Oliver Walton. The GSDRC appreciates the contributions of Claire Vallings and Lina Payne (DFID) and Hugh Waddington and colleagues at 3ie. Comments, questions or documents for consideration should be sent to enquiries@gsdrc.org.

About the GSDRC

The Governance and Social Development Resource Centre (GSDRC) provides cutting-edge knowledge services on demand and online. It aims to help reduce poverty by informing policy and practice in relation to governance, conflict and social development. The GSDRC receives core funding from the UK Department for International Development (DFID) and from the Australian Agency for International Development (AusAID).
www.gsdrc.org

Contents

Context and definitions

The changing context of monitoring and evaluation

The international context for M&E is changing, with increasing focus on measuring results and critically analysing aid effectiveness. Several international initiatives and agreements over the past decade have implications for development evaluation:

- The Millennium Development Goals now form the basis of progress indicators at the national level, and have meant much greater interagency cooperation around data collection.
- The Monterrey Declaration (2002) acted as the major impulse towards results-based monitoring and evaluation and the promotion of the allocation of aid based on the development of measures of effectiveness and results.
- The Paris Declaration (2004) committed OECD countries to increasing the country ownership of programmes, encouraging donors to harmonise and align with country monitoring and results frameworks.
- This commitment was re-affirmed at the Accra High Level Forum for Aid Effectiveness (2008) which highlighted measuring development results.

A number of new trends have emerged during this period:

- From monitoring and evaluating project processes, inputs and outputs to an emphasis on measuring results, outcomes and impact.
- From monitoring and evaluating projects, to a new emphasis on evaluating the combined effects of aid – in part prompted by the increase in multi-donor programmes and sector-wide approaches.
- From M&E being predominately donor-led to increased interest in country-led approaches, with evaluations increasingly conducted in partnership with a broader range of stakeholders, including the programmes' intended beneficiaries.
- An increased emphasis on attribution issues and a growing commitment to high quality analysis that is capable of demonstrating 'value for money'.
- A growing interest in participatory approaches to M&E.
- A growing focus on measuring less tangible aspects of development such as governance and empowerment, and an increased use of mixed methods to capture complexity.
- A renewed emphasis on sharing and using the results of evaluations effectively.

While there has been tangible progress in recent years, such as a growth in resources committed to M&E and growing independence of evaluation bodies, several challenges remain. Current key issues include the question of how to improve joint working with other donors and how to ensure more coherent follow up and post-evaluation activities.

OECD, 2010, 'Better Aid: Evaluation in Development Agencies', OECD DAC Network on Development Evaluation, OECD Publishing, Paris
http://www.gsdrc.org/go/display&type=Document&id=4088
How is evaluation managed and resourced in development agencies? What are the major trends and challenges in development evaluation? This report reviews evaluation in 23 bilateral aid donors and seven multilateral development banks. It finds that significant progress has been made towards creating credible, independent evaluation processes. Challenges include the need to improve collaboration with other donors and country partners and to support the use of findings and take-up of recommendations. The use of evaluations needs to be systematised, for instance, by integrating the consultation of relevant reports into the planning process for new programmes or country strategies. Lessons from evaluations need to be better formulated and targeted to specific audiences in accessible and useable ways.

Common tools and approaches

Monitoring and Evaluation can take place at the project, programme, sector, or policy level. It is essential to develop an M&E framework (the strategy for planning and organising the vital M&E activities) during the design phase of projects or programmes. Designing evaluation alongside a programme can facilitate the development of a counterfactual through control groups.

Most projects or programmes formulate a 'logical framework', which hypothesises the causal link between inputs and outcomes. Nevertheless, there is no universal approach to designing or implementing M&E, and a wide range of tools are used.

World Bank Operations Evaluation Department, 2004, 'Monitoring and Evaluation: Some Tools, Methods and Approaches', The World Bank, Washington DC, Second Edition
http://www.gsdrc.org/go/display&type=Document&legacyid=1130

Monitoring and Evaluation (M&E) is an area of growing importance for the development community. It allows those involved in development activities to learn from experience, to achieve better results and to be more accountable. This report provides an overview of some of the M&E tools, methods and approaches on offer to development practitioners.

DFID, 2009, 'Guidance on Using the Revised Logical Framework', How to note, Department for International Development, London
http://mande.co.uk/blog/wp-content/uploads/2009/06/logical-framework.pdf

Terms and definitions

Aid agencies use their own distinct terminology to describe their M&E systems and processes. Nevertheless, the OECD DAC's definitions of key terms are widely accepted:

Organisation for Economic Co-operation and Development – Development Assistance Committee (OECD-DAC), 2002, 'Glossary of Key Terms in Evaluation and Results-based Management', Working Party on Aid Evaluation, OECD-DAC, Paris
http://www.oecd.org/dataoecd/29/21/2754804.pdf
The DAC Working Party on Aid Evaluation (WP-EV) has developed this glossary of key terms in evaluation and results-based management to clarify concepts and reduce the terminological confusion frequently encountered in these areas. With this publication, the WP-EV hopes to facilitate and improve dialogue and understanding among all those who are involved in development activities and their evaluation. It should serve as a valuable reference guide in evaluation training and in practical development work.

Monitoring

'Monitoring' is the ongoing, systematic collection of information to assess progress towards the achievement of objectives, outcomes and impacts. It can signal potential weaknesses in programme design, allowing adjustments to be made. It is vital for checking any changes (positive or negative) to the target group that may be resulting from programme activities. It is usually an internal management activity conducted by the implementing agency.

Evaluation

'Evaluation' is defined by the OECD-DAC (2002) as:

> "The systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability."

Evaluations can be periodic/formative – conducted to review progress, predict a project's likely impact and highlight any necessary adjustments in project design; or terminal/summative – carried out at the end of a project to assess project performance and overall impact. They can be conducted internally, externally by independent consultants (particularly in the case of impact evaluations), or as a joint internal/external partnership.

Evaluations are independent, and are published in order to share lessons, build accountability, and investigate the theory and assumptions behind an intervention. A review, by contrast, tends to be undertaken by people involved in a policy or programme. It focuses on whether or not a programme has met its stated objectives, and is usually intended for internal use.

Impact

'Impact' is defined by the OECD-DAC (2002) as:

> "Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended."

Impact has another, conceptually different definition, however, which is the difference made by an intervention. An impact evaluation refers to "an evaluation concerned with establishing the counterfactual, i.e. the difference the project made (how indicators behaved with the project compared to how they would have been without it)" (White 2006). A counterfactual is defined by the OECD-DAC as, "The situation or condition which hypothetically may prevail for individuals, organizations, or groups were there no development intervention" (2002). The conceptual difference is important, since technically speaking an impact evaluation can measure the difference the project or programme made in short or longer term, or for intermediate or final, or direct or indirect outcomes, for example. However, this definition and the OECD

definition of impact are not necessarily at odds, and have for example been brought together in the NoNIE guidelines on impact evaluation (Leeuw and Vaessen, 2009).

Where is a good place to start?

For an introduction to both the context and application of M&E in development, see:

Kusek, J., and Rist, R., 2004, 'Ten Steps to a Results-based Monitoring and Evaluation System', World Bank, Washington DC
http://www.gsdrc.org/go/display&type=Document&id=2668
Governments and organisations face increasing internal and external pressures to demonstrate accountability, transparency and results. Results-based monitoring and evaluation (M&E) systems are a powerful public management tool to achieve these objectives. This handbook presents a ten-step model that provides extensive detail on building, maintaining and sustaining a results-based M&E system.

Impact evaluation

Impact evaluation: an introduction

The recent emphasis on accountability and results-based management has stimulated interest in evaluating not just the process, outputs and outcomes of development programmes, but also their impact (ultimate effect) on people's lives. Impact evaluations go beyond documenting change to assess the effects of interventions on individual households, institutions, and the environment, relative to what would have happened without them – thereby establishing the counterfactual and allowing more accurate attribution to interventions.

This counterfactual approach to evaluation is increasingly advocated as the only reliable way to develop an evidence base on what works and what does not in development. There are some 800 quantitative impact evaluations in existence across a wide range of sectors, and more are in progress or being commissioned. There is growing consensus that more rigorous quantitative approaches such as randomised control trials should be used more widely, but they are not appropriate in all contexts.

There is growing consensus that where RCTs are not appropriate, there remain a range of quantitative counterfactual approaches for large *n* interventions (where there are many units of assignment to the intervention – such as families, communities, schools, health facilities, even districts). It is possible to collect outcomes data using qualitative methods, within the context of a counterfactual evaluation design. For small *n* interventions (where there are few or only one unit of assignment – such as an intervention carried out in just one organisation, or one which affects everyone in the relevant population – mixed methods that combine quantitative and qualitative methods may be appropriate. All impact evaluations should collect information along the causal chain to explain not just whether the intervention was effective, but why, and so enhance applicability/generalisabiltiy to other contexts.

Lucas, H. and Longhurst, H., 2010, 'Evaluation: Why, for Whom and How?', IDS Bulletin, vol. 41, no. 6, pp. 28-35
http://www.gsdrc.org/go/display&type=Document&id=4040
This article discusses theoretical approaches to evaluation and draws on experiences from agriculture and health. It notes that different stakeholders may have varying expectations of an evaluation and that alternative approaches to evaluation are more suited to meeting some objectives than others. Randomised control trials, or well-designed quasi-experimental studies, probably provide the most persuasive evidence of the impact of a specific intervention but if the primary aim is systematic learning a Theories of Change or Realistic Evaluation approach may be of greater value. If resources permit, different approaches could be combined to cover both accountability and learning objectives. As there will be trade-offs between objectives, transparency and realistic expectations are essential in evaluation design.

White, H., 2009, 'Some Reflections on Current Debates in Impact Evaluation', Working Paper 1, 3ie, New Delhi
http://www.gsdrc.org/go/display&type=Document&id=4103
There is a debate in the field of impact evaluation between those promoting quantitative approaches and those calling for a larger range of approaches to be used. This paper highlights four misunderstandings that have arisen in this debate. They involve: 1) crucially, different definitions of 'impact' – one based on outcomes and long term effects, and one referring to attribution; 2) confusion between counterfactuals and control groups; 3) confusion of 'attribution' with sole attribution; and 4) unfounded criticism of quantitative methods as 'positivist' and 'linear'. There is no hierarchy of methods, but quantitative approaches are often the best available.

Asian Development Bank, 2011, 'A Review of Recent Developments in Impact Evaluation', Asian Development Bank, Manila
http://www.gsdrc.org/go/display&type=Document&id=4220
How can impact be credibly attributed to a particular intervention? This report discusses the merits and limitations of various methods and offers practical guidance on impact evaluation. A rigorously conducted impact evaluation produces reliable impact estimates of an intervention through careful construction of the counterfactual using experimental or non-experimental approaches.

Attribution and the counterfactual: the case for more and better impact evaluation

Development interventions are not conducted in a vacuum. It is extremely difficult to determine the extent to which change (positive or negative) can be attributed to the intervention, rather than to external events (such as economic, demographic, or policy changes), or to interventions by other agencies.

Impact evaluations attempt to attribute change to a specific programme or policy and establish what would have happened without the intervention (the counterfactual) by using scientific, sometimes experimental, methodologies such as randomised control trials or comparison groups.

A number of organisations and networks have emerged in recent years to make the case for more rigorous evaluation methods. These include the Network of Networks for Impact Evaluation (NONIE), the Abdul Latif Jameel Poverty Action Lab (J-PAL) and the International Initiative for Impact Evaluation (3ie). The final report (2006) of the Evaluation Gap Working Group at the Centre for Global Development is a seminal document calling for greater use of impact evaluation:

Center for Global Development, 2006, 'When Will We Ever Learn? Improving Lives Through Impact Evaluation', Evaluation Gap Working Group, Centre for Global Development, New York
http://www.gsdrc.org/go/display&type=Document&id=2685
Despite decades of investment in social development programmes, we still know relatively little about their net impact. So why are rigorous social development impact evaluations relatively rare? This paper examines this question and provides recommendations for more and better evidence for policymaking and programme planning. A new, collective approach is needed, in which developing country governments, bilateral and multilateral development agencies, foundations and NGOs work together to define an agenda of enduring questions and fund the design and implementation of rigorous impact evaluations in key sectors.

The NONIE guidelines reflect the views of a number of impact and evaluation networks. They stress that there is no single 'best' method for assessing the impacts of interventions, but that some methods have a comparative advantage over others in analysing particular objectives. Quantitative methods (experimental or quasi-experimental methods) have a comparative advantage in large n interventions and in addressing the issue of attribution. Certain tools or approaches may complement each other, providing a more complete 'picture' of impact. The guidelines present a list of eight experimental and quasi-experimental methods for causal attribution, but acknowledge that other methods may be needed for more complex or small n interventions.

Leeuw, F. and Vaessen, J., 2009, 'Impact Evaluations and Development: NoNIE Guidance on Impact Evaluation', NONIE - The Network of Networks on Impact Evaluation, Washington DC
http://www.gsdrc.org/go/display&type=Document&id=4099
This report provides a guide to evaluating the impact of a project or programme. Impact evaluation is about attributing impacts to interventions, and can play a key role in development effectiveness. No single analytical method is best for addressing all aspects of impact evaluations, but some methods have an advantage over others in addressing a particular question or objective. Different methods can complement each other to provide a more complete picture of impact. Impact evaluations provide the greatest value when there is an articulated need to obtain the information they generate.

Randomised Control Trials: the gold standard?

Randomised Control Trials (RCTs) are often referred to as the 'gold standard' of impact evaluation, but whether or not they are always feasible, appropriate and rigorous is the subject of some debate. Since RCTs seek to measure a counterfactual, they often require data collection before the start of an intervention, which makes it difficult to apply this approach after an intervention has begun. One of the key issues with RCTs is the problem of 'external validity', or the degree to which the findings of one study can be generalised to other contexts. Some argue that while RCTs may be suitable for measuring simple, short-term development interventions, they are less suitable for more complex, long-term interventions, where many factors seek to produce change.

Duflo, E., and Kremer, M., 2003, 'Use of Randomization in the Evaluation of Development Effectiveness', Paper prepared for the World Bank Operations Evaluation Department Conference on Evaluation and Development Effectiveness 15-16 July, 2003, Massachusetts Institute of Technology, Cambridge, US
http://www.gsdrc.org/go/display&type=Document&id=2626
Just as randomised pharmaceutical trials revolutionised medicine in the 20th Century, randomised evaluations could revolutionise social policy in the 21st. This paper draws on evaluations of educational programmes. It argues that there is an imbalance in evaluation methodology and recommends greater use of randomised evaluations. As credible impact evaluations, these could offer valuable guidance in the search for successful programmes.

Case Study: Progresa in Mexico

The Progresa case is considered one of the most successful examples of the application of a randomised control trial in a development context:

Attanasio, O., et al., 2005 'Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progresa', Institute for Fiscal Studies, London
http://www.gsdrc.org/go/display&type=Document&id=2647
What impact have monetary incentives had on education choices in rural Mexico? How can the design of educational interventions aimed at improving educational participation be improved? This paper analyses the education component of the Mexican government's welfare programme, Progresa, which aims to reduce rural poverty. It argues that increasing the grant for secondary school children while eliminating it at the primary age would strengthen Progresa's impact.

For further case studies of randomised control trials, see: www.povertyactionlab.com.

Adapting to time, budget and data constraints

Ideological positions can obscure the issue of which methodologies are actually feasible. Scientific approaches can be costly, time consuming, and therefore unrealistic. Many organisations do not have the resources to carry out the ideal evaluation, and an M&E framework needs to be designed with organisational capacity, human and financial resources and political context in mind. Although it will not be feasible to rigorously evaluate all projects, donors such as DFID now require that all projects must be considered for some form of evaluation as part of the design process. The rationale for the decision to evaluate or not must be defensible to the UK's new Independent Commission for Aid Impact.

It is important to understand the minimum methodological requirements for evaluation rigour in cases where it is not possible to use strong evaluation designs.

Bamberger, M., 2006, 'Conducting Quality Impact Evaluations Under Budget, Time and Data Constraints', World Bank, Washington DC
http://www.gsdrc.org/go/display&type=Document&id=2657
How do cost, time and data constraints affect the validity of evaluation approaches and conclusions? What are acceptable compromises and what are the minimum methodological requirements for a study to be considered a quality impact evaluation? This booklet provides advice for conducting impact evaluations and selecting the most rigorous methods available within the constraints faced. It provides suggestions for reducing costs and increasing rigour and clarifies the nature of trade-offs between evaluation rigour and budgets, time and data.

Evaluation designs need to be adapted to local realities: experience demonstrates that no single methodology is applicable in all cases.

White, H., 2006, 'Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank', World Bank, Washington DC
http://www.gsdrc.org/go/display&type=Document&id=2628
Aid spending is increasingly dependent on proof that interventions are contributing to the attainment of Millennium Development Goals (MDGs). Yet there is still debate over the definition of impact evaluation and how it should be carried out. This paper defines impact evaluation as a 'counterfactual analysis of the impact of an intervention on final welfare outcomes' and recommends a theory-based approach. Two sources of bias are highlighted: contamination and self-selection bias.

Mixed-method designs

As discussed in the introduction to this section, quantitative and qualitative methods can be used to answer different evaluation questions. Quantitative methods are more useful for answering 'what works' questions, while qualitative methods are more useful for answering 'why' questions. Quantitative methods are more suited to large *n* interventions, while mixed methods or just qualitative methods should be used in small *n* interventions. Mixed methods can help to bolster findings where there are gaps in the data. These approaches are particularly useful for assessing aspects of poverty that are not easy to quantify, such as governance, trust, empowerment and security.

Garabino, S., and Holland, J., 2009, 'Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results', Issues Paper, Governance and Social Development Resource Centre, Birmingham
http://www.gsdrc.org/docs/open/EIRS4.pdf
This paper reviews the case for promoting and formalising qualitative and combined methods for impact evaluation and measuring results. The case for qualitative and combined methods is strong. Qualitative methods have an equal footing in evaluation of development impacts and can generate sophisticated, robust and timely data and analysis. Combining qualitative research with quantitative instruments that have greater breadth of coverage and generalisability can result in better evaluations that make the most of their respective comparative advantages.

A growing issue in development evaluation is the need for M&E practices to reflect and embrace the insights of complexity science. This involves recognising that the contexts within which governance and development interventions are conducted are unstable and unpredictable, and that existing linear models of causality are ill-equipped to understand change in these contexts. A complexity perspective implies the need for more flexible and adaptive approaches.

Ramalingam, B. and Jones, H. et al., 2008, 'Exploring the Science of Complexity: Ideas and Implications for Development and Humanitarian Efforts', Working Paper 285, Overseas Development Institute, London, Second Edition
http://www.gsdrc.org/go/display&type=Document&id=4093
What is complexity science? How can it contribute to development and humanitarian efforts? This paper explores the key concepts of complexity science and shows how they might help development practitioners engaged in reform. The concepts highlight that the best course of action will be context-dependent, and they offer new ways to think about

questions that should be posed. Development practitioners need to recognise that they live with complexity on a daily basis, and to use the 'complexity lens'.

Toolkits

Baker, J., 2000, 'Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners', World Bank, Washington DC
http://www.gsdrc.org/go/display&type=Document&id=2631
There is broad evidence that developmental assistance benefits the poor, but how can we tell if specific projects are working? Have resources been spent effectively? What would have happened without intervention? This comprehensive handbook seeks to provide tools for evaluating project impact. It advises that effective evaluations require financial and political support, early and careful planning, participation of stakeholders, a mix of methodologies and communication between team members.

World Bank, 2006, 'Impact Evaluation and the Project Cycle', PREM Poverty Reduction Group, World Bank, Washington DC
http://www.gsdrc.org/go/display&type=Document&id=2707
The goal of an impact evaluation is to attribute impacts to a project using a comparison group to measure what would have happened to the project beneficiaries had it not taken place. The process of identifying this group, collecting the required data and conducting the relevant analysis requires careful planning. This paper provides practical guidance on designing and executing impact evaluations. It includes some illustrative costs and ideas for increasing government buy-in to the process.

Duflo, E., Glennerster, R., and Kremer, M., 2006, 'Using Randomization in Development Economics Research: A Toolkit', J-PAL paper.
http://www.povertyactionlab.org/sites/default/files/documents/Using%20Randomization%20in%20Development%20Economics.pdf

Theory-based evaluation

Impact evaluations have increasingly focused not simply on the question of *what* works, but also *why* an intervention achieved its intended impact – or why it did not. There has been growing emphasis on incorporating analysis of causal chains and a growing interest in 'theory-based' impact evaluation. Despite this interest, few studies apply the approach in practice.

White, H., 2010, 'Theory-Based Impact Evaluation: Principles and Practice', 3ie Working Paper 3, New Delhi
http://www.gsdrc.org/go/display&type=Document&id=4102
How can impact evaluation identify not just what does – or does not – work, but why? A theory-based approach to impact evaluation maps out the causal chain from inputs to outcomes and impact, and tests the underlying assumptions. Despite wide agreement that this approach will address the why question, it has not often been effectively used. This paper outlines six principles for successful theory-based impact evaluation: (1) map out the causal chain (programme theory); (2) understand context; (3) anticipate heterogeneity; (4) rigorously evaluate impact using a credible counterfactual; (5) use rigorous factual analysis; and (6) use mixed methods.

One popular technique has been to use a 'theory of change' to help design and implement development programmes. While a logical framework graphically illustrates programme components and helps stakeholders to identify inputs, activities and outcomes, a theory of change links outcomes and activities to explain how and why the desired change is expected to come about, drawing on underlying assumptions.

Retolaza, I., 2011, 'Theory of Change: A thinking and action approach to navigate in the complexity of social change processes', Humanist Institute for Development Cooperation / United Nations Development Programme
http://www.gsdrc.org/go/display&type=Document&id=4095
What is a Theory of Change (ToC) and why is it important? This guide to understanding and developing a ToC shows how a ToC helps to configure the conditions needed to achieve desired change, using the experience of a given context. This is done partly by making assumptions explicit and by analysing them critically. It is also a monitoring tool that facilitates accountability. A good ToC allows development practitioners to handle complexity without over-simplification.

More detail about the 'theory of change' approach and examples of how it can be applied can be found at www.theoryofchange.org.

Key monitoring and evaluation tools

Participatory Approaches

Monitoring and evaluation (M&E) has typically been led by outside experts, measuring performance against pre-set indicators and using procedures and tools designed without the participation of key stakeholders such as the programmes' intended beneficiaries. Evaluations in particular, because they are very often conducted by external consultants, can be seen as a form of control.

For a number of years, there has been widespread recognition that M&E should take a more inclusive, participatory approach. Participation in this sense means the involvement of stakeholders in deciding how the project or programme should be measured, in identifying and analysing change, and in acting on results. Nevertheless, there are few empirical studies of the effectiveness, quality and objectivity of participatory approaches. Questions such as how to operationalise participation and which methods work in which contexts are the subject of ongoing debate.

The case for participatory M&E

Proponents of participatory monitoring and evaluation (PM&E) argue that it is more cost-effective, accurate and sustainable than conventional approaches. Participation in decision-making processes can also motivate people to want to see those decisions implemented effectively. Another motivation for PM&E is to strengthen organisational and institutional learning.

Since the 1990s there has been growing innovation in the use of quantitative approaches to PM&E. A range of methods and tools have emerged including counting, mapping, valuing and scoring. While these approaches have proved effective in some contexts, their success in empowering communities ultimately remains dependent on official attitudes and acceptance.

Chambers, R., 2007, 'Who Counts? The Quiet Revolution of Participation and Numbers', IDS Working Paper No. 296, Institute of Development Studies, Brighton
http://www.gsdrc.org/go/display&type=Document&id=3002
How do participatory approaches generate numbers? A range of participatory methods have been developed, many visual and tangible, by which local people themselves produce numbers. This paper provides an overview of the field, building a case for more widespread use of participatory numbers in development practice and research. It argues that in many contexts, processes generating participatory numbers are better than traditional questionnaires. They provide greater validity, better insights, and increased knowledge gains for local people.

Challenges in using participatory approaches

Whilst the ideological case for participation is widely acknowledged, PM&E is not without its critics. Crucially, labelling M&E as 'participatory' does not necessarily guarantee that all stakeholder groups have participated, and there are often issues around who participates and who is excluded from these processes. Subsequently, the representativeness of the findings and recommendations of participatory evaluations have been criticised.

Gregory, A., 2000, 'Problematizing Participation: A Critical Review of Approaches to Participation in Evaluation Theory', Evaluation, vol. 6, no. 2, pp. 179-199
http://www.gsdrc.org/go/display&type=Document&id=2678
It is widely accepted that evaluation is a social process which implies the need for a participatory approach. But what is understood by 'participation'? This review argues that the blanket use of the term has masked the heterogeneity evident in its realisation in practice, and highlights a lack of transparency in participatory evaluation methods.

Operationalising PM&E can also be problematic. There is a need to be sensitive to the local socio-economic and political situation, and consider under what conditions PM&E approaches can be used without increasing the vulnerabilities of already marginalised groups.

Estrella, M., 2000, 'Learning from Change: Issues and Experiences in Participatory Monitoring and Evaluation', International Development Research Centre, Ottawa
http://www.gsdrc.org/go/display&type=Document&id=2708
Since the 1980s, concepts of participatory monitoring and evaluation (PM&E) have entered the policymaking domain of larger donor agencies and development organisations. This introductory chapter draws on twelve case studies to describe how different stakeholders have applied PM&E approaches across a range of purposes and contexts. It outlines some of the key concepts and differences between participatory and conventional approaches to M&E and highlights emerging issues.

Participatory tools: case studies

There is no single definition or methodology of PM&E, and it encompasses a wide range of tools and approaches. Common among these approaches are values such as shared learning, democratic processes, joint decision making, co-ownership, mutual respect and empowerment. Below are some examples of participatory tools and how they have been used in practice.

Participatory Rural Appraisal

Participatory Rural Appraisal (PRA) encompasses a broad range of methods to enable local people to analyse their own realities as the basis for planning, monitoring and evaluating development activities. PRA uses group exercises to facilitate information sharing, analysis, and action among stakeholders.

Chambers, R., 2007, 'From PRA to PLA and Pluralism: Practice and Theory', IDS Working Paper, no. 286, Institute of Development Studies, Brighton
http://www.gsdrc.org/go/display&type=Document&id=2786
How have the theory and practice of participatory methodologies in development activities changed since the mid 1970s? What variants and applications of these methodologies have emerged? This paper traces the spread of Participatory Rural Appraisal (PRA) and Participatory Learning and Action (PLA), and identifies strengths and weaknesses in the ways they have been implemented. Whilst inflexible applications of PRA and PLA may produce disappointing results, when executed with spontaneity and creativity, these approaches can be a source of community empowerment.

Beneficiary Feedback

Beneficiary feedback systems seek to collect key stakeholders' views about the quality and impact of a development agency's work. This approach has grown in popularity and has been supported by a range of donors. While beneficiary feedback systems are likely to improve sustainability and can empower beneficiaries, they may present only a partial impression of beneficiaries' views, and there has been a lack of rigorous evaluation of their impact. While there is growing interest in the role that new information and communication technologies can play in scaling up PM&E techniques, these advances do not provide a 'magic bullet' to improving participation.

Jacobs, A., 2010, 'Creating the Missing Feedback Loop', IDS Bulletin, Vol. 41, No. 6, pp.56-64
http://www.gsdrc.org/go/display&type=Document&id=4090
How can development agencies implement feedback systems so as to hear systematically from the intended beneficiaries of their work? How can such systems improve aid accountability, and thus effectiveness? This article focuses on agricultural development, arguing that in most agricultural development projects, quantified summaries of smallholder farmers' views can be collected. Such data can provide real-time performance indicators that incentivise staff to focus on beneficiaries' priorities. The feedback process can be empowering in itself, and acting on the feedback can significantly improve project impact and sustainability. While measuring performance against plans encourages supply-led and agency-centred development, using feedback systems can encourage demand-led and people-centred development.

GSDRC, 2010, 'Participatory M&E and Beneficiary Feedback', Helpdesk Research Report, Governance and Social Development Resource Centre, Birmingham
http://www.gsdrc.org/go/display&type=Helpdesk&id=619
PM&E and beneficiary feedback approaches can improve effectiveness and sustainability and reduce the costs of monitoring, but risk generating unrepresentative results and increasing tensions between stakeholders. PM&E should be understood as social and political processes: PM&E and beneficiary feedback mechanisms should be tailored to the local context and integrated into local political structures and processes.

Key Informant Interviews

Key informant interviews are a rapid assessment methodology that can be used as an intermediate indicator of outcomes as an alternative or supplement to full impact assessments.

Price, N., and Pokharel, D., 2005, 'Using Key Informant Monitoring in Safe Motherhood Programming in Nepal', Development in Practice, vol. 15, no. 2, pp. 151-164
http://www.gsdrc.org/go/display&type=Document&id=2627
The Nepal Safer Motherhood Project (NSMP) works to improve maternal health and contribute to programme development at district and national level. This article discusses the project's use of Key Informant Monitoring (KIM). KIM is an adapted version of the peer ethnographic research method. Data is collected by community-based Key Informant Researchers (KIRs) and used for monitoring and planning. KIRs have proved useful sources of information and acted as change agents by spreading safer motherhood messages.

Most Significant Change Technique

The Most Significant Change (MSC) technique involves the collection of change stories from the field level, and the systematic selection of the most significant of these. These selected stories are then discussed and critically reflected on to help determine the impact of the development programme or activity.

Davies, R., 2005, 'The 'Most Significant Change' (MSC) Technique: A Guide to its Use', MandE, London
http://www.gsdrc.org/go/display&type=Document&id=2672
This paper outlines an innovative qualitative monitoring technique known as the 'most significant change' (MSC) approach. The MSC technique is a participatory method of collecting and analysing stories from the field which focuses on monitoring intermediate outcomes and impact. It provides a simple means of making sense of a large amount of complex information and is best suited to large-scale, open-ended projects that would be difficult to monitor using traditional methods.

There is evidence that the most significant change technique can enhance organisational learning and performance.

Wrigley, R., 2006, 'Learning from Capacity Building Practice: Adapting the 'Most Significant Change' (MSC) Approach to Evaluate Capacity Building', INTRAC Praxis Paper no. 12, International NGO Training and Research Centre, Oxford
http://www.gsdrc.org/go/display&type=Document&id=2669
There is growing recognition of the need to take a multi-stakeholder approach to evaluation, which promotes local ownership and builds capacity for reflection, learning, improved performance and self-determination. This paper reflects on the use of the 'Most Significant Change' (MSC) methodology to evaluate the capacity building services of CABUNGO, a local capacity building support provider in Malawi.

Outcome Mapping

Outcome Mapping is an alternative to theory-based approaches to evaluation that rely on a cause–effect framework. It recognises that multiple, non-linear events lead to change. It focuses on people and changes of behaviour and how far development interventions have built the capacity of the local community. Outcome mapping assumes only that a contribution has been made, and never attempts attribution.

Earl, S., Carden, F., and Smutylo., T., 2001, 'Outcome Mapping: Building Learning and Reflection into Development Programs', International Development Research Centre, Ottawa
http://www.gsdrc.org/go/display&type=Document&id=2687
This paper argues that impacts are often the product of events for which no single agency can claim full credit. Outcome mapping moves away from impact assessments to focus on changes in the behaviour of the people it works with directly.

The following paper reviews the use of outcome mapping, examining its key advantages, the contexts in which it is most useful and how donors can facilitate its use. It finds that outcome mapping works best in situations where the implementing agency is working in partnership with a range of other stakeholders, is seeking to build capacity, or when tackling complex problems.

Jones, H. and Hearn, S., 2009, 'Outcome Mapping: A Realistic Alternative for Planning, Monitoring and Evaluation', Overseas Development Institute, London
http://www.gsdrc.org/go/display&type=Document&id=4098
What is Outcome Mapping (OM) and why is it valuable? When does it work best? How can donors facilitate its use? This note draws on case studies to review OM – a flexible, actor- and learning-centred approach to planning, monitoring, and evaluating social change initiatives. It finds that adopting OM for appropriate projects could help development agencies to increase their effectiveness and meet commitments to managing for results. OM is well-suited to areas involving complex change processes, capacity building work, and knowledge and decision-making processes. Shifting to OM's learning-oriented mode requires donors to adopt more realistic expectations and to dispense with the idea of 'controlling' change processes. Crucially, OM must be underpinned by real trust between the donor, project implementers and partners.

The Outcome Mapping Learning Community is an informal community that allows practitioners to share knowledge and experiences. A number of case studies are available on its website: www.outcomemapping.ca

Gender analysis

Development interventions can have differential impacts on men and women: men and women have different needs and constraints, different opportunities to participate in programme design and implementation, and benefit differently from outcomes and impacts. A gender analysis framework should therefore be a component of all evaluation designs. It is also important to have a gender-balanced evaluation team.

World Bank, 2002, 'Integrating Gender into the PRS Monitoring and Evaluation' in Chapter 10: Gender of A Sourcebook for Poverty Reduction Strategies, World Bank, Washington DC, pp.360-366
http://www.gsdrc.org/go/display&type=Document&id=2705

There is growing evidence that gender-sensitive development strategies contribute significantly to economic growth and equity objectives by ensuring that all groups of the poor share in programme benefits. Yet differences between men's and women's needs are often not fully recognised in poverty analysis and participatory planning, and are frequently ignored in the selection and design of Poverty Reduction Strategies (PRSs). A full understanding of the gender dimensions of poverty can significantly change the definition of priority policy and programme interventions supported by the PRS. This chapter provides practical guidance on identifying and implementing country-level policies and programmes that will benefit both men and women, and maximise potential benefits for poor families.

Evaluators have had limited success in introducing gender sensitive approaches to M&E. This is particularly evident in the use of household surveys which either only interview the household head (usually male) or which interview women in contexts where they are not able to speak freely.

Evaluations should address critical gender issues such as time poverty, participation in household decision-making and women's multiple roles (e.g. production, social reproduction and community management).

Bambrilla, P., 2001, 'Gender and Monitoring: A Review of Practical Experiences', report prepared for the Swiss Agency for Development Cooperation, BRIDGE Report 63, BRIDGE, Institute of Development Studies, Brighton
http://www.gsdrc.org/go/display&type=Document&id=2649
How can monitoring and evaluation (M&E) processes be made gender-sensitive? What measures have organisations taken to assess their effectiveness in mainstreaming gender? This report provides a tool for integrating a gender approach into M&E mechanisms.

Value for Money

'Value for money' (VFM) is a term used to describe an 'explicit commitment to ensuring the best results possible are obtained from the money spent' (Barnett et al 2010). The term has been used by a number of donors and multilateral agencies as part of their appraisal and evaluation procedures. These include DFID, USAID, DANIDA and the Millennium Challenge Commission. There is also growing interest in applying VFM frameworks to development programmes.

Whilst always present in aid effectiveness discussions, the VFM agenda has increased in prominence as a result of growing concern with transparency and accountability in government spending. DFID has stated that improving VFM is not simply an exercise in cost-cutting. It involves addressing economic, efficiency and effectiveness considerations to ensure that fewer resources are spent where possible, and that these resources are spent productively and are focused on achieving objectives.

The following study highlights several options for improving and assessing value for money. While some approaches are more focused on management (they are easier to implement and involve a strengthening of existing processes), others focus on measurement. The latter represent a more radical departure from existing practice and require more rigorous data collection. The 'ratings and weightings' approach, where programmes are rated according to pre-determined standards of economy, efficiency and effectiveness, is arguably the most promising.

Barnett, C., et al., 2010, 'Measuring the Impact and Value for Money of Governance Programmes', ITAD
http://www.gsdrc.org/go/display&type=Document&id=4115
How can value for money best be measured in governance and conflict programming? This study reviews options for a VFM approach in relation to governance programmes, including those in conflict-affected and failed states, for the UK's Department for International Development. VFM involves examining economy, efficiency and effectiveness, identifying the links between them and drawing conclusions based on evidence about how well they perform together. It is an optimal balance that is important, as opposed to a maximum productivity ratio. The cheapest option does not always represent the best value for money.

Donors are using the term 'value for money' in different ways. DFID's approach involves assessing whether results achieved represent good value against the costs incurred. Other agencies, such as USAID and the World Bank, aim to achieve VFM through rigorous economic analysis or results-based management.

GSDRC, 2010, 'Value For Money', Helpdesk Research Report, Governance and Social Development Resource Centre, Birmingham
http://www.gsdrc.org/go/display&type=Helpdesk&id=624
How are donors approaching 'value for money' in their aid programming? DFID appears to have gone the furthest among aid agencies in developing the concept of 'value for money' (VFM). Processes include the use of logframes, economic appraisals and portfolio reviews. Newer initiatives include the adoption of a business case model for project approval and the development of unit cost metrics in key sectors. Other donors, while not explicitly adopting 'value for money' terminology, aim to achieve VFM through rigorous economic analysis and results-based management.

Managing monitoring and evaluation

Badly designed and managed evaluations can do more harm than good: misleading results can undermine the effective channelling of resources for poverty reduction. Designing effective M&E systems is not enough: donors need to commit to supporting M&E at all stages of implementation, which include selecting appropriate indicators, establishing baselines, collecting quality data, and reporting and using findings effectively. Sustaining an M&E system within an organisation also presents considerable ongoing challenges.

Establishing international standards for methodological rigour, ethical practice and efficient management processes in M&E is another critical challenge. Key issues include how aid agencies should oversee evaluations outsourced to consultants, how to build country ownership of M&E processes where there are significant capacity constraints or limited buy-in, and how to co-ordinate evaluations of joint donor programmes effectively.

Steps in planning and design

Monitoring and evaluation activities are usually broken down into the stages of planning, implementation, analysis, dissemination and use.

Kusek, J., and Rist, R., 2004, 'Ten Steps to a Results-based Monitoring and Evaluation System', World Bank, Washington DC
http://www.gsdrc.org/go/display&type=Document&id=2668
Governments and organisations face increasing internal and external pressures to demonstrate accountability, transparency and results. Results-based M&E systems are a powerful public management tool to achieve these objectives. This handbook presents a ten-step model that provides extensive detail on building, maintaining and sustaining a results-based M&E system.

Aid agencies largely work to their own internal requirements for reviewing, reporting on and evaluating the inputs, process and results of their activities, producing internal guidance notes that describe the practical steps involved.

United Nations Development Programme Evaluation Office, 2009, 'Handbook on Planning, Monitoring and Evaluating for Development Results', UNDP, New York
http://www.undp.org/evaluation/handbook/

Ensuring evaluation quality

Ensuring the quality and integrity of evaluation design is vital for reaching accurate and reliable conclusions about what works and what does not. A key mechanism for ensuring evaluation quality is to establish internal quality control panels. DFID has had such a panel since 2009, providing expert advice, guidance and quality assurance.

International standards emphasise the need for impartiality, accountability, appropriately skilled experts conducting the evaluation, participation, country ownership and timeliness (evaluations should be appropriately timed to influence policymaking). The OECD DAC Network on Development Evaluation's principles for evaluation of development assistance are widely cited.

OECD-DAC, 1991, 'Principles for Evaluation of Development Assistance', OECD, Paris
http://www.gsdrc.org/go/display&type=Document&id=2688
This paper presents a set of principles on the most important requirements of the evaluation process. Development assistance is a cooperative partnership between donors and recipients. Both must take an interest in evaluation to improve the use of resources through learning and to ensure accountability to political authorities and the public.

Most donors now also make substantial use of the harmonised DAC Quality Standards for Development Evaluation. While the OECD Principles focus on the management and institutional set up of evaluation systems, the Standards inform evaluation processes and products:

OECD-DAC, 2010, 'Quality Standards for Development Evaluation', DAC Guidelines and Reference Series, OECD, Paris
http://www.gsdrc.org/go/display&type=Document&id=4114
The OECD-DAC Quality Standards for Development Evaluation, built through international consensus, provide a guide to good practice. They are not intended to be used as a development evaluation manual, but they outline the key quality dimensions for each phase of a typical evaluation process: defining purpose, planning, designing, implementing, reporting, and learning from and using evaluation results. Principles informing the whole of the evaluation process are transparency and independence; integrity and respect for diversity; partnership, coordination and alignment; capacity development; and quality control.

There is also a need to ensure that evaluations are conducted ethically, in a culturally sensitive manner that protects the anonymity and confidentiality of individual informants.

United Nations Evaluation Group, 2005, 'Standards for Evaluation in the UN System', UNEG, New York
http://www.gsdrc.org/go/display&type=Document&id=2684
This document offers solid guidelines for evaluation planning, design, implementation and reporting. Fundamental requirements include: institution-wide support, clearly-defined and transparent responsibilities, appropriately qualified staff, and a constant commitment to the harmonisation and updating of methods used.

Identifying threats to the validity of evaluation findings

Significant criticism has been levelled against the development community for failing to adopt methodologically sound approaches to evaluating their activities. These include weak analysis of qualitative data and not paying enough attention to mapping the causal chain from inputs to impacts.

White, H., 2005, 'Challenges in Evaluating Development Effectiveness', IDS Working Paper 242, Institute of Development Studies, Brighton
http://www.gsdrc.org/go/display&type=Document&id=2662
Evaluation has a crucial role to play in today's results-based culture and in the context of the Millennium Development Goals. How, then, can the quality of evaluation be improved? This paper argues that there has been inadequate investment in methodology, often resulting in low quality evaluation outputs. It discusses techniques in three areas: measuring agency performance; evaluation methods at the project level; and sustainability analysis.

The validity and usefulness of an evaluation are determined, among other things, by its statistical validity, use/action orientation, transferability and fittingness.

Bamberger, M., Rugh, J., and Mabry, L., 2006, 'Strengthening the Evaluation Design and the Validity of the Conclusions' Chapter 7 in Realworld Evaluation: Working Under Budget, Time, Data and Political Constraints, Sage Publications, California
http://www.gsdrc.org/go/display&type=Document&id=2761
How can threats to the validity of evaluations be identified and addressed? This chapter outlines some of the most common threats to the validity of both quantitative and qualitative evaluation designs. It offers recommendations on how and when corrective measures can be taken to ensure validity.

See also: **Bamberger, M., 2007, 'A Framework for Assessing the Quality, Conclusion Validity and Utility of Evaluations', Paper presented at the American Evaluation Association Conference of 2007.**
http://www.realworldevaluation.org/uploads/A_framework_for_assessing_the_quality_and_utility_of_evaluations_GAO_AEA_2007.doc

Evaluating multi-donor programmes

The Paris Declaration commits donors to co-operation and harmonisation in all stages of the development cycle. Joint evaluations are necessary where multiple agencies are involved in a chain of interventions to pursue similar outcomes, or to understand the combined effects of all interventions across a particular sector.

Joint evaluations present opportunities for donors to pool their technical and financial resources for more rigorous, in-depth and longer-term evaluations, and to reduce the multiple information demands on governments and stakeholders. However, they require agencies to reconcile their often divergent mandates and preferred evaluation approaches.

OECD-DAC, 2006, 'Guidance for Managing Joint Evaluations', DAC Evaluation Series, OECD, Paris
http://www.gsdrc.org/go/display&type=Document&id=2663
Joint evaluations have become central to development practice in recent years. Collective assessment of agencies' combined work minimises transaction costs for developing country partners and addresses the large aid-giving role of joined-up modalities such as basket funds and joint assistance strategies. This booklet provides practical guidance for making joint evaluations efficient, educational and collaborative.

The politics of evaluation

Evaluations are more than a technical process. They have the capacity to determine access to resources and the funding fate of programmes. It is inevitable therefore that they will be subject to pressures from different stakeholders to produce favourable assessments or to avoid addressing sensitive issues.

Bamberger, M., Rugh, J., and Mabry, L., 2006, 'Reconciling Different Priorities and Perspectives: Addressing Political Influences', Chapter 6 in Realworld Evaluation: Working Under Budget, Time, Data and Political Constraints, Sage

Publications, California
http://www.gsdrc.org/go/display&type=Document&id=2663
No evaluation can ever be value free and completely objective. Decisions such as what to study, which methods to use and whose criteria define programme success all involve human judgement. This chapter discusses how political factors affect evaluation. It provides a detailed analysis of possible pressures and constraints in evaluation design, implementation, dissemination and use.

Why do the evaluation, monitoring and accountability gaps persist decades after aid agencies started emphasise impact assessment? The following paper argues that the incentives in favour of the status quo are far stronger than the incentives to institutionalise stronger accountability and more rigorous evaluation.

Morton, J., 2009, 'Why We Will Never Learn: A Political Economy of Aid Effectiveness', James Morton & Co
http://www.gsdrc.org/go/display&type=Document&id=4094
Why is M&E still not being carried out effectively? This paper is a practitioner's response to current debates on M&E and aid effectiveness. It examines the technical side of M&E and the latest thinking on complexity, arguing that current methodological debates are red herrings. It highlights the underlying incentive structures that create the need for M&E, but do not create a sincere demand for the impact assessments that M&E is designed to produce. The aid industry's default setting is lesson suppression, not lesson learning.

Promoting the use of evaluation findings

M&E should ultimately result in improved policy and practice. Yet the findings and recommendations of evaluations are frequently under-used. In order for evaluations to be influential, it is important to consider: how to integrate them into the policymaking cycle; the political incentives to take up findings; and how the report is presented and understood by different stakeholders.

World Bank Operations Evaluation Department, 2004, 'Influential Evaluations: Evaluations that Improved Performance and Impacts of Development Programs', World Bank, Washington DC
http://www.gsdrc.org/go/display&type=Document&id=1063
Evaluations can be a cost-effective way of improving the performance and impact of development activities. However, they must be conducted at the right time, focus on key issues and present results in an accessible format. This report presents eight examples of evaluations that have had an important impact, and summarises lessons learned.

The extent to which impact evaluations are used can be enhanced by ensuring greater stakeholder involvement, ensuring that impact evaluations contain clear policy lessons and that these are disseminated to a wide audience.

Jones, N., et al., 2009, 'Improving Impact Evaluation Production and Use', Working Paper 300, Overseas Development Institute, London
http://www.gsdrc.org/go/display&type=Document&id=4100
What can be learned from the many impact evaluations (IEs) carried out by development agencies in recent years? This review highlights the need and growing demand for greater and more strategic coordination of IE efforts, and notes insufficient attention to diverse methodological approaches to evaluation. It is important in all policy sectors to reflect on the suitability of methods to development questions and to invest in the development of impact evaluations informed by methodological pluralism. Developing country evaluation capacity, early stakeholder involvement in IEs, and the dissemination of clear policy implications should also be supported.

Strengthening national M&E capacity

Evaluations of development programmes have historically been driven and designed by donors – primarily to satisfy their own accountability needs. However, it is increasingly recognised that both monitoring and evaluation should be a country-led process, not least because country ownership is a major factor in determining whether evaluation findings are then used in a national context.

Technical barriers to country-led evaluations centre on lack of human and financial resources, but M&E is also a highly political issue. The incentives, or lack of incentives, for evaluations to be conducted (e.g. fear of aid being withdrawn as a result of negative evaluation results) also need to be considered.

There is evidence that capacity development programmes work best when adapted to local governance structures, professional capacity and evaluation culture.

Schiavo-Campo, S., 2005, 'Building Country Capacity for Monitoring and Evaluation in the Public Sector: Selected Lessons of International Experience', World Bank Evaluation Capacity Development Working Paper no. 13, World Bank, Washington DC
http://www.gsdrc.org/go/display&type=Document&id=2706

The Evaluation Capacity Development (ECD) unit of World Bank's Operations Evaluation Department is designed to help countries strengthen their M&E capacity. The unit targets 'high-intensity' support to Uganda and Egypt and various other types of support to an additional 32 countries. This paper from the World Bank collates some of the main lessons learned from ECD activities and outlines major issues to be addressed.

Strengthening national statistical capacity

Good statistics are vital for the effective monitoring of development programmes and strategies, and ultimately to support evidence-based policymaking. However, many developing countries lack the institutional capacity or effective systems for gathering data. Many existing statistical systems were predominately designed to meet immediate rather than long-term data needs, and therefore lack coordination.

There is increasing recognition of the need for a strategic approach to statistical capacity development, particularly following the Second International Roundtable on Managing for Development Results in February 2004 and the resulting Marrakech Action Plan for Statistics (MAPS). This emphasised the need to develop National Strategies for the Development of Statistics (NSDSs).

Paris21 Secretariat, 2004, 'A Guide to Designing a National Strategy for the Development of Statistics (NSDS)', Prepared by the Partnership in Statistics for Development in the Twenty-first Century (PARIS 21), Paris
http://www.gsdrc.org/go/display&type=Document&id=2650
There is increasing awareness of the need to strengthen statistical capacity to support the design, monitoring and evaluation of national development plans. National Strategies for the Development of Statistics (NSDS) are designed to achieve this goal. This guide aims primarily to assist developing countries to design their NSDSs, but will also be useful to development partners.

Applying monitoring and evaluation tools

Monitoring and evaluation tools and approaches are applied in a range of programmes and contexts. This section provides a guide to key resources on using monitoring and evaluation tools in relation to specific governance and social development programmes.

Evaluating governance programmes

To date, most impact evaluations have focused on economic and social development programmes in the fields of health, nutrition, water, agriculture and education. Relatively few impact evaluations have been conducted on governance programmes. This is due to several factors. Governance programmes are complex and combine a number of objectives, many of which may be difficult to measure. There is also a shortage of reliable baseline data for many governance interventions. Despite these problems, experimental methods such as Randomised Control Trials have been used to evaluate corruption, community development and election programmes. Where it is not possible to use experimental methods, evaluators should consider using quasi-experimental or mixed methods approaches.

Garcia, M., 2011, 'Micro-Methods in Evaluating Governance Interventions', Evaluation Working Papers, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn
http://www.gsdrc.org/go/display&type=Document&id=4101
Although billions of dollars have been invested on improving governance in developing countries in the past decade, few of the programmes that have received funding have been subjected to rigorous impact evaluation. The aims of the paper are to: (i) discuss the challenges in evaluating governance programmes, (ii) identify some aspects of governance that have been rigorously evaluated, and finally (iii) provide practical recommendations based from previous evaluation experience.

In recent years, a growing number of studies have used experimental methods to assess the impact of a range of governance programmes and processes. The most common sub-sectors for RCTs are elections, community-driven development and service delivery. Although there are a number of challenges associated with using RCTs to measure the impact of governance programmes, these experiments also present 'immense opportunities'.

Moehler, D., 2010, 'Democracy, Governance and Randomised Development Assistance', Annals of the American Academy of Political and Social Science, vol. 628, no. 1, pp.30-46
http://www.gsdrc.org/go/display&type=Document&id=4097
Are field experiments useful in studying the impact of development assistance on democracy and governance (DG)? This article reviews recent and ongoing DG field experiments. It considers the prospects for and obstacles to the development of a useful body of experimental evidence on the political economy of development. There are significant challenges related to the difficulty of generalising from small samples and micro-level projects. However, although the field experiments have tended towards village-level interventions, they suggest that impact evaluations can potentially address higher-level interventions and theories.

E-Gap is a research network for scholars and practitioners engaged in field experiments on topics of governance, politics and institutions. Its website provides news and information on current research in this area.

Governance indicators

One of the most contentious debates surrounding the M&E of governance programmes is the use of governance indicators. Increasing emphasis on the need to measure 'good governance' and how it relates to poverty reduction has led to a proliferation of worldwide data sets, guidelines and frameworks for assessment. There is considerable debate about the validity of different methodological approaches to measurement, and increasing recognition that measuring governance is itself a political process.

UNDP, 2007, 'Governance Indicators: A User's Guide', Second Edition, United Nations Development Programme, Oslo
http://www.gsdrc.org/go/display&type=Document&id=1155
Demand for ways to measure different aspects of democracy, human rights and governance is increasing. This has led to a rapid growth in the number of indicators to assess government performance, the quality of institutions and people's perceptions. This guide offers advice on where to find and how to use governance indicators.

When used carelessly, governance indicators can 'misinform and mislead'. It is important to recognise that governance indicators are used for a range of purposes and can be distinguished according to their source and objectivity, level of impact, comparability, degree of aggregation and 'actionability'. The following paper presents several pitfalls associated with the use of governance indicators and proposes guidelines to ensure more careful use. It highlights the need for providers of indicators to be subject to greater transparency, scrutiny and peer review.

Williams, G., 2011, 'What Makes a Good Governance Indicator', Policy Practice Brief 6, The Policy Practice
http://www.gsdrc.org/go/display&type=Document&id=4035
What are the most common pitfalls associated with the use of governance indicators? How can these be avoided? This brief examines the different types of governance indicator, their use and their misuse. Users need to ask critical questions about what exactly indicators are measuring, and to remember the limited ability of broad governance indicators to demonstrate causality. Process indicators are most relevant to understanding causal mechanisms. Providers of indicators should ensure full transparency in the publication of methodology and source data, and should be subject to greater scrutiny, evaluation and peer review.

Much of the debate around governance indicators has focused on country-level governance monitoring and assessments. Relatively few studies have addressed the question of which kinds of governance indicators are most useful for governance programming. Meso-level indicators that link project/programme results to higher-level impacts have been largely absent, and macro-level indicators have sometimes been used by evaluators to judge projects.

In response to these issues, a number of agencies have sought to develop 'improved' indicators, by ensuring that these indicators are either more linked to concrete problems or based on a more participatory approach. DFID has recently commissioned an assessment of existing programme-level governance and conflict indicators, (Barnett et al., 2011), which includes a revised list of suggested indicators.

GSDRC, 2010, 'Evaluation of Governance Programme Indicators', Helpdesk Research Report, Governance and Social Development Resource Centre, Birmingham
http://www.gsdrc.org/go/display&type=Helpdesk&id=608
While there is a large body of literature providing advice in selecting and designing indicators, there is little evaluating governance programmes indicators per se. In general, the literature criticises the application of quantitative indicators to governance programmes, or criticises the confusion on which aspects to monitor. Recently, there has been a growth in 'improved' indicators. There is a debate surrounding their validity, however. Participatory and mixed method approaches seem to respond best to criticism of governance assessments.

Barnett, C., et al., 2011, 'Governance and Conflict Indicators Report', ITAD
http://www.gsdrc.org/go/display&type=Document&id=4113
How can a robust and clear set of indicators be established to monitor effectively the progress of a programme? Which indicators show best that a programme has achieved what it set out to achieve? This study tests the relevance and robustness of a list of indicators for the Department for International Development's (DFID's) conflict and governance programmes. The list consists of fifteen separate sets, or suites, of outcome and output indicators and covers a spectrum of related programme areas, including security and justice, elections, civil service reform and corruption. It suggests that a good results framework that enables programme progress to be effectively monitored and explained is critical, particularly in times of financial austerity.

In another important initiative, the World Bank has recently sought to catalogue those governance indicators that can be used to measure concrete governance issues that can be addressed directly in governance projects and programmes. The AGI Data Portal consolidates information on actionable governance indicators, provides a one-stop-shop platform to navigate these indicators and their documents, and offers customised tools for data management, analysis and display. Another useful resource is the UNDP's Governance Assessment Portal, which provides a database on country-level governance assessments, a library of publications on governance assessment and a toolbox of frameworks and indicators.

Further discussion of governance assessments and governance indicators can be found in the **Political Economy Analysis Topic Guide**: http://www.gsdrc.org/go/topic-guides/political-economy-analysis/governance-assessment

M&E in fragile and conflict-affected states

M&E faces unique challenges in fragile and conflict-affected contexts. These include:

- Lack of security (both for researchers and the people who talk to them)
- Distrust of outsiders and reluctance to talk, or fear of reprisals for talking
- Shame in acknowledging sexual violence
- Rapidly fluctuating populations (due to migration) and a lack of adequate documentation on the population
- Lack of adequate statistical data to generate baselines

Government unwillingness to recognise the level of violence or the failure of programmes which might in turn affect aid flows.

Elkins, C., 2006, 'Monitoring and Evaluation (M&E) for Development in Peace-Precarious Situations', RTI International, London
http://www.gsdrc.org/go/display&type=Document&id=2664

How can M&E information systems improve programme impact and assist peaceful development in situations prone to violent conflict? This paper outlines M&E's status as a unique discipline and describes M&E strategies and tactics implemented in real-world, peace-precarious situations. Even under the stresses of violence and conflict, M&E approaches can help build knowledge of how to advance peace and development.

GSDRC, 2007, 'M&E in Fragile States', GSDRC Helpdesk Research Report, Governance and Social Development Resource Centre, University of Birmingham
http://www.gsdrc.org/go/display&type=Helpdesk&id=399
There is a general consensus on the need to move M&E beyond the project level to the sector and strategic level in fragile states. It is important to assess the overall impact of interventions on stabilising fragile states and promoting sustainable peace. Joint evaluations and agreed upon objectives, among government departments (in the case of 'whole-of-government' approaches) and with other governments and relevant actors, are useful in providing this comprehensive picture.

Conflict prevention, conflict resolution and peacebuilding all aim to have long-term impact in terms of sustainable peace and development. Given the time-frames and the convergence of a multitude of activities in conflict-affected areas, it may be difficult to attribute quantifiable results to specific conflict prevention, conflict resolution and peacebuilding activities. This should not, however, be a deterrent, or an excuse for not conducting assessments.

There is consensus that many of the standard criteria for monitoring and evaluation (M&E) are applicable to interventions in conflict-affected areas. In addition, the OECD-DAC provides a series of common monitoring and evaluation principles that can be applied to different types of conflict prevention and peacebuilding interventions. These are: inclusiveness; the testing of underlying theories of change; the use of qualitative and quantitative methods; testing assumptions and learning; ethical standards for approaching informants and handling the reporting of findings.

The following are a selection of M&E frameworks, tools and recommendations developed by international organisations, donor agencies, academics and non-governmental organisations. They cover conflict prevention, conflict resolution, conflict transformation, stabilisation, and peacebuilding interventions.

Conflict prevention and peacebuilding

OECD-DAC, 2007,'Encouraging Effective Evaluation of Conflict Prevention and Peacebuilding Activities: Toward DAC Guidance', A Joint Project of the DAC Network on Conflict, Peace and Development Co-operation and DAC Network on Development Evaluation, Paris
http://www.gsdrc.org/go/display&type=Document&id=2958
The past decade has seen growing numbers of governments and organisations devote resources to interventions intended to avert or end conflict. How can these be evaluated? This report represents a step in the development of practical evaluation guidance. Scope, conflict analysis, impacts, skills and tools all need to be considered. Next steps should include donor experimentation with evaluations, an annual review of evaluations and a policy review.

Scheers, G., 2008, 'Assessing Progress on the Road to Peace Planning, Monitoring and Evaluating Conflict Prevention and Peacebuilding Activities', Issue Paper, no. 5, European Centre for Conflict Prevention
http://www.gppac.net/uploads/File/Programmes/PM_L/Issue%20paper%205.pdf

Conflict resolution, conflict transformation and peacebuilding

Çuhadar-Gürkaynak, E. , Dayton, B., and Paffenholz, T., 2009, 'Evaluation in Conflict Resolution and Peacebuilding', in Handbook of Conflict Analysis and Resolution, eds., D. J. D. Sandole, S. Byrne, I. Sandole-Staroste, and J. Senehi, Routledge, Oxon and New York, pp. 286-299
http://www.gsdrc.org/go/display&type=Document&id=3660
Why is evaluation essential in conflict resolution and peacebuilding work? How can traditional evaluations be adapted for this purpose? This chapter examines the difficulties and possibilities of evaluating conflict resolution and peacebuilding initiatives. Renewed attention to evaluation strengthens connections between peacebuilding theory and practice.

Lanz, D., Wählisch, M., Kirchhoff, L., and Siegfried, M., 2008, 'Evaluating Peace Mediation', Proposal of a General Framework for Evaluating International Mediation Activities, Initiative for Peacebuilding (IfP) - Mediation Cluster
http://www.gsdrc.org/go/display&type=Document&id=3470
How can accountability mechanisms be established for international peace mediation given that it takes place in extremely complex contexts and its contributions are difficult to grasp? This study has developed a framework for evaluating international mediation activities that differ from standard methodologies. It proposes a series of non-suggestive evaluation questions that allow a systematic but flexible assessment of aspects of peace mediation.

De Coning, C. and Romita, P. 2009, 'Monitoring and Evaluation of Peace Operations' International Peace Institute, New York
http://www.gsdrc.org/go/display&type=Document&id=3777
It is difficult to make causal linkages between inputs and outcomes due to the complex nature of conflict. Donor countries and NGOs would do better to focus instead on the contributions particular activities make towards a desired outcome. Sustainable DDR requires achievements in other areas. DDR evaluation should look across a variety of sectors and consider the political dynamics that affect DDR processes. Identifying M&E benchmarks and indicators can help reconcile competing perspectives of strategic decisions. Other suggestions to improve M&E are: a web-based database of M&E lessons learned; a network to debate M&E issues; policy discussions on M&E at high profile discussion-making forums; and mechanisms to ensure that M&E results are incorporated into decision-making processes.

Church, C. and Rogers, M. M., 2006, 'Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs', Search for Common Ground, Washington DC
http://www.sfcg.org/programmes/ilr/ilt_manualpage.html

Stabilisation

Cohen, C., 2006, 'Measuring Progress in Stabilisation and Reconstruction', Special Report, United States Institute of Peace, Washington DC
http://www.gsdrc.org/go/display&type=Document&id=3663
Measuring progress is essential to the success of stabilisation and reconstruction efforts. This report examines the shortcomings of current processes, including the tendency to measure implementation rather than impact. Proper assessment tools and reliable measures of progress are needed to enable policymakers to take stock of the challenges before intervening and to continuously track the progress of their efforts towards stabilisation. Political will is also essential to ensure leadership and cooperation across organisational boundaries

For more detailed guidance on measuring and assessing fragility, see the GSDRC's **Fragile States Topic Guide**:
http://www.gsdrc.org/go/fragile-states/chapter-3--measuring-and-assessing-fragility/measuring-fragility

M&E of empowerment and accountability

Donors have been increasingly concerned with empowering poor people in their economic, social and political relationships, and ensuring that their programmes (and the governments that they support) are accountable to beneficiaries. These processes are complex and can be difficult to measure. There is a tension for donors between the need to pursue a flexible approach to supporting empowerment processes and the need to demonstrate results and accountability to their constituents. Combining quantitative and qualitative methods and using locally-generated indicators can help to overcome some of these issues.

A growing number of studies have sought to use Randomised Control Trials to measure empowerment and accountability processes and programmes. The following report outlines some of the key challenges associated with using RCTs for these programmes and notes that scholars are divided about the ability of these methods to generate reliable results.

GSDRC, 2011, 'RCTs for Empowerment and Accountability Programmes', Helpdesk Research Report, Governance and Social Development Resource Centre, University of Birmingham
http://www.gsdrc.org/docs/open/HD756.pdf
This report examines the extent to which RCTs have been used successfully to measure empowerment and accountability processes and programmes. Field experiments present immense opportunities, but they are more suited to measuring short-term results with short causal chains and less suitable for complex interventions. The studies have also demonstrated divergent results, possibly due to different programme designs. The literature highlights that issues of scale, context, complexity, timeframe, coordination and bias in the selection of programmes also determine the degree of success reported. It argues that researchers using RCTs should make more effort to understand contextual issues, consider how experiments can be scaled up to measure higher-order processes, and focus more on learning. The report suggests strategies such as using qualitative methods, replicating studies in different contexts and using randomised methods with field activities to overcome the limitations in the literature.

McGee, R., and Gaventa, J., 2011, 'Shifting Power? Assessing the Impact of Transparency and Accountability', Working Paper No. 383, Institute of Development Studies, Brighton
http://www.gsdrc.org/go/display&type=Document&id=4209
What does impact mean in relation to accountability programmes and projects? This paper argues that current approaches to impact assessment in this field are inadequate: methodological wars are overshadowing key issues of power relations and politics. A learning approach to impact assessment is needed that gives power and politics a central place in monitoring and evaluation systems. Instead of looking at the extent to which the desired impact was achieved, it is important to look at what happened as a result of the initiative, how it happened and why. It is also important to test and

revise assumptions about theories of change continually and to ensure the engagement of marginalised people in assessment processes.

A more detailed discussion of measuring impact and change in empowerment and accountability programmes is provided in the '**Empowerment and Accountability' Topic Guide:**
http://www.gsdrc.org/go/topic-guides/empowerment-and-accountability/measuring-empowerment-and-accountability

M&E of policy influence

Influencing policy has become increasingly central in the development strategies of aid agencies and NGOs. Monitoring advocacy is challenging, however, because these interventions are highly complex and involve a range of interacting forces and actors. Key responses to these challenges include developing a theory of change, user surveys and collecting significant change stories.

Jones, H., 2011, 'A Guide to Monitoring and Evaluating Policy Influence', Background Note, Overseas Development Institute, London
http://www.gsdrc.org/go/display&type=Document&id=4096
Is it possible to monitor and evaluate effectively efforts to influence policy? This paper provides an overview of approaches to the monitoring and evaluation (M&E) of policy influencing activities. It suggests that while M&E in this field is challenging, information can be generated that can be used to improve programmes and provide accountability for funds. The key is for policy influencing teams to recognise the value of M&E to their work and to incorporate it into their practices from the beginning of a project or programme.

M&E in governance and social development sub-sectors

For brief overviews of some of the specific issues associated with M&E in key governance and social development sub-sectors, see the following sections in other GSDRC Topic Guides.

Monitoring and Evaluating Institutional Change

The European Foundation for Quality Management (EFQM) Excellence Model
http://www.efqm.org/en/tabid/132/default.aspx
DFID promotes the use of the European Foundation for Quality Management (EFQM) Excellence Model, which provides a framework for assessing the aspects of performance that make an organisation successful.

The model was designed to be a generic model and is as applicable to 'non-business' oriented organisations such as government departments and agencies and non-profit organisations as to private sector organisations. It recognises that there are many approaches to achieving sustainable excellence in organisations, and provides considerable freedom for interpretation to reflect strategies which are appropriate for an individual public sector entity given where it has come from, the cultural climate, country diversity, modernisation progress, and political climate. It can be used to help organisations recognise institutional issues that must be addressed as well as provide measures for evaluating the impact of change on performance.

Monitoring and Evaluating Human Rights Programmes

A key concern is assessing the intentional and unintentional effects (positive or negative) of a human rights programme. This has been developed further by approaches which use impact assessments to monitor and evaluate the impact of human rights interventions and/or the human rights components of interventions. Thinking about impact is also thought to contribute to awareness about human rights concerns, and therefore facilitate the integration of rights-based principles into the overall policy process.

Department for International Development, 2009, 'How to Note: A Practical Guide to Assessing and Monitoring Human Rights in Country Programmes', DFID, London
http://www.stabilisationunit.gov.uk/stabilisation-and-conflict-resources/thematic/doc_details/303-how-to-note-a-practical-guide-to-assessing-and-monitoring-human-rights-in-country-programmes-.html

Harrison, J., 2010, 'Measuring Human Rights: Reflections on the Practice of Human Rights Impact Assessment and Lessons for the Future', Warwick School of Law Research Paper No. 2010/26, University of Warwick
http://www.gsdrc.org/go/display&type=Document&id=4226
This study examines the practice of Human Rights Impact Assessment (HRIA). It argues that, while a uniform HRIA process in all fields will not be appropriate, eight core elements represent the 'minimum core' of a valid process: screening, scoping, evidence gathering, consultation, analysis, producing policy-oriented recommendations, publication, and monitoring and review. Overall, better performance monitoring is crucial, and the people undertaking HRIAs need a deep

understanding of human rights. More reflection is also required on the connection between HRIAs and the people affected by the policies and practices that HRIAs seek to influence.

NORAD, 2001, 'Handbook in Human Rights Assessment. State Obligations, Awareness and Empowerment', Norwegian Agency for Development Cooperation, Oslo
http://www.gsdrc.org/go/display/document/legacyid/1581
How do you enhance the human rights profile of development programmes? How do you identify the need for human rights impact analysis? This handbook assists the user in addressing human rights concerns by recording the potential, planned or likely positive or negative effects of the programme under review. It is not a manual on how to conduct a full-scale human rights impact analysis but a guide to identifying the need for such analysis.

UNHCHR, 2003, 'Human Rights-Based Reviews of UNDP Programmes. Working Guidelines'. Draft, UNDP, Geneva
http://www.gsdrc.org/go/display/document/legacyid/1788
The UN operates a Common Understanding of the human rights-based approach to development (HRBA). This paper sets out working guidelines for a human rights-based review of UNDP country offices and projects based on the Common Understanding. The guidelines aim to support reviews at each phase of programming, strengthen existing activities and assist in the design of new programmes from a human rights perspective.

Monitoring and Evaluating Justice Interventions

Where is a good place to start?

Vera Institute of Justice, 2003, Measuring Progress toward Safety and Justice: A Global Guide to the Design of Performance Indicators across the Justice Sector, Vera Institute of Justice.
http://www.gsdrc.org/go/display/document/legacyid/843
How can progress towards delivering safety, security and justice be measured? What are the key steps in designing a system of indicators to gauge the impact of sector reforms? This guide sets out a process that can be used to develop appropriate indicators. It is intended as a practical tool for use around the world.

Carothers, T., 2003, Promoting the Rule of Law Abroad: The Problem of Knowledge, Carnegie Endowment for International Peace
http://www.gsdrc.org/go/display/document/legacyid/448
The causal relationship between rule of law promotion and economic development and democracy is not as direct as it might appear but remain donors' twin rationale for these activities. Is there a problem of knowledge concerning rule of law projects? What should practitioners know when they promote these activities?

Legal and judicial sector

These documents provide guidance on evaluations that relate specifically to the legal and judicial sector.

World Bank, 2002, Legal and Judicial Sector Assessment Manual, The World Bank, Washington DC
http://www.gsdrc.org/go/display/document/legacyid/1038
How can the rule of law be achieved? What is the best way to reform a judicial and legal system? This manual, from the World Bank Legal Vice Presidency provides a tool to carry out a diagnosis of a country'slegal and judicial system prior to reform. The promotion of the rule of law is distinct from other public sector reform. It requires long-term commitment, and must be approached strategically, with a holistic and complete reform programme.

Messick, R., 2001, 'Key functions of legal systems with suggested performance measures,' draft paper hosted on the World Bank 'Legal Institutions of the Market Economy.'
http://www.gsdrc.org/go/display/document/legacyid/282
This paper attempts to help reformers devise evaluation techniques that can be adapted to specific situations in different countries. The first section asks how well the legal system as a whole is performing its four key functions: the deterrence of wrongful conduct, the facilitation of voluntary transactions, the resolution of private disputes, and the redress of the abuse of government power. The second section attempts to assess how well the key institutions of the juridical system, such as the courts, the private bar, and the public prosecutors, are working.

Hammergren, L., 'Diagnosing Judicial Performance: Toward a Tool to Help Guide Judicial Reform Programs', paper prepared for Transparency International 9th International Anti-Corruption Conference.
http://www.gsdrc.org/go/display/document/legacyid/440
This paper provides a checklist for diagnosing judicial performance. It aims to evaluate the transparency and related aspects of judicial performance and is intended to promote reform programmes. The list is composed of the characteristics believed to be critical in producing the desired patterns of behaviour. It is intended to be applied globally, and was not

written with any specific legal system or tradition in mind. It aims to capture universal factors that will help identify real or potential problems in judicial operations.

GSDRC, 2009, 'Institutional Assessment of the Police and Justice Sectors', Helpdesk Research Report, GSDRC, Birmingham
http://www.gsdrc.org/go/display&type=Helpdesk&id=515
This report includes coverage of a number of donor designed frameworks for assessing the policing and justice sector. According to much of the general academic and policy literature on SSAJ programmes, substantial reform of the police force is only possible when reform of the justice system is administered at the same time. However, whilst the underlying principles for the institutional assessment of policing and justice may be similar, the specific frameworks espoused by donors appear to tackle the institutional assessment of policing and justice separately.

Monitoring and Evaluating Climate Change Adaptation

Measuring the effectiveness of adaptation projects, programmes, policies and national systems is inherently complex. To begin with, there remains a great deal of conceptual uncertainty about what to measure (adaptive capacity, resilience, vulnerability reduction etc.). Adaptation interventions tend to cut across many sectors, are implemented at different scales (from international to household level), over different timescales, and take a broad range of approaches (from hard structural adaptation measures, e.g. infrastructure and technological projects, to soft policy measures e.g. information exchange and behavioural change). Thus, a range of different approaches are needed depending on where interventions sit on the development – adaptation continuum.

Adaptation Indicators may be process-based (to measure progress in implementation) or outcome-based (to measure the effectiveness of the intervention). Developing indicators at the project or programme level is relatively straightforward, as many projects are undertaken within sectors where established monitoring and evaluation systems with proven indicators already exist. However, monitoring and evaluation of policies and national systems is more complex as it requires strong coordination across sectors and levels and is more susceptible to external factors. There are additional challenges with regards to attributing cause and effect in adaptation interventions and accounting for unintended consequences. Practical difficulties in undertaking assessments stem from a general lack of financial, human and technical resources and capacities, a lack of baseline data and historical trends, uncertainty of projected climate change impacts, and insufficient sharing of information across stakeholder groups, levels and sectors. As a result, monitoring and evaluation (M&E) of adaptation is one of the weakest areas of adaptation practice. Of those evaluations carried out to date, most have been undertaken as part of ongoing implementation, whilst only a few have focused on evaluating interventions after completion.

Given this panorama, there are increasing calls for an integrated M&E framework for adaptation which is more closely aligned with development planning, through, for example, the incorporation of adaptation M&E into existing national poverty reduction frameworks such as Poverty Reduction Strategy Papers (PRSPs) and sectoral plans (see national-level adaptation planning). This would enable adaptation interventions to make use of existing monitoring and evaluation systems rather than create an additional layer of reporting. There are also calls to incorporate M&E approaches from the field of disaster risk reduction (DRR) given that many of the existing DRR indicators and data are relevant for adaptation.

Spearman, M. and McGray, H., 2011, 'Making Adaptation Count: Concepts and Options for Monitoring and Evaluation of Climate Change Adaptation', Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Eschborn, Germany
http://www.gsdrc.org/go/display&type=Document&id=4203
This publication offers guidance for designing M&E systems for climate change adaptation. It argues that M&E systems need to enable results-based management, promote flexibility, and support iterative learning. Achieving these goals requires development practitioners to carefully articulate their adaptation objectives, clarify the basis for their project design, and make their assumptions transparent. With this foundation, project managers can select indicators and build information systems that are able to track adaptation success.

Hedger, M., Horrocks, L., Mitchell, T., Leavy, J. and Greeley, M., 2008 'Evaluation of Adaptation to Climate Change from a Development Perspective', Desk Review, Institute of Development Studies, Brighton
http://www.gsdrc.org/go/display&type=Document&id=3961
This paper reviews the current state of the evaluation of climate change adaptation interventions (CCAI). It finds that while development agencies are scaling up the funding and delivery of adaptation interventions, few systematic assessments of CCAI have been undertaken. The authors propose a pyramid of indicators which might provide a framework to measure the accumulation and culmination of effort at multiple levels. This allows for a variety of monitoring and evaluation tools to cope with the complexities of CCAI and to improve the overall quality of assessments. Five key factors for successful adaptation – effectiveness, flexibility, equity, efficiency and sustainability – will need to be reflected in indicators.

UNFCCC Secretariat, 2010, 'Synthesis report on efforts undertaken to monitor and evaluate the implementation of adaptation projects, policies and programmes and the costs and effectiveness of completed projects, policies and programmes, and views on lessons learned, good practices, gaps and needs', UNFCCC, Bonn

How effective is the monitoring and evaluation (M&E) of adaptation implementation by the EU and its member states? This report examines the range of M&E efforts being undertaken and finds that some areas, such as the monitoring of policies, are more advanced than others. Furthermore, the majority of adaptation projects are still under development or implementation. This is an evolving area and further focus is needed on issues such as defining appropriate climate change indicators and metrics.

Frankel-Reed, J., Brooks, N., Kurukulasuriya, P. and Lim, B., 2009, 'A Framework for Evaluating Adaptation to Climate Change', Chapter 18 in R. Van Den Berg and O N. Feinstein (eds.) Evaluating Climate Change and Development, Transaction Publishers, New Brunswick, New Jersey, pp. 285-298
How can climate change adaptation be tracked in a structured way? Standard development and environment indicators are unable to reflect the nature of adaptation, which is about capacity, behaviour and risk-reducing measures for the advancement of development outcomes. This chapter presents seven considerations for establishing adaptation monitoring and evaluation (M&E) approaches, and outlines an M&E framework for adaptation – the outcome of applying these considerations in the UNDP context.

Prowse, M. and Snilstveit, B., 2010, 'Impact Evaluation and Interventions to Address Climate Change: A Scoping Study', The International Initiative for Impact Evaluation (3ie), New Dehli
The selection and design of climate change mitigation and adaptation interventions should be based on evidence of what works (and what does not), under what circumstances and at what cost. Currently, evidence on the impact of such interventions appears limited, and there is a strong case for the wider application of rigorous impact evaluation. New studies should evaluate positive and negative impacts of climate change interventions on both environmental and welfare outcomes. Programme planners and evaluators should work together to accommodate rigorous impact evaluation from the start. While appropriate outcome indicators will differ between interventions, future evidence syntheses will be improved by work to develop a consensus on a set of common outcome indicators.

Monitoring and Evaluating Social Protection Programmes

Due to the diverse nature of social protection instruments, available guidance focuses on individual instruments rather than addressing social protection as a whole. Impact evaluations are ideal but have large data collection and analysis requirements.

International Labour Organisation (ILO), 2007, 'Health Microinsurance Schemes: Monitoring and Evaluation Guide, Volume 1: Methodology', Strategies and Tools against Exclusion and Poverty Programme (STEP), International Labour Organisation (ILO), Geneva
What are the required capacities and key indicators of a viable health microinsurance scheme (HMIS)? This guide provides a tool to help in overcoming the lack of evaluation data on health microinsurance schemes. In countries with low levels of health insurance coverage, many health microinsurance schemes designed to reach the poor are emerging. There is growing recognition that health microinsurance schemes constitute a complementary and valuable strategy in extending social security.

Robalino, D., et al., 2009, 'Ex-Ante Methods to Assess the Impact of Social Insurance Policies on Labor Supply with an Application to Brazil', Social Protection Discussion Paper No. 0929, World Bank, Washington DC
This study has developed a behavioural model to assess how changes in the rules of pensions and unemployment benefit systems could affect savings rates, the share of time that individuals spend outside of the formal sector, retirement decisions, and system costs. Key parameters are: (1) preferences regarding consumption and leisure; (2) preferences regarding formal versus informal work; (3) attitudes towards risks; (4) the rate of time preference; and (5) the distribution of an outside shock that affects movements in and out of the social insurance system, given individual decisions. Simulations suggest, among other findings, the importance of joint policy analysis of unemployment benefits and pension systems.

Paes-Sousa, R. and Santos, L. M. P., 2009, 'Measuring the Impact of Bolsa Familia Program Based on Data from Health and Nutrition Days (Brazil)', Working Paper No.7, Iniciativa Latina y Caribe Sin Hambre, Food and Agriculture Organisation of the United Nations, Santiago
Does Brazil's Bolsa Familia conditional cash transfer programme reduce children's malnutrition and food insecurity? This study assesses the programme's impact on the nutritional status of zero to five year olds. Data on 22,375 children's height/age, weight/age and weight/height shows that the PBF does improve child nutrition. To ensure an increase in

beneficiaries' health levels, families need greater access to goods and services which interact with improved nutrition. The provision of more and better basic services and initiatives for inclusion in the labour market would ensure the PBF's effectiveness.

Card, D., Ibarrarán, P. and Villa, J. M., 2011, Building in an Evaluation Component for Active Labor Market Programs: a Practitioner's Guide', Inter-American Development Bank, Washington DC
http://www.iadb.org/document.cfm?id=36479813