DLP

Policy and Practice for Developmental
Leaders, Elites and Coalitions

DEVELOPMENTAL LEADERSHIP PROGRAM

Background Paper **11**

The Evaluation of Politics and the Politics of Evaluation

Chris Roche, Oxfam Australia
Linda Kelly, Praxis Consulting
August 2012

DLP

Policy and Practice for Developmental
Leaders, Elites and Coalitions

DEVELOPMENTAL LEADERSHIP PROGRAM

Contents

Executive Summary

The Developmental Leadership Program (DLP) argues that the real binding constraints on development are 'political' rather than technical. Donors therefore need to understand, engage with and influence political processes.

This raises many challenges for donor staff and systems, especially the challenge of how to assess programs that are thinking and working politically. The challenges are magnified because political change is likely to be long-term, with intermediate steps that are important, but possibly intangible - such as changes in processes, relationships and power. The context for such change is also likely to be dynamic and complex and it will therefore be difficult to draw a simple linear attribution between donor inputs and the results achieved.

This paper proposes that using a mixture of assessment methods, with an appreciation of the strengths and limitations of each, provides the most effective solution to the assessment challenge. For example, utilising quantitative methods such as randomised controlled trials can contribute to an understanding of the causal effects of a limited number of variants of an intervention over relatively short time periods. Qualitative methods such as outcome mapping or action-reflection approaches can help explain *why* change happened and *how* donor inputs influenced change or supported key actors in this situation. Utilised together, with an understanding of the value of the data that they produce, this mixture of methods can provide the donor with a more comprehensive picture of the changes that have happened, and the reasons for those changes, across short, medium, and longer term time frames. This is particularly the case for more complex programmes involving non-linear changes, as a mixed method approach allows us to understand these changes from different perspectives and therefore allows for a more rounded view.

However, the application of mixed methods requires skill and care. Good quality performance assessment starts with asking the right questions, utilising an appropriate mix of good quality data-collection methods, and then bringing diverse information together. Particular skill is required to effectively synthesise and analyse the data.

Donors have been reluctant to utilise a mixed methods approach. In part, we argue that this is because development practitioners are poorly informed about various methodologies and their limitations. But we would also argue that evaluation itself is a political process. Political interests, such as the current donor focus on short-term results, can influence the assessment processes of development agencies and skew their focus towards the use of single assessment methods. This has several implications for the quality of data and assessment. A discussion to be continued in a later paper.

Introduction

The aim of this paper is to contribute to the argument for a 'mixed methods' approach to monitoring and evaluation, particularly for programs that seek to 'work politically'[1]. We suggest that this is particularly relevant where the object of evaluation is a process or a practice that is intended to change power relations, rather than the more easily quantifiable measurement of the delivery of a service (such as clean water) or specific items (such as kilometres of road or bed-nets). We argue that different policy, program and operational objectives require different methods of evaluation and, often, mixed or combined ones, depending on the range of objectives and the intended sequence of outcomes. The paper outlines a range of these methods and illustrates their usefulness for the evaluation of different processes and interventions. It also suggests that part of the explanation for the resistance by parts of the development community to some forms of evaluation, as discussed in this paper, can be attributed to the political imperatives of donors and other development organizations. This proposition will be taken up more fully in a second paper to be published by DLP later in 2012.

The paper is intended for policy makers and operational staff who may be involved not only in the design of projects and programs, but also in deciding what kind of monitoring and evaluation components to commission for those programs. Much of this paper is drawn from the material presented to, and discussed at, a specialist DLP workshop held in Canberra on "Innovations in Monitoring and Evaluation as if Politics Mattered", on 17-18th October 2011.[2]

---

1   For a fuller discussion of 'working politically' see Leftwich (2011).

01

Context

The **Developmental Leadership Program (DLP)** addresses an important gap in international thinking and policy about the critical role played by leaders, elites and coalitions in the politics of development. This developing program brings together government, academic and civil society partners to explore the role of **human agency** in the processes of development.

At the core of DLP thinking is the proposition that the real binding constraints on development are political – understood in the broadest sense – rather than technical, and that political processes shape developmental outcomes at all levels, nationally and sub-nationally, and in all sectors and issue areas. Accordingly, DLP seeks to address the policy, strategic and operational implications of 'thinking and working politically' – for example, how to help key players solve collective action problems, forge developmental coalitions, negotiate effective institutions and build stable states.

Two recent DLP workshops that brought together researchers, policy-makers and practitioners,[3] identified a number of implications for International Aid agencies. These include the following.

- If the nature and quality of leadership matters for shaping institutions and developmental outcomes, then this will require **much longer-term strategies** (in the order of 10-20 years) than those usually adopted by donors.

- **All aid is political.** No matter how neutral assistance appears to be, it will have a political impact in the recipient country and this needs to be recognized.

- Therefore, **donors and their intermediaries need a deep understanding of the political context** and how to frame development issues in order to facilitate and support local developmental leaderships and coalitions at all levels and in all sectors and issue areas.

- **Do no harm.** Aid should not support the political status quo if it is considered to be anti-developmental.

- Donors can play an important role in facilitating and brokering the emergence of developmental coalitions, but **processes and practices need to be locally led.**

- **Aid programs need to be more flexible.** Critical junctures, windows of opportunity and processes of political change do not align easily with rigid time-bound project designs or log-frames. There should always be room for long-term and slowly disbursed funding, as well as a capacity to respond quickly when opportunities open up.

- **Staff within donor agencies** need to be better equipped and trained so as to ensure better political analytical skills and experience, and to promote experimentation with new programs.

- **This means moving beyond political economy,** with its focus on formal institutions and incentives, to a much more nuanced **political analysis** and appreciation of the structures and distribution of power (both formal and informal), the key players in any sector or issue area and their relationships, and the nature of the political practices and ploys used within or beyond the rules of the game.

---

3   For more detail see Leftwich & Wheeler (2011) and Developmental Leadership Program (2012)

DLP research findings to date include the following.

1. Development is a political process in which leadership and agency matter, although development is clearly not achieved through politics alone.

2. Effective and lasting institutional and policy innovation and change is only achieved and sustained where it is locally appropriate and legitimate, and achieved through endogenous political processes.

3. Coalitions of a variety of kinds play a particularly important and often under-recognized role in institutional formation (or the 'rules of the game'), policy reform, and development processes generally. Prior networks can be very important in facilitating the emergence of such coalitions.

4. 'Critical junctures', 'windows of opportunity' and 'trigger events' often provide opportunities for developmental change. Seizing such moments for reform can be critical. What matters is whether leaders have the capacity to respond to these opportunities.

5. Civil society and other coalitions need to know - and be able to access - key gatekeepers, connectors and 'champions' in positions of power and influence. Informal networks help.

6. Framing of objectives can be critical for success. How leaders and developmental coalitions 'frame' their campaigns, strategies and communication can turn out to be very important in building or undermining support. Likewise, the way donors and external agencies frame their support matters too.

7. Donors need the analytical skills and political understanding to be able to recognise the opportunities for developmental change and to be astute about what can be done in support of reform or developmental leaders and coalitions. Without this understanding, international aid can bolster harmful political settlements.

8. A lack of understanding of political context by international donors compromises their ability to support development effectiveness, nationally and sub-nationally, as well as within sectors and issue areas. In particular, it is crucial that they understand the 'micro-politics' of a given situation – what the formal and informal institutional context is; who the key players are; what their sources, forms and degrees of power consist of; what their interests, incentives and ideas are; where their support and opposition comes from; and how they typically work politically through or behind the prevailing formal and informal institutional arrangements.

9. Investing so as to help facilitate the emergence and support of local developmental leaderships and coalitions, and investing in processes that enable them to work politically in locally effective ways, is at least as important as investing in projects with pre-planned end points.

See http://www.dlprog.org/ftp/ for the full list of DLP publications. For more detail on this see Leftwich (2011).

1.1   Monitoring and evaluation

In addition, monitoring and evaluation practices have been identified as an area requiring attention. This paper speaks to that issue. It looks at how donors can better assess the effectiveness of working and thinking politically. It looks in particular at the appropriate methods and approaches for monitoring and evaluation of donor programs that support developmental leaderships and coalitions and the political processes through which they work.

02

Evaluating working and thinking politically

Evaluating the emergence of developmental leaders and coalitions requires long-term and flexible strategies based on a deep understanding of political context and supportive of locally led practices. This requires monitoring and evaluation processes that provide 'real-time' feedback, that allow for changing strategies and objectives over time, and that can help to build the wider evidence base for transformational change, by answering questions on 'how' and 'why' things change, or don't. This presents several challenges for more orthodox approaches to monitoring and evaluation.

2.1   The results and value for money requirement

There is currently pressure on development actors to demonstrate results and value for money within very short time frames and certainly within the life of programs. The reality for programs working in political processes is that change is likely to be long term, non-linear, often slow to mature and difficult to predict. The most immediate challenge for programs focused on working politically is therefore to demonstrate *during the program life* that they are leading to tangible outcomes and benefits for people. A further challenge is to demonstrate, in the short to medium term, the value of contributing to the growth of stable institutions and systems, when the value of those institutions or systems, for developmental change, may not be evident until well beyond the program life.

While the DLP research suggests that engaging in political processes can produce tangible outcomes for people in the long term, intermediate outcomes are more likely to be changes in political organisation, capacity and space to create momentum for action around development issues. For example,, in South Africa women's coalitions have made significant strides for women's rights and access to improved services, including reproductive healthcare, increased protection of women in the workplace, better education for women and girls, and the reform of rape laws. This had been achieved through intermediate steps, in which these coalitions had drawn on and expanded their networks, and then exploited political and institutional arrangements to build developmental partnerships that spanned both civil society and government, thereby creating the space and capacity required to achieve these tangible outcomes in the long term (Hodes *et al.,* 2011).

Understanding these intermediate outcomes is an important outcome of monitoring, which helps to ensure that programs remain relevant to changing contexts, and adapt and respond to new political opportunities as necessary. This sort of monitoring provides programs with the opportunity to learn and improve, and provides direction for programs that seek to be flexible and responsive to dynamic contexts. Monitoring and evaluation processes therefore need to balance the donor request for tangible 'results' against the need to assess less tangible, but critical, processes of change, which produce these results.

2.2   Causality and attribution challenges

A further challenge, and one which is of concern to donors, is the issue of attribution. While DLP research makes it clear that positive changes are driven by political processes, managed and influenced by coalitions and networks of development actors, the research is less clear about the role and contribution of external agencies and donors in these situations. In the context of the current focus on results, donors are very interested in being able to draw a direct link between their inputs and the results identified.

The reality of a program that takes a political approach to understanding and managing change is that the donor role is likely to be one of facilitation and support for local actors to bring about change. The donor program may be one, albeit a significant one, of the many contributory factors in that situation. A simple attribution between outcomes and donor inputs is not possible in this situation. Rather, the monitoring and evaluation has to provide evidence of donor *contribution* to change and answer questions about the effectiveness and efficiency of the contribution as analysed against the results that have been achieved. Donors need feedback to know their money was well spent. If they cannot 'own' the result, they need at least to understand the significance and quality of their contribution.

2.3   Top-down assessment and evaluation

Finally, an emerging challenge is the current push towards top-down processes of assessment and evaluation. In some cases these take the form of centralised results frameworks, which ignore the context of change and try to aggregate results over multiple locations in order to claim wide spread impact. For example, DfID's new business model has core results that set targets of, for example, millions of people assisted to access education, or millions of bed nets distributed to prevent malaria (DfID, 2011). While the figures provide comfort to the public that their money has been directed to helping people, the focus on simple results narrows the monitoring and evaluation agenda, pushing practitioners to assess short term tangible outcomes instead of the processes and other changes that would help assess the less tangible, but often critical, changes in the political and social context. It is those changes that are often required to ensure that the education services are directed towards poor people, and that the bed nets actually reach the children living in remote rural areas. Further, that the services will be sustained and that people have the freedom and ability to use the services freely. Monitoring and evaluation has to avoid being simplified and narrowed by what is essentially a political agenda.

Donors are also centralising their methodological approaches, wanting uniform approaches to assessment. USAID's recent evaluation policy, with its expressed preference for impact evaluations based on models of cause and effect and a credible and rigorously defined counterfactual, is an example of this trend (USAID, 2011). As discussed later in this paper, experience suggests that performance assessment of programs engaged in working politically, and through networks of development actors, require different sorts of assessments at different points of time, and across various different domains of programs. It will be argued that a mixture of assessment and evaluation approaches and methods, used in an informed and context-specific way, are likely to provide the most useful range of information to satisfy accountability requirements, program improvement and learning requirements. The challenge is to ensure that donors and those they work with, or through, have the range of skills, knowledge and expertise to be able to draw upon an appropriate mix and range of methods that such an approach to assessment requires. A further challenge is to ensure that donors and others are able to understand the limitations of single methodological approaches and the need for investment in a more sophisticated mix of assessment approaches.

03

Different approaches to the evaluation of 'politics'

Key to understanding the different approaches to monitoring and evaluation of politically informed programs is an understanding of the context in which developmental changes take place, alongside an understanding of the purpose and use of monitoring and evaluation. These perspectives provide a base for exploring appropriate methodologies and how these might be mixed together to address the challenges outlined above.

3.1.   Performance assessment and the context of change

David Snowden (Snowden & Boone, 2007) argues that change can be understood in terms of four possible dimensions.

- *Simple:* In this dimension change is linear and can be attributed to a particular intervention or input. 'Simple' change examples include providing vaccinations so that children are prevented from dying of vaccine-preventable diseases, or providing teacher training so that more skilled teachers are available within the national system. They generally involve short causal chains and existing evidence about how that causality produces the desired outcome. Notwithstanding the assumptions and possible risks within any type of program or project, change in a context such as this can largely be predicted and can be measured through agreed indicators and expected results. Nearly all development interventions have some aspects that have a comprehensible and predictable linear link between inputs and outcomes, and for which best practice from elsewhere is usually applicable.

- *Complicated:* In this dimension the link between the change achieved and the original inputs and activities is not a simple linear 'cause and effect'. Most often there are several intervening variables and influences involved. However, with good analysis and appropriate stakeholder consultations, the 'map' of the various connections can be developed and the system or pathway between inputs and change can be identified. Examples of this sort of change might include providing scholarships for emerging leaders to study and extend their skills in order that they might be better positioned to contribute to developmental change in their country. While past experience might provide evidence that further study opportunities are an important contribution to the development of new leaders, various factors within a particular context will also be important. Assessment processes that can track the progress between the take-up of a scholarship and the eventual contribution an individual might make to development and change require attention to several interactions in the causal chain, some of which might be predicted and assessed through predetermined indicators, while others might need to be understood through different and more exploratory methodologies.

- *Complex:* Programs that are located within the complex change dimension are those where the pathways towards change are largely unknown at the beginning of practice and may never be completely understood even at the end of the program. While experience and principles from other situations may guide the design and implementation of such work, it is often the case that it is only by probing and acting that understanding is developed. In these situations regular monitor-

ing and feedback provide the information to enable the program to assess its progress, or not, towards its objectives, and adapt as experience and learning develops. Typically this assessment is not against predetermined indicators (given that the pathway to those objectives is not clear). Rather it is exploratory, gathering information about what change has happened, then analysing this information against the broader outcomes being sought. Examples of programs like this include those working across diverse sectors, those where multiple partners are involved, or those operating in locations where the intersection of politics, culture and social and economic relations is itself complex and dynamic. For example, AusAID works to support improved law and justice in Papua New Guinea (PNG) through support to a sectoral program that brings together several PNG Government departments (the Department of Justice and the Attorney General, the National Judicial Staff Services, the Ombudsman Commission, the Royal Papua New Guinea Constabulary, Correctional Services, the Magisterial Services, the Office of the Public Prosecutor and Public Solicitor and the Department of National Planning and Rural Development). The program has several work areas, focusing on deterrence, coordination across the sector, prevention of crime, and restorative justice. It seeks to work both within the formal legal systems as well as making use of the traditional legal and community systems. In this complex web of many variables and players it is not possible to map all the interactions and influences upon change. The monitoring and evaluation processes need to build knowledge about what is working and why, assisting the decision makers to direct their resources and focus appropriately.

- **Chaotic:** in 'chaotic' situations of change it is never possible to understand the connections between inputs and outcomes. Practice is guided by instinct and principles, and measurement is usually about the quality of inputs rather than understanding their specific contribution to change and outcomes. Examples of development interventions in 'chaotic' situations might include humanitarian responses or development programs undertaken in protracted and unstable war situations.

Development programs that are seeking to engage in political processes and support local actors to work together for change may find elements of all four dimensions of change throughout the life span of the program. While much of the work being undertaken is likely to lie within the 'complex' dimension of change where the many connections, influences and interactions between actors and institutions and sustained change are only partly understood, there are likely to be some elements of the intervention or context where the connection between inputs and outcomes are 'simple' or 'complicated' and therefore can be identified and more closely managed and assessed. On the other hand, there may also be some times in the change process that are simply 'chaotic' and where the program is limited to measuring the quality of its inputs.

Box 1: AusAID support for decentralised service delivery in PNG

The Australian government, through AusAID, provides support for PNG's decentralised system of government, with the purpose of contributing to improved service delivery. The program has a range of different interventions and ways of supporting and facilitating effective change, recognising the complex political, social and cultural dynamics in the rural PNG context. Some interventions have a close and easily identified relationship between inputs and outcomes. Examples include the provision of training for provincial government staff to provide them with the skills they need to undertake their work. These interventions are closely monitored based on expected changes in the behaviours of those staff.

Other interventions have more complicated pathways for change. These include peer support between provinces and mentoring for provincial administrators. Elements such as these are focused on facilitating local actors to develop their own solutions for improved government and therefore improved service delivery, but over time the relationships and processes that contribute to making these activities successful are increasingly understood by the program. The activities can be monitored for critical indicators such as engagement by senior level staff, changes in provincial priorities and budgets.

The program is also concerned with the interaction between national government and sub-national government departments, and the way in which these relationships operate through formal systems and less formal processes to support (or not) the effective flow of funds, policy and support in one direction and performance monitoring and accountability in the other. These formal and informal relationships are not under the control of AusAID and, as an outside organisation, AusAID will never be able to fully understand or direct these critical interactions. However, understanding the importance of these relationships, the agency takes up various opportunities to influence and support change-agents as they work to improve the national to sub-national interactions. Some of this work is long-term and ongoing and some is in response to unplanned opportunities. On a regular basis AusAID and its partner agencies consider the outcomes to date of this support and the lessons learned for the ongoing program work.

Land, Dobunaba, & Fargher (2010)

Understanding the nature of change in a development program provides a basis for appropriate development of a monitoring and evaluation framework. 'Simple' and 'complicated' elements of the change process can be described in advance with clear identification of assumptions and risks and more easily assessed through predetermined signs or indicators, or methodologies that focus on proving and verifying causal links. More complex change dynamics require skills in other types of data collection and assessment methodologies, most often those associated with qualitative and exploratory types of assessment. It is this mix of methods and assessment approaches that is most likely to provide the full range of information required to account for and assess change in politically focused work.

3.2.   The purpose of performance assessment

Monitoring and evaluation of development work has to serve several needs. Jones *et al.* (2010) provide a useful summary of the different purposes of monitoring and evaluation. These can be summarised in six overlapping categories.

- **Proving and rendering judgements** i.e. what works.
- **Improving** i.e. what can be done better, quality improvement.
- **Generating knowledge** i.e. theory building, identifying principles and new ideas, creating debate and dialogue.
- **Reducing uncertainty** i.e. gradually accumulating knowledge.
- **Legitimising or symbolic** i.e. to corroborate a decision already taken, to advance political ends, or to comply with procedure.
- **Empowering people** to further progressive social change. That is, to provide feedback from those that are meant to benefit, and increase their ability to change power relations.

Monitoring and evaluation needs to provide information for program management, learning and

improvement, as well as information necessary for accountability both upwards and downwards to those being served. However, it is very rare for a single assessment process to be able to meet all of these various purposes, except in the most simple of contexts. The nature of the particular intervention and the stakeholders involved will determine the kinds of evaluative questions that should be asked. For example, a recent AusAID review identified 18 possible evaluation questions for the Law and Justice programs supported by the agency (Cox, 2010). Each question addresses an important element of law and justice work but together would be far too large for a review. Identification of the needs of stakeholders and, therefore, the purpose of the assessment has to guide the focus and questions for assessment.

This is a critical step that needs to be undertaken before any discussion of methods is broached. As Patton has recently argued "valuing is fundamentally about reasoning and critical thinking. Evaluation as a field has become methodologically manic-obsessive. Too many of us, and those who commission us, think that it's all about methods. It's not. It's all about reasoning" (2012: 105).

3.3.   Mixed method approaches in monitoring and evaluation

Finally, we contend that developing effective systems of monitoring and evaluation for politically informed programs requires a good understanding of methodologies, in particular how to draw together an appropriate 'mix' of methodologies.

As donors have moved to more complex program approaches across various sectors there has been a rise in interest in the mixed methods approach.[4] Recent assessment of monitoring and evaluation in gender empowerment suggests that:

> "...no single M&E framework can capture all aspects of the change, impact, or results of a women's rights/empowerment project or strategy – in short, one size does not fit all! Because the nature of gender and social power relations is complex, embedded in multiple social, cultural, economic and political structures and institutions, single M&E instruments – such as the logical framework, theory of change, outcome mapping, or gender impact analysis – can assess **some** dimensions of the change process, but **not all**. Consequently, a comprehensive assessment process requires the application of multiple frameworks, methods and tools, working in a complementary fashion" (Batliwala, 2011).

Using a mix of methods in monitoring and evaluation can raise the standard of the assessment (Gray, 2009). It can provide for triangulation of assessments; one method compensates for weakness in others and provides a cross-check on the findings of the study of the same phenomenon (Bamberger and Segone, 2011). It allows for complementarity; different methods can explore different elements of a program and the results of one method can inform the development of the next. Using mixed methods helps to uncover paradoxes, contradictions and new perspectives. Finally, mixed methods can broaden and widen the range of a study to include more issues, e.g. contextual understanding, and participants lived experience, as well as changes in observable phenomena.

Most significantly, a mixed methods approach provides greater rigour. Different approaches to evaluation, and indeed research, have different ways of defining rigour, each of which is determined by its theoretical provenance. However, as Patricia Rogers[5] argues, an evaluative process is not only about choosing the right evaluation method, but about choosing the right combination of evaluation methods. Rigour, in this view therefore, depends "upon both **appropriate** method choice – selecting methods based on evaluative purposes and contexts – and successful implementation that meets accepted

---

4   For a recent thorough review of Impact Evaluation methods see Stern *et al.* (2012) http://www.dfid.gov.uk/Documents/publications1/design-method-impact-eval.pdf

5   See http://betterevaluation.org/evaluation-components

standards of quality" (IE4D, 2011:1).

Effective use of mixed methods requires a good understanding of the strengths and value of particular approaches.[6] The Impact for Development Group (IE4D) notes that a key principle for action must be that impact assessment exercises:

> "...draw on the full range of methods and designs for systematic and rigorous empirical investigation, including ethnographic, case study, statistical and experimental/quasi-experimental approaches and methods, as well as those from such fields as complexity science, participatory research, and action learning" (IE4D, 2011: 1).

While there are many methods that could contribute to development assessment, the following range provides an illustration of the range of methodologies that could be used in combination for effective assessment of working politically.[7]

3.3.1.   Network mapping[8]

Social network analysis assumes that social networks – organized through both formal and informal relationships and bonds – are important. In the real world who knows who, how players are linked via other players, and how knowledge moves through these links is important. Network mapping makes these relationships and the flow of information within networks visible. The network map can then be used to identify key players (e.g. those who may not be particularly high in a hierarchy but who have numerous links to others), to identify bottlenecks in the flow of information or those who are isolated from knowledge flows, to spot opportunities for better flow of information or influence and so on. It can also identify key points of influence to be targeted for change.

A recent multi-country study of service delivery, collective action and social accountability in Brazil, India, and Mexico used network analysis to:

> "…explore changing patterns of relations among collective actors who represent the poor and in their relations to policy makers. When combined with detailed qualitative and contextual data, it allows one to identify changes in the structural capacity for action of different collective actors and of their network as a whole" (Houtzager and Acharya, 2008: 35).

This study attempted to look at both formal and informal relations between individuals and how they link together organizations of which they may be a part (rather than the formal relations between organizations that much network analysis tends to focus on). Like the case study below (Box 2) this research sought to identify changes in the capacity of networks of collective actors by constructing 'before' and 'after' reform analyses. For example, the changing role of health professionals in the *Sanitarista* movement (the public health movement) in promoting the Brazilian health sector reforms of the 1980s and 1990s that led to universal and free access to healthcare (Dowbor, 2008).

Information for network analysis can be gathered in a variety of ways, for example, through surveys and questionnaires, or through participatory exercises. A number of software tools are available to help with the actual mapping process.[9] This approach is particularly well suited to programs working with coalitions and networks of change because it recognises that relationships, both formal and informal, are important and that they need to be understood as part of the change process. The process can

---

6   Approaches go beyond methods. Understanding the epistemology of the approach is an important starting point for assessing its use. See White & Philips (2012) for a more detailed overview of methods and their epistemological base.

7   For a useful tabular summary of different approaches to Impact Evaluation and the different ways they address causality see Stern *et al.* (2012: 24).

8   See Serrat (2009) for more details on how to do network mapping; Davies (2009) for a description of the approach and some examples of how it has been used; Clark (2006) for a comprehensive overview of network mapping and how to do it using one of the software packages.

9   For details on different software and much more on social network analysis see this collection of resources from Rick Davies http://mande.co.uk/special-issues/network-models/

show unexpected connections or blockages as well as identify the key points of influence where a program might need to focus.

Box 2: Using network analysis to understand the role of civil society organisations in Indonesia

Context

Starting in 2001, Indonesia has developed one of the most ambitious decentralisation policies in the world. The proportion of public spending determined at the local level has risen from around 15% of State budgets in 2000 to 30% more recently. The Asia Foundation Civil Society Initiative Against Poverty (CSIAP) was designed in 2004 to improve district-level policies and spending to address poverty. Given that technical assistance to help with budgeting and spending was already available to local officials, the program believed that additional political incentives were needed to encourage the reallocation of local funds towards the poor. CSIAP decided to work with non-traditional NGO partners, and in particular Muslim mass-based organisations in 38 districts and municipalities.

The study

The study looked in particular at how pro-poor legislation had been brought about in two contrasting Indonesian municipalities, Semarang and Pekalongan, and how a local NGO, Pattiro, had worked to influence developmental change. In Semarang, despite an unsupportive major, within a year of Pattiro introducing the idea of a poverty alleviation bill, the law had been passed. In Pekalongan, with a pro-reform mayor, Pattiro's Director was appointed to his 'Team of Five' that developed a popular strategy to reach the poor – dubbed the Acceleration Program – which used local neighbourhood associations to plan, allocate, spend and account for block grants for community development.

As part of the study, network maps were developed in both municipalities. Different actors were positioned relative to two scales: Reform-mindedness and Power. 'Before' and 'After' maps were developed to chart change in the actors' location on these axes, as well as how the relationship between these actors had shifted.
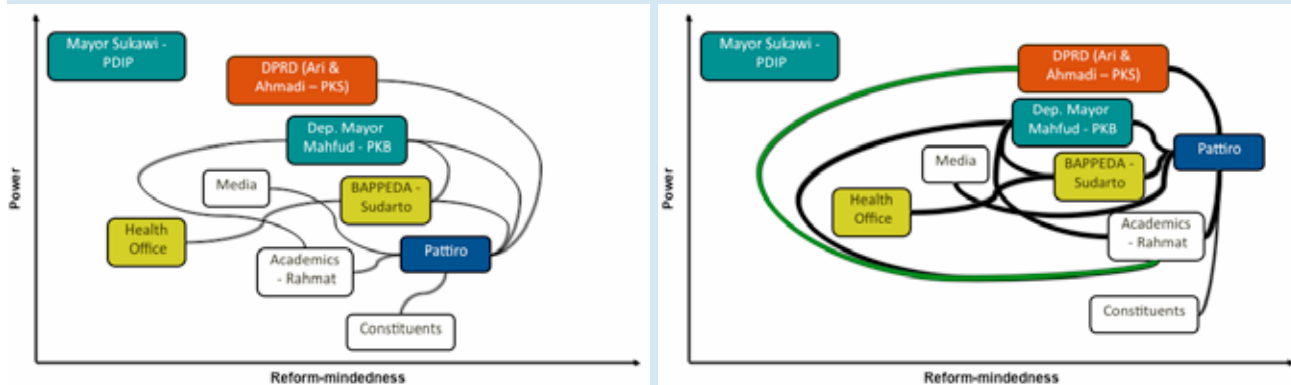
Stakeholder Map – Semarang (Before)   Stakeholder Map – Semarang (After)

Results and policy lessons

These cases indicate that a relatively small number of "champions" across civil society, the executive, the bureaucracy and Parliament were able to utilize their agency to garner broader political support for reform by offering a range of incentives to politicians. When powerful agents were more oriented toward political patronage, civil society could resort to punitive action (such as reporting violations to the Corruption Eradication Commission) alongside providing political incentives (such as mobilizing Parliamentarians against the mayor). When powerful agents were oriented toward reform, civil society effectively positioned themselves as political insiders, and significantly influenced how policy and spending were designed.

Maclaren, Putra & Rahman (2011)

There are, however, some limitations to the method. Understanding which relationships and connections to monitor over time requires informed engagement from various stakeholders. This, in turn, may raise some issues around confidentiality of information and monitoring processes. In addition, collecting the base-line data for this process can be time-consuming and careful thought needs to be given to establishing what information will be useful.

It is also important to recognise the distinction between networks and coalitions. Networks are generally a looser form of entity, for example, see Figure 1 below. From an evaluation perspective there are therefore likely to be a number of different characteristics of networks and coalitions that would need to be taken in to account, depending upon the evaluation questions and the nature of the relationships.

Figure 1: Distinguishing between networks and coalitions (Raynor, 2010)

Summarizing Distinguishing Evaluative Features

What to look at for:	Networks	Coalitions
Primary unit of analysis	Individual participants	Group as a whole
Potential units of analysis	Individual, organization, IOR as a whole	Same
Mission/vision	Buy-in/ clarity of personal benefit	Agreement on Goal destination and clear value proposition (why the right vehicle)
Leadership	Less centralized, facilitating "nodes" (boundary spanners)	More centralized (core team)
Progress lens	Outputs (# members, amount of exchange, etc.)	Outcomes (goal destination and value proposition)
Outcomes	Knowledge gains, belonging/ confidence, mobilizing ability (?)	Progress on agenda

tcc group

3.3.2. Outcome mapping[10]

Outcome mapping is an integrated planning, monitoring and evaluation tool that encourages agencies to think holistically and strategically about how they intend to achieve results. It focuses on changes in the behaviour, relationships or actions of 'boundary' partners (i.e. those partners who are in a direct relationship with a given program), as outcomes. The methodology seeks to assess the program's contributions to the achievement of outcomes. This is an approach that takes account of the broader development context but focuses the assessment on an agency's sphere of influence. The method uses

---

10 See the Outcome Mapping Learning Community website http://www.outcomemapping.ca/; the GSDRC website http://www.gsdrc.org/go/display&type=Document&id=2687 and link to the book on outcome mapping by Earl *et al.* (2001); analysis of the strengths and weaknesses of outcome mapping compared to logframes and the 'Most Significant Change' (MSC) method http://pme.global-connections.nl/advantages-and-weaknesses-different-methods

'progress markers', a set of graduated indicators of behavioural change, identified in the design stage to clarify direction with its primary partners and to monitor outcomes.[11]

This approach is strong on providing learning for program improvement and adaptation but weaker in generating evidence of impact to meet accountability needs. The strengths of outcome mapping include its attention to non-linearity and the complexity of the real world. The assessment process is a participatory process that enhances ownership and mutual accountability, as well as producing findings that are of use to both the agency managing the programme and their 'boundary partners' and use of findings. It provides a framework for continuous monitoring and iterative feedback, allowing learning and adaptation as the external environment changes. These features are particularly suited to programs trying to work responsively in dynamic political contexts.

On the other hand outcome mapping can be lengthy and in many situations it can be difficult to involve boundary partners in the assessment process. The approach does not measure relevance or cost effectiveness and provides limited information around attribution or measurement of impact.

Box 3: Outcome mapping & assessment: Smallholder dairy sector change in Kenya

'Rapid Outcome Assessment' is a visual tool that combines the outcome-mapping concept with a case study and episode study approach. It uses a systematic approach to collecting information about changes in the behaviour of key project partners, that contributed to a policy change that has already occurred.

The dairy sector in Kenya is mainly based on smallholder production with informal marketing by small-scale traders. The pre-2004 policy environment was based on an industrial cold-chain model with sales of raw milk effectively prohibited in urban areas. The Kenya Dairy Board (KDB) was the main regulator, and tended to harass and arrest informal traders. Powerful private sector actors put pressure on KDB to stamp out informal trade. By 2005 a more positive engagement by KDB with small-scale milk vendors was evident, including the establishment of a training and certification system, and business development services available to the informal sector. A new Dairy Policy had been approved, which explicitly recognised the role of small scale traders, committed to engaging with the informal sector for training and quality improvement, and saw the transition of KDB to a stakeholder-managed structure.

A rapid outcome assessment process undertaken as part of ODI's RAPID research enabled a greater understanding of how this process occurred and the role of a smallholder dairy project – implemented by the Ministry of Livestock and Fisheries Development, the Kenya Agricultural Research Institute and the International Livestock Research Institute – in achieving this. In particular, it provided insights into how the linkages between research, local campaigning and changes in the political environment combined to produce the changes in policy. The mapping process illustrates the connections between the various actors and their behaviours and the identified changes in the policy process. Based on interviews, case studies and episode analysis it traces the relative influence of the actors on the policy outcomes and therefore helps to unravel the way change took place.
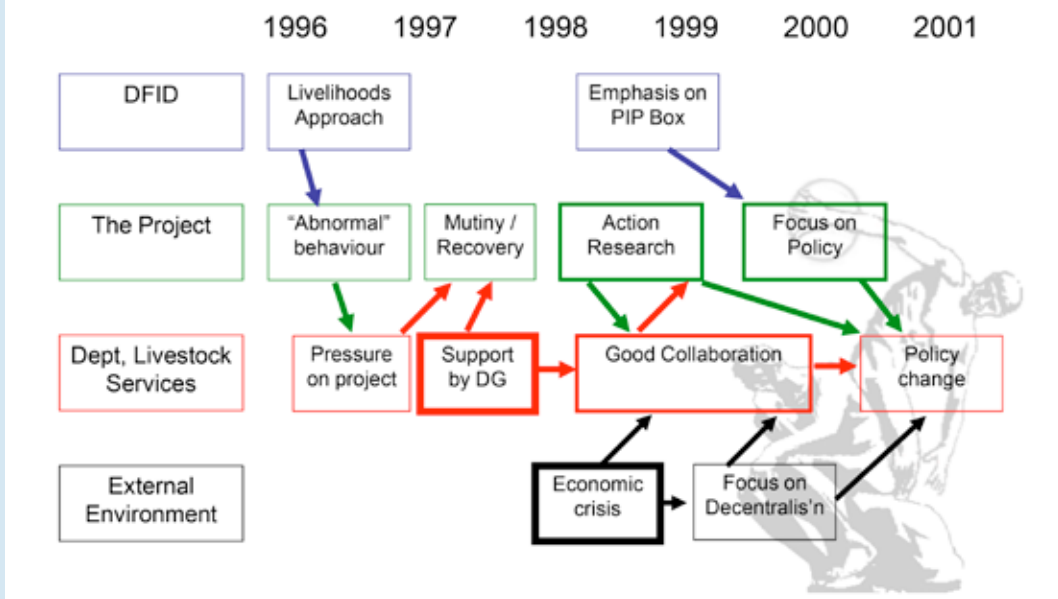
(Continued Overleaf)

---

11   See Earl *et al.* (2001) for full details about the tools.

(Box 3 Continued)
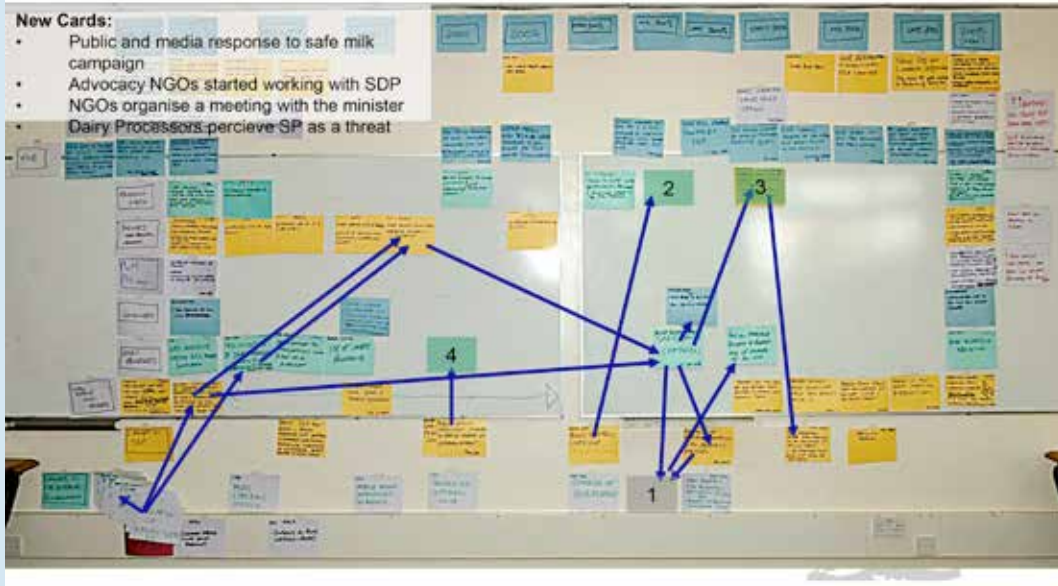
C. Leksmono et al. (2006) and See powerpoint of case study here, and methodology brief here.

3.3.3. 'Randomised Controlled Trials'[12]

The randomised controlled trial (RCT) method is an experimental approach, widely used in the medical world (for example, in drug trials). At its simplest, individuals or groups, are randomly allocated to either a treatment group (that receives the 'treatment'- the drug, input, resource, activity or whatever is being tested) or to the control group. Since, on average, the only difference between the treatment group and the control group is that one has received the treatment and the other has not, it is assumed that any difference in the outcomes is due to the treatment. In medicine RCTs are 'blind' – neither the subject nor the person recording the results knows who has received the 'real' treatment and who has not. In social policy evaluation it is not possible to conduct 'blind' RCTs, but attempts are made to reduce "Hawthorne effects"– people changing their behaviour as a result of being observed or part of a trial – and "John Henry effects"– people changing their level of effort as a result of their treatment or control assignment.

There are several limitations of using RCTs in assessment of development programs. RCTs are usually very expensive – average cost estimated to be about US$250,000 (Davies, 2011: 22) – and usually require large numbers of observations to obtain statistically significant results. For example, Olken's work on combating corruption in Indonesia involved 608 villages (Olken, 2007), while the study of community-based monitoring of primary healthcare providers in Uganda, described in Box 4 below (Björkman, M. & Svensson, J., 2009), involved 50 dispensaries and roughly 5,000 households in 9 districts. They are technically challenging, and require specialist input at all stages, from design to interpretation.

RCTs assume that changes in the political and social environment would, on average, have happened equally to the treatment and control groups in the absence of the experiment. Problems can arise when these changing conditions are related to treatment/control status, but are not the product of the treatment itself. So for example, if certain villages are in the comparison group in a microfinance study, but then a bank targets these villages for expansion of its services, that can invalidate the experiment. Furthermore, RCTs have to be implemented as part of program design, so as to ensure randomization in the delivery. Once a program has started it is too late to implement RCTs.

A further critique of RCTs relates to ethical concerns, i.e. in withholding 'treatment' from certain people or groups. Proponents of the method argue that (a) the rationing of 'treatment' is a normal part of programs, which always have limited resources, and random assignment can be a transparent and fair way of allocating scarce resources; (b) ethical concerns can be overcome by those in comparison groups being 'treated' in later program phases; and, (c) subjecting people to ineffective policies is itself ethically problematic, and RCTs help generate the knowledge policy makers need to avoid doing so.

RCT results may only be applicable to the very specific context in which the trials are carried out. This limits the external validity of the results of RCTs, i.e. whether results in a particular location, or for a particular group, are more broadly generalisable (Cartwright, 2007). Furthermore, RCTs can only test a relatively limited number of variables at the same time. So, whilst it may be possible for perhaps three or four variations in treatment to be applied while other things are held constant, more than this is liable to be impossible and/or prohibitively expensive. Proponents of RCTs argue that external validity concerns apply to any method, and that they can be addressed in part by replicating experiments in different contexts, and by designing RCTs that evaluate mechanisms and theories rather than just programs.

Well-designed RCTs can provide robust evidence that clearly establishes causality. Utilising this approach in a targeted way can produce evidence-based information that directly addresses the accountability

---

12   See Asian Development Bank (2011: 12-15), and http://www.povertyactionlab.org – the J-PAL website;

needs of donors and governments. Using RCTs to assess specific elements of programs seeking to change political processes may, therefore, provide compelling evidence around key causal relationships in specific contexts – the short term and tangible evidence that donors often require.

Patton (2011) has argued that RCTs may be appropriate when there is a discrete, concrete intervention (treatment); singular, well-specified implementation can be standardized; implementation can be controlled and there is no significant feedback between the intervention group and external environments; valid and reliable measures exist for the outcome to be tested; and, random assignment is possible and ethical.

Box 4: 'Power to the People': a field experiment on community-based monitoring in Uganda

Context

In the early 1960s Uganda had a functioning health care system. This collapsed in the 1970s and 1980s as Uganda went through political upheaval, and health indicators fell dramatically. After peace was restored in the late 1980s, the Ugandan government invested heavily in public health, which has been free of charge since 2001. However, health care delivery at the primary level is perceived as weak. There is a system of committees with community representatives to monitor the dispensaries that deliver primary health care, but few people actively participate. In addition, there is limited information available to the community about what standard of service they could expect and it is therefore difficult for the community to pressurise service providers to provide a good, or better, service.

The study

The study involved 50 public dispensaries and the health care users in their catchment areas covering nine districts and all four regions in Uganda, involving some 55,000 households. The dispensaries were stratified by district and then by population, and randomly allocated to either a 'control' or 'treatment' group. Quantitative information on service delivery was collected from the dispensaries. Since health workers might have incentives to misrepresent key data, the information was collected directly from the information that the dispensary kept for its own records (daily patient records, stock cards, etc.), rather than from administrative records. A stratified random survey of 5,000 households was used to collect data on health outcomes and the experience of users of the health system. Data on under-five mortality was collected and all infants in the surveyed households were weighed.

In the treatment group, a report card was prepared for each facility in the local language giving the results of the surveys, including information on the utilisation and quality of the service provided compared to other facilities. Local meetings were held to discuss the report cards and to develop an action plan detailing the issues and services that the community had identified as the most important to address, how this could be done and how they could monitor improvements. Community representatives met with the healthcare provider to develop a shared contract outlining what needed to be done, how, when and by whom. After the initial meetings the communities were in charge of establishing ways to monitor the provider.

Results and policy lessons

A year after the intervention, the study found a statistically significant difference in the weight of infants and a 33% reduction in under-5 mortality in the treatment communities. Utilisation for general outpatient services was 20% higher compared to the control facilities and the overall effect across a set of utilization measures was large and (statistically) significantly positive. A calculation by the authors suggests that the intervention cost circa US$3 per household or US$160,000 in total. The results of this study suggest that community monitoring can play an important role in improving service delivery and that enabling (poor) people to scrutinise whether those in authority have fulfilled their responsibilities can have a significant impact.

Björkman & Svensson (2009)

3.3.4. Quasi-experimental approaches[13]

Randomised controlled trials can be difficult to carry out in the real world of social development. Therefore, a number of experimental designs and methods have been developed that approximate RCTs but do not include the randomisation element. They allow comparisons between groups (e.g. program vs. no program) and can produce strong statistical evidence of cause and effect relationships. Garcia (2011) argues that despite the increasing interest in interventions aimed at supporting governance in development, donors have been reluctant to apply rigorous quantitative assessment to these programs. She argues that while such approaches are difficult, they are feasible particularly when the full range of quasi-experimental approaches is understood and the various strengths and possibilities of these approaches utilised throughout an assessment process in a way that complements other methodologies.

There are various types of quasi-experimental approaches. Methods include:

- **'Before and After Comparison'** – compares pre-program and post-program measures and assumes that any changes are due to the program.

- **'With and Without Comparison'** – compares the two groups – one of which has received the treatment and one of which has not – at the same point in time after treatment and uses any difference in the groups to estimate the effect of the treatment.

- **'Difference-in-Differences'** – looks at the mean change over time in two groups (treatment and non-treatment) and then assumes that the difference in these differences is due to the program treatment.

- **'Regression Discontinuity Designs'** – assume that for interventions where a cut off point can be identified, those individuals who fall just below the cut off point (and do not receive the treatment) should be statistically similar to those who are just above the cut off point (and do receive the treatment). Therefore any difference in the two groups is due to the treatment. The method assumes that people do not have precise control over the variable that is the basis of assignment, and therefore cannot easily manipulate their treatment status.

- **'Propensity Score Matching' designs** – a statistical method that attempts to compensate for the fact that the individuals or groups have not been randomly assigned to treatment or control groups and that there may therefore be differences between the two groups. Factors that are believed to be likely to affect the response to treatment are estimated and a 'propensity score' is calculated using statistical methods. Subjects in the treatment group can then be matched with controls that have similar propensity scores, in a way that approximates to a random allocation.

- **'Natural' experiments** – mimic randomised trials, in that groups or individuals are effectively randomly assigned to a treatment and control group through some means other than via the program (for instance by some sort of policy change outside of the program being investigated). One of the more famous examples of this was Robert Putnam's study 'Making Democracy Work' that used Italy's 1970 devolution of authority to 15 newly created regional governments with essentially identical institutions, as means to explore differences in political culture and social capital across the country (Putnam, 1993).

Quasi-experimental approaches, amongst other things, require statistical analysis to provide credible results. As with RCTs, this can provide strong statistical evidence for causal links and the 'internal validity' of a specific program. They therefore produce 'evidence-based' information that donors and governments tend to prefer.

The weaknesses of these approaches are similar to those of RCTs: they produce valid estimates of impact to the extent that the underlying assumptions are correct. They can be expensive and can be technically difficult, requiring specialist input. Changing conditions in the surrounding context can

---

13   See http://www.povertyactionlab.org – the J-PAL website, or Asian Development Bank (2011: 12).

introduce confounding factors and unintended impacts on outcomes can be missed. Similar to many qualitative and quantitative methods, the results are often only applicable to the specific context in which the trial was carried out.

Perhaps more significantly, it is important to understand that the observed effects or results in both RCTs and quasi-experimental approaches cannot be assumed to be permanent, nor can they be simply generalised. Booth (2011) notes in reference to the two studies from Uganda (discussed in Box 4 and Box 5) that these methodological approaches are often described as rigorous because of the care taken in the statistical design and testing techniques, but the same rigor is often not applied in extracting the policy relevant messages from the evidence. In other words they often 'draw mono-causal conclusions from what are clearly multi-causal situations' and assume that understanding one part of a causal chain provides evidence for explanation of the whole system.

Box 5: Fighting corruption to improve schooling: evidence from a newspaper campaign in Uganda

Context

A public expenditure tracking survey (PETS) in Uganda in the mid 1990s found that the average school received only about 24% of the allocated funding from central government. When this was discovered, the government began to publish monthly information in local newspapers about the transfer of grants to the local government so that communities, parents and head teachers could monitor the funds that they should be receiving. In addition, there was a government crackdown on local officials where funds were found to be missing.

The study

To assess the effects of improved access to public information on capture of funds a repeat PETS was carried out in 2002. The original 1996 sample consisted of 250 schools, randomly drawn from 18 districts. The 2002 survey revisited these original schools plus an additional 170 schools from nine of the original 18 districts surveyed. This repeat study showed that schools were now receiving more than 80% of their entitlements, although around 30% of schools were still receiving less than two thirds of their entitlement.

A quasi-experimental technique - a difference-in-differences method - was used (i.e. an assessment of the differences between two groups 'before' and 'after' the newspaper campaign was launched). Schools were divided into two groups: the treatment group, where the head teacher reported access to at least one newspaper in 2001, and a control group, where the head teacher reported no access to newspapers in 2001. The study compared the two groups to see how the percentage of their funding entitlement that was actually received had changed between 1996 and 2001. The study also looked at the relation-ship between how close a school was to a newspaper outlet and the funds it received, since there are some methodological problems with using head teacher access to newspapers as the measure. First, the publicity campaign was country-wide, so it potentially affected every school; second, head teachers without access to newspapers may have found out about the campaign from parents or others with access to newspapers; and, third, there might be a variation in how often the head teachers in the group reporting access to newspapers were able to access to newspapers.

Results and policy lessons

The study concluded that schools that were more exposed to the newspaper campaign experienced a statistically significantly larger reduction in local capture of funds. Schools where the head teacher reported access to newspapers received a (statistically) significantly higher proportion of the funds to which they were entitled.

Preliminary evidence suggested schools with higher access to newspapers, and with more funds, also had significantly increased enrolment and test scores. The study demonstrates that a relatively inexpensive policy change that improves public access to information can significantly reduce corruption.

Reinikka & Svensson (2005)

3.3.5. 'Most Significant Change' and 'Narrative Analysis'[14]

In narrative analysis, programs are viewed as practices constituted by stories. For the evaluator, stories are interesting not so much for the facts that they may contain (although the fact that two different accounts of the same event are factually different may be interesting in itself) but for the social construction that is placed on them.

There are various methods that are used, including interviews, story-telling circles, and theatre. One type of story-telling method that has become well known is the Most Significant Change (MSC) method. This involves the search for outcomes and impact through the collection of stories that stakeholders consider illustrate the most significant changes encountered. A key part of the MSC process is for stakeholders to explain why they chose a particular change story, and in so doing reveal the basis upon which they believe it to be more significant than others.

Narrative analysis provides an understanding of change from the perspective and interpretation of stakeholders, including project beneficiaries. It respects different ways of viewing the world and can allow for different interpretations that can be cross-checked to ensure a solid understanding of the impact and significance of change.

On the other hand, the process is not simple and it can be difficult to analyse and produce a coherent synthesis of the rationale behind particular stories and narratives. However, there are some attempts to combine the story telling aspects of MSC with more sophisticated statistical analysis and use of mapping and visualisation techniques (see the example from Kenya below, Box 6). Considerable cross cultural knowledge and sensitivity is often required to interpret data.

Furthermore, as with many techniques, it can be difficult to ensure that the voices of women, children and other less powerful people are sufficiently represented in the various narratives that are collected, if particular care is not taken. The process can be time-consuming, and while it can provide good insights into the real impact as expressed by stakeholders, the process can be difficult to communicate in a simple way to donors and governments. Stories are considered by some to be 'anecdotal' or 'subjective', and these methods are thus considered often less rigorous than more 'objective' approaches.

---

14   See Sukop (2007) for an outline of various techniques; http://www.mande.co.uk/docs/MSCGuide.pdf - explaining MSC.
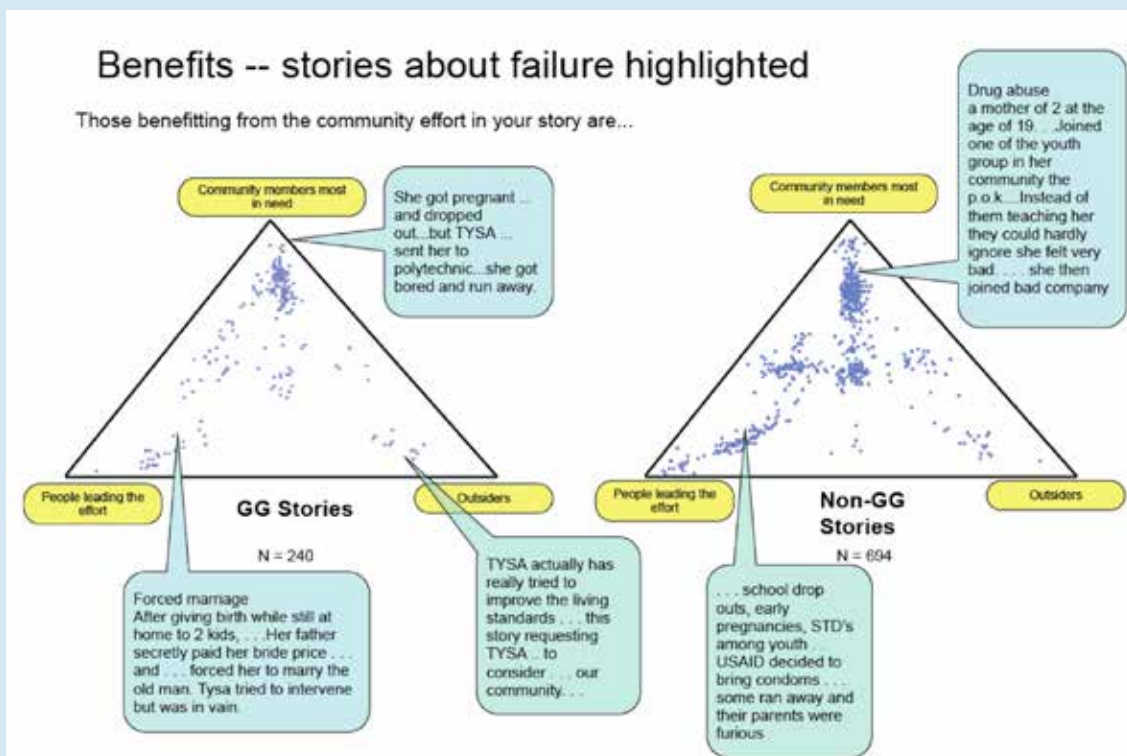
Box 6: Micro-narrative inquiry in Kenya

Context

Global Giving is a charity that uses a website to allow social entrepreneurs and not-for-profit organisations the opportunity to seek funding support directly from individual donors. Global Giving has been using a number of monitoring and evaluation techniques, including 'self-reports', visitor postcards, crowd-sourced feedback and formal evaluations. This produced considerable information and stories, but the quantity and quality varied widely between projects and there was no consistent structure for the stories. This made it difficult for the Global Giving staff in Washington, D.C., to make sense of all the information that they possessed. They wanted to improve the usability of the information, but also to keep the costs down, and they therefore decided to investigate crowd-sourcing more systematically. They decided to trial the use of a software program called Sensemaker developed by Cognitive Edge that provides a platform on which to gather, process, and visualise knowledge. Stories and anecdotes collected using this software are 'self-tagged' by the respondents and the results of this tagging are then analysed statistically.
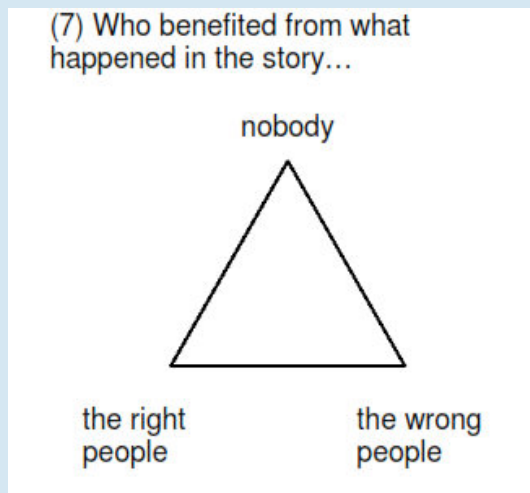
The study

In 2010 and 2011, teams of over 1,000 local Kenyan and Ugandan scribes collected 36,988 stories from over 5,000 community members by asking respondents to "tell us about a time when a person or an organization tried to change something in your community". They did not specifically ask about Global Giving projects, and a series of standard questions followed to gather more information about the story. The storyteller was then asked to self-index or tag the story, against different dimensions (e.g., whether the effort succeeded or failed, what the story related to, who the storyteller is, etc.) or through the use of triads (i.e., by placing a bubble or a cross in a triangle representing three dimensions). The stories were collected on paper and entered onto the Sensemaker software online, and are also available via on-line maps.

Example of analysis of stories, Guijt (2010)

(Continued overleaf)

(Box 6 Continued)

(7) Who benefited from what
happened in the story…

nobody

the right
people

the wrong
people

Example of triad, Guijt (2010)

see http://www.globalgiving.org/story-tools/. For information on Sensemaker see http://www.sensemaker-suite.com/

Results and policy lessons

This process allowed an aggregated visual picture of the stories showing clusters, distributions and patterns. It also allowed users to drill down to the individual stories. It revealed significant differences between the results from community stories compared to how outsiders tagged the same project. The stories also highlighted issues that Global Giving had not previously recognised as important in the community and informed them about other organisations working with the same communities.

3.3.6. Participatory evaluation[15]

Participatory evaluation is an inclusive method that involves the participation of a wide range of stakeholders – including in particular the intended beneficiaries – in monitoring and evaluating impact and outcomes. The stakeholders are involved in deciding what is to be measured, what information to collect (and in collecting it), analysing the data, and recommending further action. The actual level of involvement of the stakeholders in all parts of the evaluation can vary. Sometimes stakeholders coordinate, control and carry out the whole process; in other cases they are somewhat less involved and the process is facilitated by an external evaluator. However, whatever the exact form, participatory evaluation is more than simply asking beneficiaries their opinions - they are actively involved in the control of the process.

Various types of participatory evaluation methodologies have been developed – examples include participatory monitoring and evaluation (PM&E), Participatory Rural Appraisal (PRA), and Participatory Learning and Action (PLA). Whilst they differ in detail, all have the fundamental participatory aspect at their heart. Within these various forms, there is a range of techniques that can be used for gathering information, both qualitative and quantitative.

The approach is particularly valuable in working with networks of development actors where those actors need to understand and monitor the change process they are engaged in. Well applied, the process can build capacity and can help stakeholders deal with critical and less positive outcomes and findings.

The process requires an organisational culture that enables participants to share power over data collection and analysis with other stakeholders. It is sometimes very difficult to involve the very groups – such as poor or marginalised people, or women – who would provide the greatest insight into evalu-

---

15   See http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20190347~menuPK:4121 48~pagePK:148956~piPK:216618~theSitePK:384329,00.html for a World Bank site with description of participatory evaluation and links to some associated techniques; http://www.gsdrc.org/go/topic-guides/measuring-results/participatory-mande-tools-and-approaches for a GSDRC site with links to various documents on participatory evaluation, and Chambers (2007) describing the use of quantitative methods in participatory approaches.

ation of change. It requires support to develop skills in data collection and analysis among stakeholders. A frequent criticism of the approach is that data collection undertaken by local beneficiaries can be less rigorous or technically flawed, and indeed can reinforce existing power relations (Cooke and Kothari, 2001).

A particular tool often utilised by participatory approaches is appreciative enquiry.[16] Appreciative Inquiry (AI) focuses on the strengths of an organisation, system or group and systematically and consciously searches for what is working well, and the reasons for that success. The approach is particularly useful in 'complicated' change situations where analysis of successful outcomes can help unravel the causal factors and provide an understanding of particular contributions by donors and others.

AI uses a variety of data collection techniques to elicit information, but its particular strength is that it builds connection and cohesion between groups of stake-holders as they are able to collectively explore the reasons for successful outcomes. In turn, this can identify processes and influences that may be overlooked or misunderstood by external stakeholders such as donors.

The process does need to be well managed to ensure that all voices, not just the powerful, are heard. More significantly, poorly managed processes can be uncritical and can lead to incomplete assessments of program performance.

The approach lends itself to program learning and improvement, but also provides a good basis for explaining attribution in 'complicated' change situations.

3.3.7. Assessing cost effectiveness and value for money

Increases in aid budgets and growing scrutiny of aid effectiveness have led to an increased emphasis on the assessment of Value for Money (VfM) of aid programs and modalities. It is recognised that for interventions in the 'complicated' and 'complex' domains, as referred to above (section 3.1), *how* value is assessed and *who* assesses it are not straight-forward issues.

> "The development of a Governance programme requires sufficient range and depth in the context analysis (social, economic and political) to formulate relevant responses to very different circumstances, and as a consequence Governance programmes display necessarily diverse and complex pathways to successful outcomes. Building strong business cases for investment which respond constructively and realistically to the focus on managing and measuring for results, and obtaining 'more for the same' or 'more for less' is a particular challenging". (Barnet *et al.*, 2010:7).

This study of governance and value for money for DfID (*Ibid.*: 11) presents four options for improving and assessing value for money.

1. **'Improved Results Based Management'** in which improvements in results based management and other project management processes are used to demonstrate improved VfM.
2. **'Rating and Weighting' approach,** in which key processes associated with Economy, Efficiency and Effectiveness are identified and used to rate programs, and where these processes and measures can be weighted to reflect their relative importance.
3. **Trend Analysis,** in which a comparison of progress against indicators is used to measure effectiveness and in which baselines, targets and underlying trends are all taken into consideration.
4. **Cost-Benefit Analysis (or Social Return on Investment),** in which outcomes are monetized or financial proxies are introduced in order to make a direct comparison between costs and the

---

16 See http://www.appreciativeinquiry.net.au/aidev/ website of the Appreciative Inquiry Network; Serrat (2008b); Hammond (1996) – a 'where to start' text.

financial value of benefits.

It suggests that these options are cumulative, progressing from an ability to demonstrate that programs are **managed** with VfM in mind, to an ability to **measure** (and compare) the value programs generate with the money spent (Barnet *et al.*, 2010). We describe below (Boxes 7 and 8) two examples of attempts to explore the value for money of programs that have a similar focus to DLP, i.e. concerned with leadership, politics and human agency.

Box 7: DfID governance and conflict programs

In their work with DfID, Barnet *et al.* (2010) developed a process for assessing programmes to support the first two options, described above (improved results-based management, and rating and weighting approach). These were identified as the most useful starting point by DfID staff, and could be adapted to use *ex ante* or *ex post*. A table of criteria related to the three VfM elements of economy, efficiency and effectiveness and a rating scale based on a number of score descriptors were seen as an initial step, and presented an example of how programs might start to undertake VfM judgements.

From the perspective of DLP it is of interest to note that one of the key criteria of effectiveness was **leverage and replication**; that is, the extent to which the programmes help to build individual, organisational, administrative and political capacity for the long term. This was in recognition of the fact that governance and institutional support programs, which tend to focus on capacity development, and enhanced accountability and responsiveness, can often:

> "...produce spin-offs, multiplier or replication effects, and leverage further benefits through advocacy and the success of a good example...[t]hese additional benefits should be factored into valuing outcomes on the basis that the strength of the evidence provided indicates that it is reasonable to conclude that they will occur" (*Ibid.*: 14).

Two further effectiveness criteria were identified: **theory of change,** and strength of **indicators**. These were seen to provide a test of the assumptions made, and of how far outputs could be expected to deliver the desired results.

Table: Barnett *et al.* (2010) VfM scoring sheet (without descriptors)

		Score Descriptors				
	Vfm Criteria	1	2	3	4	5
Effectiveness	Leverage/ Replication (Weighting X 3)					
	Theory of Change (Weighting X 3)					
	Relevance and Robustness of Indicators (Weighting X 3)					
Efficiency	Productivity measure (Weighting X 3)					
	Risk Analysis and Mitigation (Weighting X 3)					
Economy	Procurement (Weighting X 2)					
	Unit Costs (Weighting X 3)					

This prototype was tested with DfID staff who found the table useful. It was felt that it provided "a structured, systematic way to analyse programmes for VfM" (*Ibid.*: 19). Many suggestions were made for how it might be improved or adapted including how criteria might be developed in different ways. In particular, it was recognised that more work needed to be done on the theory of change criterion and descriptors (which were seen as too linear), perhaps through the provision of better guidance on writing and expressing theories of change related to governance and conflict programs.

Barnet *et al.* (2010)

AusAID used a survey-based approach to assess the cost-effectiveness of its *Pacific Leadership Program* (PLP). The program is designed to foster developmental leadership in the Pacific, and works closely with DLP. The program has developed a hybrid management arrangement whereby AusAID and program contractor staff are co-located and work closely together in an attempt to ensure the responsiveness of the program. As part of the evaluation this partnership model was compared with a purely contractor-managed partnership program and a grants program. This exercise also used a rating approach, whereby different stakeholders were asked to assess their satisfaction with PLP and determine the importance of the service/benefit to the achievement of common outcomes for each of the programs. These are summarised in the Table below.

Table: Elements of Pacific Leadership Program outcomes survey

Target	Key Focus	Sample
Partners	Satisfaction and importance to partner rated for: • Funding • Flexibility and responsiveness • Financial support • Improved accountability • Access to networks • Assistance with advocacy • Reduced reporting burden • Improved trust and transparency	All partners (12) including sampling all key people in each partner that are involved in the partnership
Advisory Panel	Higher level survey with no more than 15 questions: • Difference between engagement with AusAID management and with contractor • Main source of value in partnership approach	All Advisory Panel Members
AusAID	Benefits accruing to AusAID due to: • Knowledge sharing/retention • Alignment • Networking	All nominated AusAID staff that have direct contact with the PLP

At the same time, the costs of the program were compared with other similar programs delivered in a different way, e.g. through contractors, or grant programs. The approach proved useful in that it allowed both quantitative and qualitative analysis of the value as assessed by different stakeholders in a relatively simple manner (and at a relatively low cost), allowing managers to arrive at a reasonable, but evidence-informed, judgement of the relative merits of working in this way.

AusAID (2012)

Both these examples suggest that for interventions that involve complex partnerships engaged in un-predictable and non-linear process of change, it is possible to include a value for money assessment, but tailored approaches are required. Whilst quantifying inputs might be relatively easy, the assessment of 'effectiveness' or 'value' is clearly more complex. These case studies suggest the need for techniques that allow for scoring, rating and weighting of different criteria - in essence because they allow for a negotiation of 'value' and the development of evaluation 'rubrics' (i.e. descriptions of levels of 'perfor-

---

17   Methodology paper value for money analysis of the Australia Africa community engagement scheme, 18 April 2012

mance' from 'good' to 'bad'), which can differ by context and can evolve over time. Rubrics can often be more appropriate than narrowly defined objectives that address only one part of performance, or than attempts to inappropriately 'dollarise' all outcomes in order to be able to compare different interventions.[18]

3.3.8. 'Action-Reflection' and 'Action Research' approaches[19]

'Action Reflection' or 'Action Research' approaches assume that data collection and analysis are an ongoing part of the program management process. Any intervention or action undertaken by the program contributes to reflection on and further refinement of the program (Reason and Bradbury, 2008). 'Participatory Action Research' builds on the work of Paulo Freire, and Orlando Fals-Borda who sought to value ordinary people's lived experience as a legitimate way of knowing - as opposed to solely valuing intellectual or academic knowledge (Fals-Borda, 1987).

While there are no specific tools and techniques for data collection, the approach requires management commitment to analysis and consideration of program outcomes on a regular basis. The strength of the approach is that it helps programs to respond to unexpected and new information. It particularly encourages learning from experience and supports programs to focus on learning and improvement. An action-reflection approach complements a mixed methods approach as it encourages attention to multiple sources of data to challenge stakeholders to further reflection and enquiry about the outcomes of their action(s).

This approach is challenging for many donor agencies as it can lead to the need to change their objectives in response to new information and learning. The major risk for this approach is that, given the demands of program implementation, time is never allocated for adequate data collection and analysis and therefore reflection is superficial and makes limited difference to the program direction.

---

18   See this website for examples of Social Return on Investment methods http://www.thesroinetwork.org/sroi-analysis
19   See Serrat (2008a), for an introduction to action learning, and http://issuu.com/catholicreliefservices/docs/reflective_peacebldg on the use of action learning with a focus on peace-building.

Box 9: The Pacific Leadership Program

The AusAID funded Pacific Leadership Program (PLP) works to supports influential Pacific leaders to shape and lead the developmental change they want to see in their region. The program has a broad design, currently focused around four objectives:

- to support leaders to lead effective influential organisations
- to support coalitions of leaders to identify national goals and drive developmental change
- to equip influential leaders to be effective, with tailored resources relevant to their context
- to build evidence, explore new leadership development initiatives and promote knowledge to strengthen leadership practice and assist stakeholders engaged in leadership development.

The objectives are operationalised through various program strategies. The program has changed considerably since its original design, as it has learned and developed appropriate strategies to support the emergence of developmental leadership in different locations in the Pacific.

Central to the development of the program and its ability to change and respond within its dynamic context has been an action-reflection process. Every six months the entire implementation team come together with management to reflect on the evidence of outcomes and learning from the past six months. This meeting includes feedback from stakeholders, detailed program monitoring and evaluation reports, as well as insights and observations from staff themselves. Increasingly, information is drawn from PLP-supported research and from relevant research insights from elsewhere, such as the **Developmental Leadership Program (DLP)**.

Up to five days are allowed for this meeting to ensure that evidence and information are carefully examined and the implications for programs and strategies fully considered. At the end of the meeting a plan of action for the next six months is developed. This, in turn, becomes the basis for assessment of progress at the end of the time. This process has been complemented by conventions and symposia of eminent and emerging Pacific leaders from a diverse range of sectors including private, government and civil society. The purpose of the Convention is to facilitate high-level dialogue between influential leaders on the impact that leadership has on achieving developmental change.

Using this approach the program has developed a reputation for being responsive and relevant to different contexts and events in the Pacific, and is increasingly valued by its diverse range of stakeholders and partners as a program that is open to feedback and adaptation.

For more on this program see http://www.plp.org.fj/

3.4. Application of mixed methods in politically informed programs

Mixed methods approaches to monitoring and evaluation are particularly appropriate for politically informed programs. The mix of methods allows for the collection of different types of data over time, addressing different needs for both short and longer-term accountability as well as the need for information for program learning and improvement.

The application of mixed methods for assessment of politically informed programs assumes the ability to draw from the best of different methodologies and match the information these methodologies provide to assessment and learning needs. The following example shows how different methods might come together to provide different types of information for different audiences.

Figure 2: Example of mixed methods approach
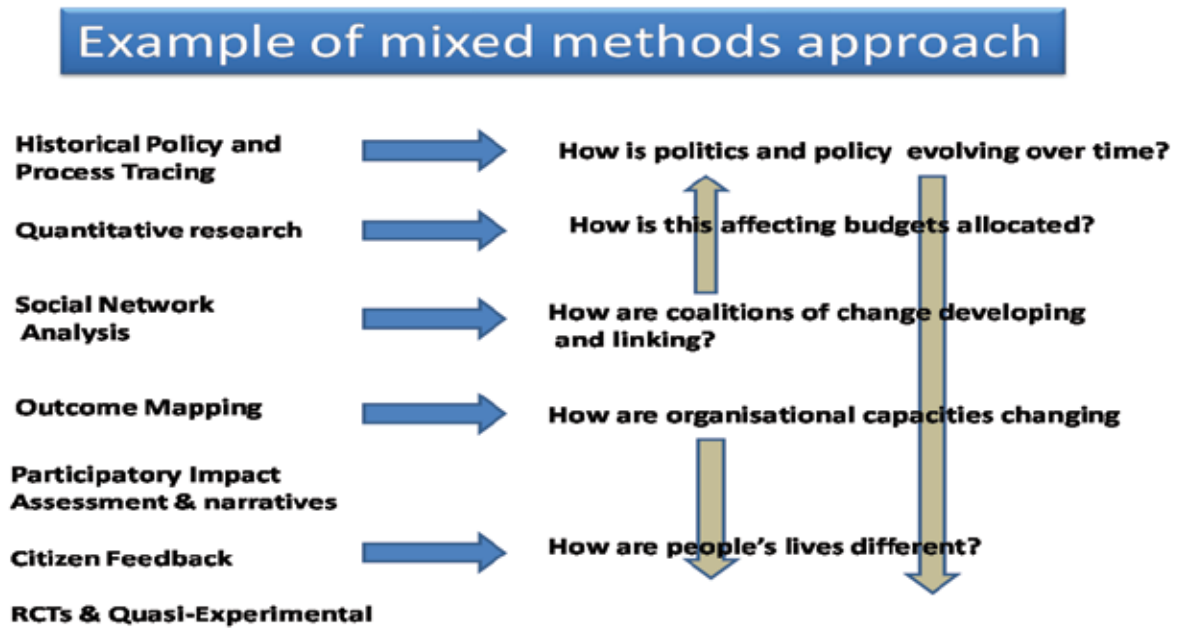
Mixed methods approaches need to be applied with care and skill, however (Bryman, 2007). If not managed well, they can produce products that are less than the sum of their parts – leading to a lot of information that is impossible to make sense of. As Stern *et al.* (2012: 81) note "[t]here are few ground-rules for how methods should be combined."

Mixed methods approaches generally require a 'fundamental logic of design', often based on a theory of change and clear evaluative questions, to hold the evaluation and the different methods used together. This design needs to be matched to the context and intervention, and to explore how different methods compensate for strengths or weaknesses in others, and how these combinations would strengthen causal claims (Stern *et al.*, 2012; White *et al.*, 2012). The example of the Wajir impact assessment below (Box 10) illustrates some of these points.

Box 10: Mixed methods example: Impact assessment of the Wajir Pastoral Development Project

The Wajir Pastoralist Development Project (WPDP) was a long term initiative undertaken by Oxfam Great Britain, which began in 1993 and was designed to reduce poverty and vulnerability of pastoralists and settled communities in the Wajir district of Kenya. Preliminary monitoring of progress during the first phase (1994-1996) highlighted the relative success of the project. However, against this background of reported, but not properly quantified, success a fuller evaluation of impact was undertaken.

The objectives of this study were to: (1) determine the extent and reasons for any increases in wealth and security of pastoral livelihoods; (2) assess the efficiency and effectiveness of service delivery to poor pastoral communities in the project area; (3) examine issues of institutional economics and capacity building as a consequence of the interventions under the programme; and, (4) assess the comparative value of the project in order to clearly delineate project impacts from the wider influences on economic livelihoods.

The study was undertaken by the Kenyan based Resource Management and Policy Analysis Institute. The method developed for the study used a theory-based mixed methods approach, which involved the following.

a. Developing, in a participatory manner with project staff, a theoretical framework outlining the supposed sequence of *"cause - effect"* relationships between the project components and pastoral livelihoods. This provided the basis for relating any observed changes in welfare to changes in behaviour that could be directly and discretely attributed to the project. Indicators were then developed to measure *impact* and the *means of achieving impact* at three levels: evidence for impact on pastoral welfare; the differences in the livelihoods of pastoralists in project and non-project sites, responsible for differences in welfare; the differences in the capacity of local institutions to deliver effective services.

b. Data on impact collected through a two-week survey of pastoral households stratified by: project and non-project sites, each approximately equivalent in agro-ecological resources and market access; family wealth; and, membership of pastoral associations. Two hundred pastoral families were interviewed, comprising some 24% of households in selected non-project sites, and 15% of households in selected project areas. Additional focus-group discussions with members of pastoral associations provided contextual information while a restocking programme was evaluated through a combination of interviews with 60 restocked families and analysis of project monitoring data that tracked flock size changes for 30 families every three months for two years. In addition, 184 female members of credit and saving schemes were interviewed to assess returns to the credit programme.

c. Analysis of the institutional capacity of pastoral associations, community based organisations and local government technical services, and an assessment of the cost-benefit of the investments in these agencies.

d. Analysis of project documentation and monitoring reports to provide further data and information. This material, based on systematic participatory monitoring processes, provided key data, which had been used for regular ongoing assessment of progress, but was also critical in informing the evaluation, which was focused on assessing longer term change.

This mix of methods allowed specific project benefits to be identified, which could be divided into those that yielded measurable economic gains, and those that were less easy to quantify but nonetheless made significant contributions to livelihood security, through, for example, improved water supplies and education facilities. Furthermore it allowed project participants to provide their perspective on relative changes in their conditions over the life of the project, compared to those in non-project sites. The assessment of institutional capacity development also made it possible to assess where the greatest contribution was made to benefits that accrued to the pastoralists.

Finally, because of its robust nature, the exercise proved useful in being able to communicate the clear linkages between institutional strengthening and pastoral welfare, and to raise awareness within the donor community. At the national level, the results formed a strong basis for policy reform and advocacy work on pastoral development issues, under the umbrella body of the Kenya Pastoralists Forum.

Odhiambo *et al.* (1998)

Mixed methods approaches can be expensive and difficult to manage, demanding high skill levels across multiple domains. The capacity of organisations and partner agencies may be a constraint when setting up any monitoring and evaluation system, let alone complex mixed method approaches. Consideration of what is going to be manageable and feasible, as well as what ongoing support might be required, is therefore critical. Without a cross-disciplinary team that has the ability to understand the potential of mixed methods, the design, implementation, and in particular analysis, of findings can lack integration. As Davidson (2012) notes "[t]he evidence is but one ingredient in the evaluation; the evaluative reasoning is how we determine what evidence to gather in the first place, and how to interpret it once we have." This concern is possibly one of the major reasons for the limited adoption of mixed methods in many donor approaches to development assessment. This may suggest that collaboration with researchers or universities may be required.

The example below (Box 11) of a multi-country study on social accountability, for example, was undertaken by "well-coupled interdisciplinary research teams" (Houtzager & Acharya, 2008: 27) and organized by the Centre for the Future State (CFS), based at the Institute of Development Studies at Sussex University, and research networks in eight countries through its partner institutions.

Finally, some mixes can demand too much of the participants who are faced not only by researchers engaging them in participatory exercises and focus group discussions, but also questionnaire surveys and randomised experiments!

Box 11: Multi-country study of modes of service delivery, collective action and social accountability in Brazil, India, and Mexico (BIM), Centre for the Future State.

This study explored how reforms of "the institutions that provide public services shape the capacity of organisations representing the poor to negotiate public policy and engage in social accountability of providers" (Houtzager & Acharya, 2008: 27-28). The method used was designed to address conceptual, measurement, and causal validity by outlining a comparable process of analysis across different sites, using mixed methods.

These methods were (a) case studies of process tracing of the negotiations of policy reforms; (b) 'before' and 'after' snapshots of the network of actors representing the poor, and their links to policy makers; and (c) quantitative snapshots, to assess whether social accountability was more likely when local organisations were connected to actors who negotiated policy, and whether social accountability leads not to just uptake of reforms, but improved services as well.

Figure 1 **Project multi-level analysis**

The diagram illustrates an increase in capacity of civil society actors, as the number of actors and their connectedness and ties to public officials grow. T₁ represents the period prior to reforms and T₂ the period after.

Analytic levels

1. **Narrative account** of *national-level* politics of reforms and *municipal-level* case studies of designing/implementing programmes that embody logic of the reforms

2. **Network analysis** of changing state−society relations
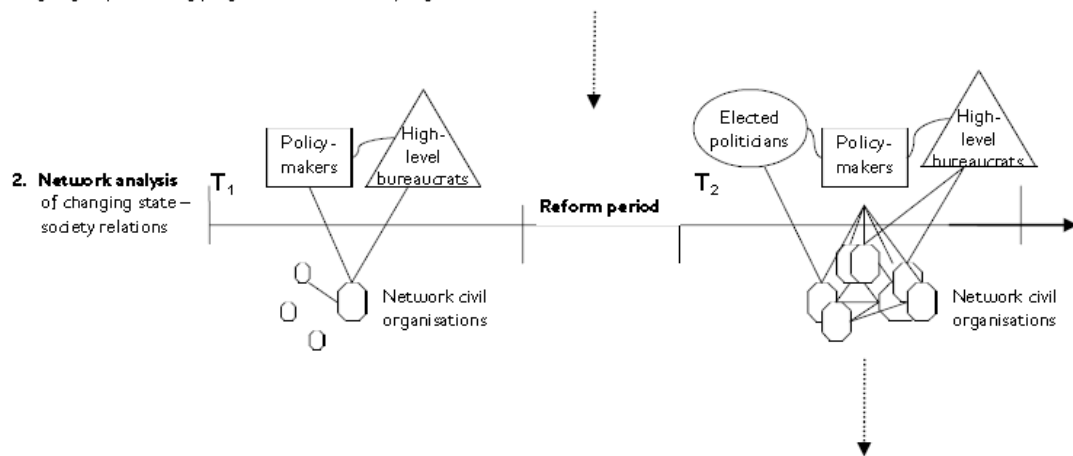
3. **Cross-sectional analysis** of 40 catchment areas in low income areas that vary in levels o f collective action to identify any impact on (i) policy implementation and (ii) service outputs

Houtzager & Acharya (2008)

04

And the politics of evaluation?

In this section we touch on some points that will be explored in greater detail in a second DLP paper, but which start to emerge from the discussion so far.

While the methodological and technical limits of different methods discussed above are relatively clear – even if some are more contested than others – many donors continue to privilege some methods over others. For example, USAID evaluation policy states:

> "Impact evaluations in which comparisons are made between beneficiaries that are randomly assigned to either a treatment or a control group provide the strongest evidence of a relationship between the intervention under study and the outcome measured" (USAID, 2011: 2).

Given the narrow information available through a single method approach, the question is: **why do development agencies default to these lower quality positions?**

We propose that there are several reasons for this limited methodological approach.

- The first is a failure by development practitioners to understand the origin of different methodologies and approaches that underpin monitoring and evaluation and their appropriate application within different contexts. As discussed earlier, different methods support inquiry in different contexts, ranging from those that are more predictable through to those that are more complex or, indeed, where changing the power and established relations is part of the development intention. However, donors and others often prefer to pretend all development interventions will follow simple cause and effect rules - Chelimsky (2012: 79) talks about "a general decline in our societal ability to entertain complexity". Unfortunately, a failure to understand and appreciate the epistemological basis for particular methods can lead to the inappropriate use of, or valuing of, one method over another. As a result, agencies become fixated on certain methods. Stern *et al.* note that "[u]p to now most investment in I[mpact] E[valuation] has gone into a narrow range of mainly experimental and statistical methods and designs that...DFID has found are only applicable to a small proportion of their current programme portfolio" (Stern *et al.*, 2012: i).

- Second, we suggest that there continue to be fundamental questions about the relationship and tension between **learning** and **accountability** agendas. While the debate on accountability is evolving[20], it is still the case that 'principal-agent' notions of accountability tend to dominate much of the thinking about evaluation.[21] That is, upwards accountability, for example to funder or manager. While this is a legitimate purpose of performance assessment, unexamined it can dominate the evaluation agenda with little space left for other purposes that monitoring and evaluation should

---

20 Fukuyama (2004) has noted that what he calls the 'delegation problem' can be handled in different ways and through different forms of accountability, notably through peer or relational accountability mechanisms (see also Brown & Jagadanan, 2007) or voice or exit options, which speak more to political accountability processes (see Hirschman, 1970). These different notions of accountability and feedback potentially provide means to resolve some of the learning and accountability tensions that are the subject of much debate in the literature and in development agencies.

21 For an introduction to debates on 'Cash on Delivery Aid' see, for example, http://www.cgdev.org/section/initiatives/_active/codaid/others_write_about_cod_aid_OK

serve. The performance assessment agenda becomes focused on 'proving' results rather than trying to 'improve' the program or learn why particular results are being achieved.

- Finally, we contend that evaluation itself is a political process, subject to political pressures internal and external to development agencies. For example, AusAID have recently released a new framework for the Australian Aid program. As part of improving the predictability of aid flows, the framework includes a four year budget strategy. In order to receive Australian Government approval for this four year budget, the new aid policy includes annual reporting on 'results'. Similar to the DfID experience noted earlier in this paper (section 2), the 'results' focus on tangible and specific outcomes that can be easily 'counted'. Examples include:

  - 4 million more boys and girls will be enrolled in school
  - 24,000 classrooms built or upgraded
  - 14,000 law and justice officials trained to improve community safety
  - 12 million textbooks provided, contributing to 20 million boys and girls obtaining a better quality education
  - 9,600 km of roads constructed, rehabilitated or maintained.

Simple aggregation of short term results such as these serves the purpose of providing information to the public in terms that they can understand. It also provides easy to measure 'triggers' for the Government's annual budget cycle. Unfortunately, what it can also do is focus agency attention on the achievement of these results at the expense of the assessment of the processes and relationships that determine sustainable impact in people's lives. Monitoring and evaluation processes in AusAID will now have to adjust to provide the short term tangible numbers required to maintain this exercise. Resources will potentially be diverted from more sophisticated mixtures of methodologies that provide information for program improvement, learning, and accountability to the people being served. There is a further risk that the work itself will become skewed. The new Australian aid policy commits itself to "act swiftly when expected results are not being achieved" (AusAID, 2012a: 25). With this pressure may come a focus on those programs that can produce the results sought in a short time frame rather than those working for sustained change informed by political, social and cultural insight (Barder, 2012).

Taylor and Balloch (2005) argue that all evaluation is inherently political, because all evaluations have to operate within political and resource constraints. Continuing this theme of assessment as a political process, O'Brien *et al.* suggest:

> "Evaluation involves politics at both the micro and institutional levels, and in multiple ways, since it engages complex social relations and decisions about rules and resources between stakeholders, or different interest groups, that have vested interests in the outcomes of the evaluation" (2010: 444).

They also argue that this is particularly the case for evaluation – as opposed to other forms of research – as the outcomes of evaluations will normally entail judgments that may have far-reaching consequences, not only for the staff of particular programs, but also for the lives of those that programs seek to benefit. For example, most development organisations are committed to empowering women and to supporting the interests of excluded groups, yet the way in which programs are assessed can lead to further marginalisation of these groups, undermining the very intentions of the agency. A focus on short term change, such as ensuring delivery of drugs, may be an easy way to communicate a 'result' but it might completely overlook the experience of people with disability who may never get to access the drugs because they cannot reach the health clinics. Reporting on delivery of books to schools might provide good stories about aid for children, but could ignore the experience of children from ethnic minorities who are excluded from schools. Judgement about what to assess and how, can privilege the experience of some people over others and, in turn, influence the decision makers in their

future aid spending.

These points will be explored in greater detail in the next paper on this area from DLP.

4.1   The implications for development agency staff

The ideas raised throughout this paper suggest a number of implications for evaluation and program managers in development agencies about how they manage both the methodological and political nature of monitoring and evaluation processes.

- First, monitoring and evaluation has to serve several purposes. Development agency staff need to know the various purposes of their assessment systems, and how to ensure that all are attended to over time.

- Flowing from this, development agency staff need the skills and understanding to be able to commission and manage monitoring and evaluation processes that draw from different perspectives. They need to have the skills to ensure their agencies are good purchasers of evaluation and assessment. They need to know the value of the information that they are receiving from particular methodologies, and the limitations inherent in any single methodological approach.

- A recognition of the political nature of evaluative processes should also be a fundamental skill for development agency staff. They need to be aware that by choosing single methods they are more likely to privilege the voice and interests of particular groups and focus on only part of the developmental outcomes.

- Managers and staff of development organisations need incentives to assess the quality of assessment processes and to provide the space and resources for good quality assessment systems. Staff and organisations need to be empowered to resist pressures to narrowly define 'results' and 'value for money' in order to ensure there is space for longer, wider, and multiple processes of change and assessment systems to capture the complexity of these processes.

05

Conclusions

At the core of DLP thinking is the proposition that political processes shape developmental outcomes at all levels and in all aspects of society; in short, that it is politics that is the binding constraint on development. This proposition has several implications for development agencies, including the need to utilise appropriate monitoring and evaluation processes.

A mixed methods approach to evaluation provides a way forward through some of the challenges, methodological or otherwise, presented in the assessment of politically informed programmes. There are several possible methodologies that may be drawn together in such a mixed methods approach. The key is to understand the strengths, limitations and appropriate use of each method in order to draw them together in a way that allows for complementary data collection and analysis.

While the case for using mixed methods is strong, some donor and development agencies fail to embrace this agenda effectively. In part this may be due to a lack of knowledge and understanding about the relative strengths and limitations of particular approaches. It is also the contention of this paper that decisions to use particular methods are in themselves political decisions, driven by political pressures, and that these can in turn exclude some interests and privilege others.

References

Asian Development Bank (2011) "A Review of Recent Developments in Impact Evaluation", Manila: Asian Development Bank. http://beta.adb.org/sites/default/files/pub/2011/developments-impact-evaluation.pdf

AusAID (2012) "Methodology paper value for money analysis of the Australia Africa community engagement scheme", 18 April 2012, unpublished paper, AusAID, Canberra.

AusAID (2012a) "Helping the World's Poor through Effective Aid: Australia's Comprehensive Aid Policy Framework to 2015–16", AusAID, Canberra. http://www.ausaid.gov.au/Publications/Pages/comprehensive-aid-policy-framework.aspx

Bamberger, M. & Segone, M. (2011) "How to design and manage equity-focused evaluations", Evaluation Working Paper No. 6, UNICEF, New York.

Barder, O. (2012) 'Seven worries about focusing on results, and how to manage them', blogpost 12 April 2012. http://www.owen.org/blog/5483

Barnett, C. Barr, J. Christie, A. Duff, B. and Hext, S. (2010) "Measuring the Impact and Value for Money of Governance & Conflict Programmes", ITAD Final Report to DFID, December 2010. http://www.dfid.gov.uk/R4D/Output/186103/Default.aspx

Batliwala, S. (2011) "Strengthening Monitoring and Evaluation for Women's Rights: Twelve Insights for Donors", Association for Women's Rights in Development, Toronto. http://www.awid.org/Library/Strengthening-Monitoring-and-Evaluation-for-Women-s-Rights-Twelve-Insights-for-Donors

Björkman, M. & Svensson, J. (2009) "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda," *The Quarterly Journal of Economics*, 124:2, 735-769.

Booth, D. (2011) "Working with the Grain and Swimming Against the Tide: barriers to uptake of research findings on governance and public services in low income Africa", Africa Power and Politics Programme, Overseas Development Institute, Working Paper 18, April. http://www.institutions-africa.org/filestream/20110606-appp-working-paper-18-working-with-the-grain-and-swimming-against-the-tide-by-david-booth

Brown, L. D. & Jagadanana (2007) "Civil Society Legitimacy and Accountability: Issues and Challenges," Civicus/Hauser Centre, January 2007. https://www.civicus.org/new/media/LTA_ScopingPaper.pdf

Bryman, A. (2007) "Barriers to integrating quantitative and qualitative research". *Journal of Mixed Methods Research*, 1:1, 8-22.

Cartwright, N. (2007) "Are RCTs the Gold Standard?" *BioSocieties*, 2, 11–20.

Chambers, R. (2007) "Who Counts? The Quiet Revolution of Participation and Numbers", IDS Working Paper 296. Brighton: Institute of Development Studies. http://www.ids.ac.uk/files/Wp296.pdf

Chelimsky, E. (2012) "Valuing, evaluation methods, and the politicization of the evaluation process." *New Directions for Evaluation*, 133, 77–83.

Clark, L. (2006) *Network Mapping as a Diagnostic Tool: A Manual.* La Paz: Centro Internacional de Agricultura Tropical. http://www.dfid.gov.uk/r4d/PDF/Outputs/FIT/networkmapping_LC06.pdf

Cooke, B. and Kothari, U. (2001) *Participation: The New Tyranny?* London: Zed Books

Cox, M. (2010) "Issues Paper, law and Justice Evaluation", Office of Development Effectiveness, AusAID, November. http://www.ode.ausaid.gov.au/current_work/law-and-justice.html

Davidson, J. (2012) "Why 'What's the best tool to measure the effectiveness of X?' is totally the wrong question", *Genuine Evaluation* blogpost 4 June, 2012, http://genuineevaluation.com/whats-the-best-tool-wrong-question/

Davies, R. (2009) "The Use of Network Analysis in the Evaluation of Social Change Communications," background paper for an Expanded M&E Framework for Social Change Communication, Communication for Social Change Consortium, London.

Davies, R. (2011) "3ie and the Funding of Impact Evaluations: A Discussion Paper for 3ie's Members", Canberra: AusAID. http://www.ode.ausaid.gov.au/publications/pdf/3ie-impact-eval.doc

Davies, R. and Dart, J. (2005) "The 'Most Significant Change' (MSC) Technique: A guide to its use", http://www.mande.co.uk/docs/MSCGuide.pdf

Developmental Leadership Program (2012) "Coalitions in the Politics of Development. Findings, Insights and Guidance from the DLP Coalitions Workshop, Sydney, February 2012". A DLP Workshop Report. http://www.dlprog.org/ftp/info/Public%20Folder/Coalitions%20in%20the%20Politics%20of%20Development.pdf

DfID (2011) "Results Framework", DFID, London. http://www.dfid.gov.uk/About-us/How-we-measure-progress/DFID-Results-Framework/

Dowbor, M. (2008) "Origins of Successful Health Sector Reform: Public Health Professionals and Institutional Opportunities in Brazil", *IDS Bulletin,* 38:6, 73-80.

Earl, S., Smutylo, T. & Carden, F. (2001) "Outcome mapping: Building learning and reflection into development programs", IDRC, Ottawa, Canada.

Eyben, R. (2005) "Donors' learning difficulties: Results, relationships and responsibilities", *IDS Bulletin,* 36:3, 98-107.

Fals-Borda, O. (1987) "The Application of Participatory Action-Research in Latin America", *International Sociology*, 2, 329-347.

Fukuyama, F. (2004) *State Building: Governance and World Order in the 21st Century*, Ithaca, NY: Cornell University Press.

Garcia, M. (2011) "Improving Donor Support for Governance: The Case for More Rigorous Impact Evaluation", German Development Institute, Briefing paper 11. http://www.die-gdi.de/CMS-Homepage/openwebcms3.nsf/(ynDK_contentByKey)/ANES-8KZHSF?Open

Gray D. (2009) *Doing Research in the Real World,* 2nd ed. London: Sage Publications.

Guijt, I. & Hecklinger, J. (2010) "Making Sense of Sensemaker: Evaluating development initiatives through micro-narrative capture and self tagging in Kenya", presentation to the American Evaluation Association Conference November, 2010. http://www.globalgiving.org/jcr-content/gg/landing-pages/story-tools/files/microsoft-powerpoint---makingsenseofsensemaker.pdf

Hammond, S. (1996) *The thin book of appreciative inquiry.* Texas: Plano Publishing Co. http://www.thinbook.com/pages/books/ebooks.htm

Hirschman, A. (1970) *Exit, Voice, Loyalty: Responses to the Decline in Firms, Organisations and States,* Boston, MA: Harvard University Press.

Hodes, R., Thorpe, J. & Stern, O. (2011) "Structure and Agency in the Politics of a Women's Rights Coalition in South Africa: The Making of the South African Sexual Offences Act, 2007" Developmental Leadership Program Research Paper, 13. http://www.dlprog.org/ftp/download/Public%20Folder/1%20Research%20Papers/Structure%20and%20Agency%20in%20the%20Politics%20of%20a%20Womens%20Rights%20Coalition%20in%20South%20Africa.pdf

Houtzager & Acharya, (2008) "Big Governance Research: Institutional Constraints, the Validity Gap and BIM", *IDS Bulletin*, 38:6, 27-37.

IE4D group, (2011) "Impact Evaluation for Development: Principles for Action". http://www.scalingimpact.net/files/Impact%20Evaluation%20for%20Development%20-%20Principles%20for%20Action.pdf

Jones, N. Jones, J. Steer, L. & Datta, A. (2009) "Improving impact evaluation production and use'" ODI Working Paper 300, Overseas Development Institute, London. http://www.odi.org.uk/resources/docs/4158.pdf

Land, T., Dobunaba, F. & Fargher, J. (2010) "Papua New Guinea sub-national strategy mid-term review, November", Unpublished paper, Canberra: AusAID.

Lederach, J., Neufeldt, R. & Gulbertson, H. (2007) "Reflective Peacebuilding: A Planning, Monitoring and Evaluation Learning Toolkit", Joan B Kroc Institute for Peace Studies, Universoity of Notre Dame, Catholic Relief Services Southeast, East Asia Office, Manila, Philippines.

Leftwich, A. (2011) "Thinking and Working Politically: What does it mean? Why is it important? And how do you do it?" DLP Discussion Paper prepared for the DLP Research and Policy Workshop, Frankfurt, March 2011. http://www.dlprog.org/ftp/info/Public%20Folder/Working%20politically%20workshop%20introduction.pdf.html

Leftwich, A. & Wheeler, C. (2011) "Politics, Leadership and Coalitions in Development: Findings, insights and guidance from the DLP's first Research and Policy Workshop. Frankfurt 10 -11 March 2011", A DLP Research and Policy Workshop Report., http://www.dlprog.org/ftp/info/Public%20Folder/Coalitions%20in%20the%20Politics%20of%20Development.pdf?utm_source=N%26E&utm_medium=Download&utm_campaign=CoalitionsW

Leksmono, C., Young, J., Hooton, N., Muriuki, H. & Romney, D. (2006) "Informal Traders Lock Horns with the Formal Milk Industry: The role of research in pro-poor dairy policy shift in Kenya", Working Paper

266, Overseas Development Institute, London. http://www.odi.org.uk/Rapid/Publications/Documents/WP266.pdf

Maclaren, L. Putra, A. & Rahman, E. (2011) "How Civil Society Organizations Work Politically to Promote Pro-Poor Policies in Decentralized Indonesian Cities", Occasional Paper No. 6 , June 2011, The Asia Foundation. http://asiafoundation.org/publications/pdf/895

Maxson, M. (2012) *The 'Real Book' for story evaluation methods*, May, 2012 edition, Global Giving. http://chewychunks.files.wordpress.com/2012/05/storytelling-realbook-may-23-2012.pdf

O'Brien, T., Payne, S., Nolan, M. & Ingleton, C. (2010) "Unpacking the Politics of Evaluation: A Dramaturgical Analysis", *Evaluation,* 16:4, 431–444.

Odhiambo, O., Holden, S., & Ackello-Ogutu, C. (1998) "Economic Impact Assessment of the Oxfam Wajir Pastoral Development Project – Kenya", Unpublished Report, Nairobi: Oxfam UK/Ireland.

Olken, B. (2007) "Monitoring Corruption: Evidence from a Field Experiment in Indonesia", *Journal of Political Economy*, 115:2, 200-249.

Patton, M. Q. (2011) "The Debate about Randomised Controls in Evaluation: The Gold Standard Question", http://um.dk/en/danida-en/results/eval/reference-documents/20112904rct/

Patton, M. Q. (2012) "Contextual pragmatics of valuing", *New Directions for Evaluation*, 133, 97–108.

Putnam, R. (1993) *Making Democracy Work*. Princeton: Princeton University Press.

Raynor, J. (2010) "Networks vs. Coalitions Distinction or Distraction", American Evaluation Association Presentation November 2010, http://www.berlin-civil-society-center.org/convening/cso-academy/measuring-impact/presentations/

Raynor, J. (2011) "What Makes an Effective Coalition? Evidence-Based Indicators of Success", California Endowment, http://www.mcf.org/resources/evidence-based-indicators-of-coalition-success

Reason, P. & Bradbury, H. (2008) *Action Research: Participative Enquiry and Practice*, London: Sage.

Reinikka, R. & Svensson, R. (2005) "Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda", *Journal of the European Economic Association*, 3:2/3, 259-267.

Serrat, O. (2008a) "Action Learning", Knowledge Solutions, Manila: Asian Development Bank. http://www.adb.org/Documents/Information/Knowledge-Solutions/Action-Learning.pdf

Serrat, O. (2008b) "Appreciative Inquiry", Knowledge Solutions, Manila: Asian Development Bank. http://beta.adb.org/sites/default/files/pub/2008/appreciative-inquiry.pdf

Serrat, O. (2009) "Social Network Analysis", Knowledge Solutions, Manila: Asian Development Bank. http://beta.adb.org/sites/default/files/pub/2009/social-network-analysis.pdf

Snowden, D. & Boone, M. (2007) "A Leader's Framework for Decision Making", *Harvard Business Review*, November 2007.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. & Befani, B. (2012) "Broadening the Range of Designs and Methods for Impact Evaluations: Report of a Study Commissioned by the Department for International Development". DFID Working Paper 38, DFID, London. http://www.dfid.gov.uk/r4d/pdf/outputs/misc_infocomm/DFIDWorkingPaper38.pdf

Sukov, S. (2007) "Storytelling Approaches to Program Evaluation: An Introduction", adapted from an original report by Joseph Tobin & Gustavo E. Fischmann, The California Endowment. http://www.tcfv.org/pdf/prevention/Storytelling%20Approaches%20to%20Program%20Evaluation%20-%20CA%20Endowment.pdf

Taylor, D, & Balloch, S. (2005) "The politics of evaluation: an overview". In Taylor D. & Balloch, S. (eds.) *The Politics of Evaluation: Participation and Policy Implementation*. Bristol: Policy Press, 1–17.

Tilley, N. (2000) "Realistic evaluation: an overview", Paper presented at the Founding Conference of the Danish Evaluation Society, Haifa, Israel, 10 September 2000.

USAID, (2011) "Evaluation Learning from Experience", *USAID Evaluation Policy,* January 2011, Washington, DC. http://usaid.gov/evaluation/USAIDEvaluationPolicy.pdf

DLP Publications

Research Papers

1. Jo-Ansie van Wyk (2009) "Cadres, Capitalists and Coalitions: The ANC, Business and Development in South Africa".

2. David Subudubudu with Patrick Molutsi (2009) "Leaders, Elites and Coalitions in the Development of Botswana".

3. Eduard Grebe with Nicoli Nattrass (2009) "Leaders, Networks and Coalitions in the AIDS Response: A Comparison of Uganda and South Africa".

4. Deborah Brautigam with Tania Diolle (2009) "Coalitions, Capitalists and Credibility: Overcoming the Crisis of Confidence at Independence in Mauritius".

5. Jo Beall with Mduduzi Ngonyama (2009) "Indigenous Institutions, Traditional Leaders and Developmental Coalitions: The Case of Greater Durban, South Africa".

6. Adrian Leftwich (2009) "Bringing Agency Back In: Politics and Human Agency in Building Institutions and States".

7. Junji Banno & Kenichi Ohno (2010) "The Flexible Structure of politics in Meiji Japan".

8. Marine Destrez & Nick Harrison (2011) "Leadership Training and Network Formation: The evolution of the LEAD process".

9. Michael Bratton & Eldred Musunungure (2011) "The Anatomy of Political Predation: Leaders, Elites and Coalitions in Zimbabwe, 1980-2010".

10. Laura Brannelly, Laura Lewis & Susy Ndaruhutse (2011) "Higher Education and the Formation of Developmental Elites: A literature review and preliminary data analysis."

11. Heather Lyne de Ver & Fraser Kennedy (2011) "An analysis of Leadership Development Programmes working in the context of development."

12. Mariz Tadros (2011) "Working Politically Behind Red Lines: Structure and agency in a comparative study of women's coalitions in Egypt and Jordan".

13. Rebecca Hodes, Jennifer Thorpe & Orly Stern (2011) "Structure and Agency in the Politics of a Women's Rights Coalition in South Africa: The Making of the South African Sexual Offences Act, 2007".

14. Sarah Phillips (2011) "Yemen: Developmental Dysfunction and Division in a Crisis State".

15. Eduard Grebe & Minka Woermann (2011) "Institutions of Integrity and the Integrity of Institutions: Integrity and Ethics in the Politics of Developmental Leadership".

16. Andrew Rosser, Ian Wilson & Priyambudi Sulistiyanto (2011) "Leaders, Elites and Coalitions: The politics of free public services in decentralised Indonesia"

17. Monique Theron (2011) "African Trends and Transformation: The Profiles of Sub-Saharan African Executive Heads of State since Independence".

18. Laura Brannelly, Laura Lewis & Susy Ndaruhutse (2011) "Learning and Leadership: Exploring the Linkages between Higher Education and Developmental Leadership".

19. Monique Theron (2012) "Emerging and Non-Emerging African Countires: A Statistical Exploration of the Leadership Factor".

20. Iván Égido, Friedhelm Guttandin, Gustavo López, Jana Leutner, Ulrich Müller, Lena Lázaro (2012)

"Between Hope and Resignation: Perceptions and practices of politics and leadership among the poor in southern South America".

21. Caryn Peiffer (2012) "From Islands of Integrity to Understanding the Politics of Corruption Reduction: Phase One - Statistical identification of positive outliers".

22. Tom Harrison & Genia Kostka (2012) "Manoeuvres for a Low Carbon State: The Local Politics of Climate Change in China and India".

Background Papers

1. Adrian Leftwich & Steve Hogg (2007) "Leaders, Elites and Coalitions: The case for leadership and the primacy of politics in building effective states, institutions and governance for sustainable growth and social development".

2. Adrian Leftwich & Steve Hogg (2008) "The Politics of Institutional Indigenization: leaders, etites and coalitions in building appropriate and legitimate institutions for sustainable growth and social development".

3. Heather Lyne de Ver (2008) "Leadership Politics and Development: A Literature Survey".

4. Heather Lyne de Ver (2009) "Conceptions of Leadership".

5. Adrian Leftwich & Steve Hogg (2010) "The Leadership Program: Overview & Objectives".

6. Adrian Leftwich (2010) "An Overview of the Research in Progress for the Leadership Program: Developmental Leaders, Elites and Coalitions".

7. Isabelle van Notten (2010) "Integrity, Leadership, Women's Coalitions and the Politics of Institutional Reform. Bringing Agency Back In. Mid-Term Workshop Report, Cape Town 25-26 May 2010".

8. Edward Laws (2010) "The 'Revolutionary Settlement' in 17th Century England: Deploying a Political Settlements Analysis".

9. Edward Laws & Adrian Leftwich (2012) "Bringing History Back In: Three Big Books: Summaries and some questions".

10. Edward Laws (2012) "Political Settlements, Elite Pacts, and Governments of National Unity: A conceptual study".

DLP

Policy and Practice for Developmental
Leaders, Elites and Coalitions

DEVELOPMENTAL LEADERSHIP PROGRAM

The Developmental Leadership Program (DLP) addresses an imporant gap in international thinking and policy about the critical role played by leaders, elites and coalitions in the politics of development. This growing program brings together government, academic and civil society partners from around the world to explore the role of human agency in the processes of development. DLP will address the policy, strategic and operational implications of 'thinking and working politically' - for example, about how to help key players solve collective action problems, forge developmental coalitions, negotiate effective institutions and build stable states.

The Developmental Leadership Program

E: info@dlprog.org
W: www.dlprog.org