



Evaluating for security and justice

Challenges and opportunities for improved monitoring and evaluation of security system reform programmes



Simon Rynn with **Duncan Hiscock**

December 2009



Evaluating for security and justice

**Challenges and opportunities for improved monitoring
and evaluation of security system reform programmes**

Simon Rynn with **Duncan Hiscock**

Acknowledgements

Saferworld would like to thank the Governments of Sweden and the United Kingdom for financially supporting this project. A number of other governments and multilateral organisations have provided encouragement or technical support for the research. They include the Government of Australia (AusAID), the Council of Europe and the European Commission, the Governments of Ireland, the Netherlands, Switzerland (Swiss Agency for Development and Cooperation), the United States, and the United Nations (the United Nations Development Programme and the Department for Peacekeeping Operations).

Particular thanks for their inputs into the research process go to Nicole Ball, Mariska van Beijnum, Jenny Chapman, Diana Chigas, Angela Christie, Jim Coe, Clotilde Gouley, Phillipa Haden, Dylan Hendrickson, Eboe Hutchful, Evert Kets, Graham Mathias, Eric Scheye and Victoria Walker.

The case study research was conducted by Sébastien Babaud, Sarah Dalrymple, Catherine Flew, Dorte Hvidemose, Jérôme Mellon, Simon Rynn (all Saferworld); Peter Albrecht, Sean DeBlieck, Jeremy Smith and Jake Sherman.

This report was written by Simon Rynn, and was then revised and edited following comments from reviewers by Duncan Hiscock (Saferworld).



Contents

Executive summary	i
1 Introduction	1
2 Challenges and specifics of the M&E of SSR	4
3 Content and process: What should we be measuring and how?	22
4 Available resources for the M&E of SSR	33
5 The demand for guidance and next steps	38
Bibliography	43

Acronyms

BCPR	Bureau for Crisis Prevention and Recovery (UNDP)
DFID	Department for International Development (UK)
DRC	Democratic Republic of the Congo
EC	European Commission
DPKO	Department of Peacekeeping Operations (UN)
FCO	Foreign and Commonwealth Office (UK)
GAO	General Accounting Office (US)
INCAF	International Network on Conflict and Fragility (OECD DAC)
JLOS	Justice Law and Order Sector (Uganda)
M&E	monitoring and evaluation
MFA	Ministry of Foreign Affairs
MOD	Ministry of Defence
OECD DAC	Organisation for Economic Cooperation and Development (OECD) Development Assistance Committee
OPR	Output-to-Purpose Review
PNG	Papua New Guinea
PRSP	Poverty Reduction Strategy Paper
SIDA	Swedish International Development Agency
SILSEP	Sierra Leone Security Sector Reform Programme
SSR	security system reform
SWAp	sector-wide approach
TOR	Terms of Reference
UNDP	United Nations Development Programme
UNPOL	United Nations Police

Executive summary

THIS REPORT BRINGS TOGETHER the results of a research project on the monitoring and evaluation (M&E) of security system reform (SSR) programmes. It focuses particularly on donor-supported SSR programmes, but with reference throughout to local ownership of and capacity for M&E activities. The report seeks to answer four questions about the M&E of SSR:

- Specific challenges of the M&E of SSR: What challenges apply to the M&E of SSR and security and justice institutions and what, if anything, is distinct about this area?
- Content and process: What should we be measuring when monitoring and evaluating SSR and how?
- Available resources: What existing resources can be drawn upon from within the field or from related disciplines to assist in developing specific guidance on M&E of SSR?
- Demand: Who are the most obvious users of tailored guidance on this subject and what do they need?

The report does not in itself constitute a guidance document on the M&E of SSR, but provides material from which tailored guidance could be prepared to meet the needs of interested parties.

Specific challenges of the M&E of SSR

The research identified a large number of challenges and specific issues relating to the M&E of SSR, which can be grouped together into six broad categories:

- General challenges which are familiar when undertaking M&E in any area
- SSR-specific challenges
- Challenges associated with donor policy
- Challenges generated by donor bureaucracy and politics
- Challenges arising from poor programme design and management
- Contextual challenges in fragile and conflict-prone environments

General, 'familiar' challenges

Within the first category of general, 'familiar' challenges, the most significant is the frequent resistance to M&E processes, with project staff and even management often perceiving it to be burdensome, unnecessary or threatening. There are often weak incentives to invest in M&E systems and to ensure take-up of the information and analysis that is provided. The research also noted that donor-driven, results-oriented M&E can feel imposed and can undermine local ownership. This includes the setting

of targets. While targets can be helpful, targets can also distort priorities and ways of working as managers focus on meeting targets at the expense of other actions ('what gets measured gets done'), and targets are sometimes achieved through perverse actions. Moreover, a regular problem is that project managers have not devoted enough time to developing, agreeing and building support among key stakeholders for a methodology to measure change, which often leads to problems further into the programme.

SSR-specific challenges

There are a number of SSR-specific challenges. This does not necessarily mean that these challenges are entirely exclusive to SSR programmes, but rather that such challenges are more frequent, more pronounced and require M&E systems and processes to be adapted accordingly:

- **Complexity of the security and justice sectors** Security systems have many constituent parts and each sub-sector differs significantly in its purpose, functioning and orientation. It is thus hard to track and evaluate changes, especially as one sub-sector may be affected by several others. It is also a challenge to find evaluators who have knowledge of both (cross-sectoral) SSR and M&E, especially from developing countries.
- **No ideal security system** There is no one international blueprint for what a security system or its constituent parts should look like, and thus a lack of agreed international models against which to measure outcomes.
- **Cultures of secrecy** SSR inevitably touches on sensitive security matters, and entrenched cultures of secrecy (both in partner country institutions and donors) can restrict access to relevant information. Secrecy can also result in a restricted pool of evaluators, since evaluators sometimes require security clearance.

Challenges associated with donor policy

Donor policy on SSR also presents some challenges. The OECD DAC Handbook advocates an integrated approach to security and justice programming. This makes coherence an issue for both monitoring and evaluation. Evaluations should consider coherence across agencies, sectors and policies at both partner government and donor level, but should recognise the many practical obstacles to integrated approaches to SSR. Evaluations should also consider how SSR contributes (indirectly) to national development strategies and processes. Another point is that donors should monitor and evaluate how well their programmes uphold the principle of conflict sensitivity, yet in practice this has rarely been done. Evaluations also need to be conflict-sensitive themselves.

Challenges generated by donor bureaucracy and politics

There are also a number of issues around donor bureaucracy and politics. In some cases, M&E can identify problems and suggest better policy and practice; in other cases, there are deeper, structural issues which M&E systems can only acknowledge and adapt to appropriately. Some of the most frequent issues for SSR include:

- **Poor local ownership of M&E** Donor-supported SSR often promotes interventions more in line with the donor's needs and views than those of partner countries. Many SSR programmes do not pay sufficient attention to building partner country M&E capacities and information management systems, and thus M&E is often seen as solely a donor agenda.
- **Vague/hidden objectives** Programme goals are sometimes intentionally ambiguous, either to give the programme space to develop and/or because the programme is guided by deliberately unstated political objectives.

- **Incoherence and lack of co-ordination** Different donor departments still have different mandates and perspectives on SSR, which extends to M&E. A lack of coherence between departments means evaluators can struggle to understand multi-budget and multi-strand programmes.
- **Staff turnover and workload** High staff turnover, a lack of guidance and heavy workloads all dissuade or prevent field staff from routinely collecting and analysing information on the progress of their SSR projects.

Challenges arising from poor programme design and management

Many of the problems that arise regarding M&E of SSR programmes can be traced back to poor programme design. Many SSR programmes make little preparation for M&E during project design; some do not address the issue in project documents at all. M&E is often under-resourced, rarely adhering to the rule of thumb in other spheres that 5–10 percent of project budgets should be spent on M&E. Participation is also an issue: Reviews and evaluations usually lack sufficient input from e.g. marginalised groups, beneficiary populations, rural inhabitants, and there is often undue reliance on information from security sector institutions and partner government officials.

Contextual challenges in fragile and conflict-prone environments

Lastly, M&E should always be sensitive to the context, and this is particularly important in fragile and conflict-prone environments. Where there are multiple power centres and informal methods dominate the style of governance, M&E techniques must mirror this, but current M&E frameworks tend to be defined by the donor's institutional needs and are not sufficiently adapted to the local context. The other main challenge is to collect appropriate information for M&E purposes. Project managers find it hard to set baselines using primary research and external data sources, given time pressure and the difficulty of collecting information, especially as national systems and capacities for data collection, handling, storage and analysis (including those of non-governmental actors and oversight bodies) are often weak.

Content and process

Beyond the challenges identified above, the fundamental questions for the M&E of SSR are *what* exactly should be measured, and *how*? The report breaks this down into four main areas:

- Levels and units of analysis for the M&E of SSR
- Theories and dimensions of change
- Evaluation criteria
- Ownership of and participation in the M&E process

Levels and units of analysis

M&E can be carried out at several distinct levels and units of analysis:

- | | |
|---|---|
| Programme/
institutional
level | <ul style="list-style-type: none"> ■ Donor programming efforts, or aspects such as the performance of contractors or conduct of the donor (e.g. in line with aid effectiveness standards such as ownership, alignment, harmonisation, co-ordination or in line with stated SSR policy). ■ National security and justice institutions (their performance being subject only in part to donor efforts where any support is provided). |
|---|---|

Sector or system level	<ul style="list-style-type: none"> ■ Donor support to security and justice sectors/systems (e.g. through sector-wide approaches and sector frameworks, budget support and similar). ■ National SSR 'processes' (the broader state and trajectory of security system performance and governance in a country, region or territory). ■ Linkage of either of these to national development or peacebuilding frameworks (e.g. poverty reduction strategy papers, transitional results matrices and associated surveys and dialogue processes).
Country/context level	<ul style="list-style-type: none"> ■ May cover political and security aspects, but might also touch on socio-economic aspects – can be part of ongoing context, conflict or risk monitoring.
Policy level	<ul style="list-style-type: none"> ■ General compliance of donors or national governments with relevant policy standards on SSR across a range of programmes and contexts. ■ Relevance, appropriateness and impact of such policies.

Each level is a legitimate and important focus for M&E, but has different political significance and technicalities. In practice, most M&E of SSR is currently project- or programme-focused, since donors wish to know whether their project 'works' and the project level provides relatively easy unit of analysis through project documents and logframes. However, given the shift towards whole-of-government and cross-sectoral approaches, sector/system level M&E is likely to increase. M&E at this level is likely to be impact- as well as outcome-focused, concentrating more on how national security systems and institutions functions and how this affects beneficiary populations, working backwards to assess the contribution of donors to observed changes. Tracking progress at country/context level allows data to be gathered on often neglected questions around drivers of change, political will and conflict dynamics, but it is harder to attribute impact to individual programmes. Evaluating donor performance or policy compliance is rare, since SSR reviews tend to be initiated and owned by donors rather than partner governments.

Theories and dimensions of change

Often, the underlying theory of change informing the design of an SSR programme is opaque. Evaluations should explore theories of change in more detail, though they rarely do: reviews and evaluations tend to concentrate on tracking progress towards programme goals and objectives, and ignore important questions about whether programme design is appropriate to the context, addressing the right issues with the right methodology. Evaluators usually need to talk at length with programme staff (and possibly their predecessors) in order to unravel the thinking behind the project, before then analysing the strengths and weaknesses of this approach.

When designing monitoring systems and planning evaluations, it is also necessary to be clear about the 'dimensions' of change which the programme is trying to influence. SSR programmes can seek change along a large number of different dimensions, each of which can be addressed at several result levels (impact, outcome, activity/process). Some of the main dimensions include:

- Improved service delivery
- Enhanced security
- Enhanced and equitable access to justice
- Institutional and human capacity
- Security policy
- Good and democratic governance and rule of law
- Oversight and accountability
- Ownership of reform processes
- Enhanced civilian involvement in and oversight of security systems
- Non-state security and justice provision
- Conflict sensitivity
- Changes to political dynamics
- Sustainability

- Financial and resources management
- Cross-cutting issues such as gender, human rights, and poverty reduction.

Evaluation criteria

This report proposes eight evaluation criteria which can be used to inform the design of evaluations. These criteria build on the five standard OECD DAC development evaluation criteria, taking into account the specifics of SSR programmes and policy and the dimensions of change noted above:

- Relevance/appropriateness
- Effectiveness
- Efficiency
- Impact
- Sustainability and ownership
- Coherence
- Co-ordination/linkages
- Consistency with values.

The full report provides a list of prompt questions for each criterion (see 'Criteria and prompt questions for evaluation of SSR programmes' in Section 3.4).

Beyond these criteria, another overarching question for any evaluation is: *what has been learnt throughout the life of this programme?* Every programme and activity will provide lessons and project staff (in both the partner government and the donor organisation) are likely to have learnt a great deal that evaluators can draw out and document.

Ownership of and participation in the M&E process

How M&E is carried out is just as important as what is done. At the design stage, there may be little appetite for a Western-style M&E framework among local actors. Donors therefore need to start by raising the issue of M&E and exploring the attitudes of local actors towards what kind of M&E should be undertaken and why. Donors should avoid presenting M&E as an external imposition; stakeholders should work together to agree how change will be measured, an important step towards establishing local ownership from the start. Donor-led SSR assessments should also analyse partner country capacity to undertake M&E and where appropriate make the capacity-building of national M&E systems and integral part of SSR programme design.

For the duration of the SSR programme, joint ownership of M&E processes between the donor and the partner government is required, and also with beneficiaries and non-governmental actors as far as this is possible and appropriate. The programme should gradually transfer ownership of the M&E system as local capacity increases, so that by the end of the programme the M&E system is fully locally owned and more likely to be sustainable.

All aspects of the M&E process should encourage the greatest possible participation from different stakeholders, including all relevant national state actors (e.g. other relevant ministries and agencies beyond the security sector, parliamentarians, committee representatives, opposition politicians) and broader civil society (e.g. beneficiary representatives, civil society organisations, business representatives). Reviews and evaluations should consult with a wide variety of voices representing different social categories and constituencies, and should be presented in a format that is accessible to local actors.

Available resources for the M&E of SSR

A great deal of general information and advice on M&E is available, though there is relatively little that is specifically targeted to SSR/rule of law practitioners. The report identifies a number of publications that appear most relevant, grouped into four categories (Section 4 in the full report):

- SSR-specific material
- Performance measurement in developed countries
- M&E in fragile and conflict-affected environments
- General M&E publications.

The report does not attempt to promote particular tools for the M&E of SSR over others: all are useful, but no one document or system will be appropriate on its own. Many are not adapted to the realities of less developed, fragile and conflict affected states. Furthermore, many neglect or do not deal well with dimensions like public demand, empowerment and satisfaction, which is necessary as a way of cross-checking information and building local accountability and a voice for civil society on security and justice.

The demand for guidance and next steps

Although some guidance on aspects of security and justice sector M&E has been developed, nothing comprehensive is available that covers both process and substance, both monitoring *and* evaluation, and reflects the latest policy thinking. This research has sought to identify potential users of specific guidance on M&E of SSR, their needs, the factors that drive demand for M&E, and any obvious upcoming opportunities to incorporate new guidance into different organisations' procedures and work-plans.

In general, there is high interest in most donor countries in improving M&E practice for SSR work, driven in particular by the increasing international emphasis on results-based management frameworks for development aid and the aid effectiveness agenda. This is beginning to extend beyond aid institutions into other government departments and multilateral agencies. Moreover, there is recognition that current practice on M&E on SSR is insufficient and requires improvement.

Although guidance needs to be tailored to specific agencies and contexts, it appears that donor governments share enough understanding of the content of SSR and sufficient institutional characteristics to warrant the development of a general guidance product. The primary audience would be the donor programme staff who bear most responsibility for designing programmes and commissioning evaluations. The report sets out some essential points to include in any guidance, including:

- Definitions and purposes of M&E
- The importance of M&E for effective SSR
- What is specific about the M&E of SSR
- Levels and focuses for M&E
- Theories and dimensions of change
- Evaluation criteria and prompt questions
- Forming legitimate and skilled evaluation teams
- Participation in and ownership of M&E processes
- Building the capacity of partner country M&E systems.

Saferworld intends to develop guidance that can easily be tailored to the needs of specific users. It will contribute to the preparation of a practical toolkit on the M&E of security and justice programming for the OECD DAC's International Network on Conflict and Fragility (INCAF), and may in time produce further guidance beyond this. Saferworld also intends to consult on and test any guidance that is prepared to clarify how such products can be used most effectively.

1

Introduction

THE DONOR COMMUNITY has made good progress in recent years in developing coherent and progressive policy and guidance on aspects of security system reform (SSR). Two of the most important advances are the 2004 Organisation for Economic Development (OECD) Development Assistance Committee (DAC) policy statement and paper *Security System Reform and Governance: Policy and Good Practice* and the subsequent guidance provided in the 2007 *OECD DAC Handbook on Security System Reform: Supporting Security and Justice*. The OECD DAC Handbook is the first comprehensive framework to assist those involved in SSR work in designing and implementing SSR projects. It is also clearly and explicitly norm-driven, stipulating guiding and strategic principles for donor support to SSR processes.

On some aspects of SSR, however, practical guidance is still lacking. Monitoring and evaluation (M&E) is one such area where little or no guidance is so far on offer.¹ There are several reasons for this, including a historic lack of agreement among the SSR ‘community’ as to what the term ‘security system reform’ actually means. Views on the purpose of SSR also differ. Over the years, the discipline has variously been claimed for the development, governance, democratisation, security and conflict prevention fields, and the various donor government departments that engage in SSR may have contrasting views of its purpose and scope. Although OECD DAC policy and guidance has provided much greater clarity, the reform of security and justice systems remains a complex and challenging undertaking, with many areas of uncertainty.

Scoping research by Saferworld in early 2008 identified a lack of clarity among those involved in SSR work at both the national government and donor levels on how to measure the worth of SSR interventions and revealed an appetite for tailored guidance on the topic. A subsequent desk review examined a wide range of donor project documents, progress reports, reviews and evaluations from the SSR and rule of law fields. This concluded that where SSR interventions have been evaluated, the methods and assumptions used in these evaluations have varied considerably. It was harder to find relevant information about monitoring procedures, but it appeared that monitoring practice also varied.

In light of these initial findings, in Spring 2008 Saferworld began a research project titled ‘Evaluating for Security: Developing guidelines on monitoring and evaluating Security System Reform interventions’. Saferworld’s approach was to work collaboratively with OECD DAC members to identify lessons from past attempts to monitor and evaluate SSR and then to combine those lessons with best practice from M&E in

¹ The OECD DAC Handbook does however devote about three pages (pp 240–243) to the subject, providing a basic introduction to issues such as evaluation criteria, partner country participation in evaluations and the need to prepare for M&E early on in programmes. OECD DAC, *OECD DAC Handbook on Security System Reform: Supporting Security and Justice*, (OECD, 2007).

other fields. Using current OECD DAC policy and guidance on SSR as the normative and technical starting point, the project began with three strands of research:

- Field case studies to identify lessons learnt regarding monitoring and evaluation during past SSR/rule of law programmes;²
- Qualitative surveys of the M&E systems used by the main international donors for SSR work, together with donors' needs in this area and any obvious trends or opportunities that might be taken up;³
- Desk research covering SSR/rule of law project documents, reviews and evaluations, and a review of guidance on M&E in related fields such as governance.

In addition, throughout the project we have linked with issue specialists from the M&E, development, conflict prevention and SSR fields, either through interviews or by requesting their input as reviewers and advisers.

The research focused particularly on donor-supported SSR programmes, both because the majority of recent attempts at SSR have depended on donor assistance and because donor-supported programmes have tended to maintain much greater documentation and evidence of M&E activities (thus making case studies and related research a more manageable prospect). However, it should be emphasised from the start that M&E can take place, and be owned at two distinct levels. The first is at the level of the national government which is undertaking the security system reforms. The second, which is currently much more frequent, is at the level of the donor organisation which is providing assistance to the SSR programme.⁴ These two levels are obviously linked in various ways, as this report will discuss, but the purpose and form of M&E may differ depending on the requirements of whoever is driving the M&E process. This report is written with this in mind: many of the challenges and potential solutions that we present here apply at both levels, but where necessary, the report clarifies where an issue is only relevant to donor-assisted SSR and donor M&E processes. It also highlights challenges relating to the weak M&E capacity of many national governments, and the need for SSR programmes to address this – an issue that ultimately goes beyond M&E frameworks into the programme design of donor-assisted SSR programmes.

This report is a synthesis of the findings from the three strands of research listed above. It seeks to answer four key questions which the project identified as pressing concerns:

- 1 Challenges and specifics for SSR: What challenges apply to the M&E of SSR and security and justice institutions and what, if anything, is distinct about this area?
- 2 Content and process: What should we be measuring when monitoring and evaluating SSR and how?
- 3 Available resources: What existing resources can be drawn upon from within the field or from related disciplines to assist in developing specific guidance on M&E of SSR?
- 4 Demand: Who are the most obvious users of tailored guidance on this subject and what do they need?

The report thus identifies the key dimensions and challenges of the M&E of SSR and indicates the areas which most urgently require further attention and development. As such, the ideas presented in this report could form the basis for a guidance document on the M&E of SSR, but this report itself is not intended as a guidance product or handbook; Saferworld intends to produce more tailored guidance on the M&E of SSR in the future. In the concluding section of the report we make some initial suggestions regarding the content, format, testing and outreach of such guidance.

² The following case studies were completed in mid- to late 2008: AusAID-supported justice sector reforms in Papua New Guinea (PNG); Reforms to the Justice, Law and Order Sector (JLOS) in Uganda; SDC-supported police reform in Bosnia and Herzegovina; the UK-supported Sierra Leone Security Sector Reform Programme (SILSEP); the United Nations Development Programme (UNDP) 'Support to Security Sector Reform' project in Albania. The case study research was carried out using document reviews and semi-structured interviews with representatives from national governments and their security and justice institutions, evaluators, civil society and 'user' groups and donor governments.

³ The following surveys were completed in mid- to late 2008: European Union, Government of the Netherlands, Government of the United Kingdom, Government of the United States, United Nations.

⁴ This distinction may not hold in SSR programmes undertaken in transitional administrations such as East Timor or Kosovo, where international actors have executive functions regarding local security provision.

1.1 Defining monitoring and evaluation

Before entering into a more detailed discussion of the M&E of SSR, it is worth beginning by clarifying the meanings and purpose of monitoring and evaluation (and how they differ). While definitions of ‘monitoring and evaluation’ abound, this paper takes the OECD DAC’s standard definitions unless otherwise specified:⁵

Monitoring: *A continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds.*

Review: *An assessment of the performance of an intervention, periodically or on an ad hoc basis. Note: Frequently “evaluation” is used for a more comprehensive and/or more in-depth assessment than “review”. Reviews tend to emphasize operational aspects. Sometimes the terms “review” and “evaluation” are used as synonyms.*

Evaluation: *The systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors.*

Evaluation also refers to the process of determining the worth or significance of an activity, policy or program. An assessment, as systematic and objective as possible, of a planned, on-going, or completed development intervention. (Note: Evaluation in some instances involves the definition of appropriate standards, the examination of performance against those standards, an assessment of actual and expected results and the identification of relevant lessons.)

In addition to these definitions, it is also worth mentioning the guiding principles for evaluation of development co-operation published by the OECD DAC in 1992. These emphasise qualities such as openness, partnership and feedback and identify the main purposes of evaluation as being:

- To improve future aid policy, programmes and projects through feedback of lessons learnt; and
- To provide a basis for accountability, including the provision of information to the public.⁶

These definitions demonstrate that monitoring and evaluation are complementary but different processes. Monitoring is a continuous process throughout the implementation of a project or programme that allows managers to track how well it is performing against agreed plans, timelines, and objectives – and to highlight any challenges that may require project plans to be adapted. Its main purpose is thus to maintain and improve the quality of the intervention while it is taking place. Evaluation, by contrast, is focused more on learning lessons and thus improving the quality of future programmes.⁷

⁵ OECD DAC, *Glossary of Key Terms in Evaluation and Results Based Management*, (OECD, 2002).

⁶ OECD DAC Network on Development Cooperation, *Evaluating Development Cooperation, Summary of Key Norms and Standards*, (OECD, 2008), p 7.

⁷ Department for International Development (UK), *Guidance on Evaluation and Review for DFID Staff*, (DFID Evaluation Department, 2005).

2

Challenges and specifics of the M&E of SSR

BOTH SSR AND M&E are broad, complicated disciplines in their own right. It is hardly surprising therefore that there are many challenges and specific issues relating to the M&E of SSR. This section groups together these challenges and particularities into six broad categories:

- ‘Traditional’ challenges to effective M&E
- Sector-specific challenges
- Challenges arising from policy
- Challenges generated by donor bureaucracy and politics
- Challenges arising from poor programme design and management
- Contextual challenges

We discuss various difficulties that may arise in each category, and where relevant we identify potential means of overcoming these obstacles.

2.1 ‘Traditional’ challenges to effective M&E

This research identified various challenges that are common to M&E in many circumstances and are also relevant to the M&E of SSR. These include:

- Bureaucratic and personnel challenges
- Managerialist, results-oriented approaches undermining local ownership
- Agreeing how to measure change
- Targets distorting priorities and ways of working
- Attribution challenges
- M&E serves multiple, sometimes competing purposes
- Ensuring take-up of M&E findings

Bureaucratic and personnel challenges

Some of the biggest challenges to effective M&E are caused not by technical difficulties regarding methodologies, indicators sets, logical frameworks, etc, but by the bureaucratic and personal attitudes of staff and institutions (both in governments undertaking reform and in donor organisations) towards the M&E process. Perceptions of the incentives for undertaking M&E are particularly important. Interviewees from several quarters, including many evaluators, acknowledged that both partner government officials and donor project staff can perceive M&E as burdensome. It can even be

threatening, since they may fear that the evaluators – or decision-makers who use the evaluation findings – may not adequately understand the context and the constraints within which they are working, and thus that any ‘negative’ evaluation will be seen as a direct judgement on them. To overcome this, partner government officials and field-level donor staff need to be encouraged to engage positively with the M&E process and evaluation findings. A combination of adequate training and resourcing, prior planning, leadership and financial and managerial incentives are needed to embed and safeguard M&E.

The Uganda case study showed some of the incentives that can be used to help overcome such problems at a programme level, including output-focused national budgeting, building responsibility for M&E into job descriptions and contracts and negotiating reduced intrusion by donors in return for enhanced national attention to M&E. At donor country headquarters, the incentives are likely to be different. In most cases, donors already have established M&E systems, but they may not work as well as expected. Creating time and incentives to ensure that evaluation findings are taken up (as opposed to simply monitoring the process and briefly reviewing the final outcome) is a particular challenge at higher strategic levels.

Managerialist, results-oriented approaches undermining local ownership

Managerialist approaches to public services (including the security system) which emphasise results-oriented management may be unfamiliar to partner governments and can sometimes be seen as a foreign imposition. Hence, partner government officials may be concerned with how M&E tools will be used to manage the project, particularly if they perceive that these tools are largely designed and owned by the donor and will primarily be used by the donor for its own project management. This was illustrated by the case study in Papua New Guinea (PNG), where AusAID interviewees noted that in the PNG justice institutions with which they work, senior management historically displayed limited commitment towards outcome-oriented M&E. Research has identified other cases in which national partner institutions eschewed a commitment to measuring for results or even to setting indicators.⁸ This may reflect broader ownership problems and capacity limits, as well as a retreat from ‘accountability’.

Our Sierra Leone case study suggested that it might be useful to look at the quality of donor project management and the appropriateness of the resources and staffing allocated to programming by including these factors either in indicator sets or in guidance for external evaluators.

Agreeing how to measure change

Our field research identified several lessons regarding the methods that are used to measure change. In terms of process, the key lesson was that sufficient time must be allowed to develop, agree and build support for the methodology among key stakeholders (national government officials and decision-makers, donor representatives, local civil society, etc). These stakeholders must have a shared understanding of what research methods will be used to measure change. If indicators are part of the methodology (as they usually are), there should be mutual agreement on which indicators will be measured, why, and how. In the case studies in PNG and Uganda, which looked at longer-term, more sustained programmes, indicators underwent several revisions over time as the understanding and priorities of stakeholders changed. These cases also showed that M&E frameworks, and indicators in particular, can be an important test of national ownership. Lest donors end up measuring only for their own use, it is crucial to establish and maintain an open dialogue about what success looks like and how it should be measured.

⁸ Mancini F, ‘The Company We Keep: Private Contractors in Jamaica’ in Peake G, Hills A, and Scheye E (eds), *Managing Insecurity: Field Experiences of Security Sector Reform*, (Taylor and Francis, 2006).

Targets distorting priorities and ways of working

Another familiar challenge with results-oriented approaches is that setting targets based on indicator frameworks can at times lead to unpredictable outcomes. While setting targets helps project managers to focus on clear goals, it can also distort project implementation by skewing priorities and ways of working to suit the targets, at the expense of other actions which may be just as important ('what gets measured gets done'). Similarly, where projects have limited resources, there is a tendency to allocate resources not to the most important tasks but to those where the targets are easiest to meet. Even worse, targets may create perverse incentives as managers try to meet the target in ways that run contrary to the spirit of higher-level objectives.

This appears to be a common problem in reforming security systems, as shown by the field research carried out for this project; it is also acknowledged in some of the existing guidance products.⁹ For example, our Uganda case study found particularly illustrative examples of targets being met by perverse actions: a target to achieve a decrease in case backlog rates led to an apparent reduction in the quality of justice, while a target to reduce crime rates resulted in more illegal arrests. Such targets were inadvertently encouraging actions that go against due process and ultimately were undermining the rule of law. Hence, any agreed indicators need to be monitored carefully during their use, and reviewed over time so that they can be refined or dropped as programme staff and stakeholders better understand what is most important to measure and as monitoring reveals which indicators are most/least suitable.

Attribution challenges

As in many other fields, it can be difficult not only to measure high-level impacts and outcomes, but also to untangle causes and results and attribute change to particular programmes and interventions. As evaluators well understand, there may be any number of intervening variables between the programme inputs and any observed changes. For example, SSR programmes might use indicators related to perceptions of public safety, crime rates, or border crossing statistics to measure success or failure, but there are many factors that might influence these indicators beyond the specific programme activities.¹⁰ In addition, other interventions in and around the security and justice sectors may make it even harder to link any observed impact-level changes to specific activities.

This is a problem that evaluators have long recognised and various remedies have been developed to tackle this. These range from very thorough (and expensive) impact evaluations using statistically valid measures, control groups and so on, through to a retreat from impact assessment towards a more general focus on outcomes.¹¹ More recently, the OECD DAC has developed draft guidance on evaluating conflict prevention and peacebuilding efforts, which covers the question of impact assessment in complex environments. This has led to a more elaborate definition of impact in this area, though the claim that impact can be tracked at the level of individual projects¹²

⁹ See for example Foglesong T et al, *Measuring Progress toward Safety and Justice: A Global Guide to the Design of Performance Indicators across the Justice Sector*, (Vera Institute of Justice, 2003).

¹⁰ This can mean either that a programme may not have been effective despite apparent improvement, or that a programme may have been beneficial despite a decline in key indicators. For example, a lessons learnt paper by the Inter-American Development Bank on its violence prevention and citizen security work in several countries concluded that project work had been effective despite dire negative trends in specified indicators such as homicide and violent assault rates and public perceptions of security. The negative trends were attributed to a serious economic crisis and high unemployment but also to improved police performance which led to better recording of crime statistics. Buvinić M, Alda E and Lamas J, *Emphasizing Prevention in Citizen Security: The Inter-American Development Bank's Contribution to Reducing Violence in Latin America and the Caribbean*, (Inter-American Development Bank, 2005).

¹¹ For an example of measuring change at the outcome level, see: www.outcomemapping.com.

¹² 'In the view of the CDA team, the ... common preoccupation with the "remoteness" of impacts must not divert evaluation attention from assessing the many effects of conflict prevention and peacebuilding programmes on the conflict, whether intended or not. Evaluations should take account of individual programme impacts and cumulative, multi-programme impacts at the strategic or policy level.' Anderson M et al, *Encouraging Effective Evaluation of Conflict Prevention and Peacebuilding Activities: Towards DAC Guidance*, (OECD, 2007).

is at odds with other research findings.¹³

The literature on impact assessment suggests that though it is infinitely complex, evaluators can find simple methods for impact assessment that are acceptable given the circumstances. For SSR, particularly as seen from the donor/OECD DAC angle, one favoured solution is to start by acknowledging the reality of limited resources, weak programme design, hurried evaluations and the absence of meaningful monitoring information. Rather than undertake very detailed, World Bank-style impact assessments we suggest that donor M&E should focus first and foremost on tracking outcomes (which is challenging in itself), and then on establishing what contribution if any programme inputs and outputs have made towards any observed changes. If indeed SSR is ‘the art of the possible’¹⁴, the M&E of SSR must surely be the art of the plausible.

M&E serves multiple, sometimes competing purposes

The standard definitions of monitoring and evaluation provided above establish the purpose of M&E in positive terms, as ensuring progress and learning lessons for the future. Hence, monitoring seeks to indicate ‘the extent of progress and achievement of objectives’, while evaluation is about ‘enabling the incorporation of lessons learned into the decision-making process of both recipients and donors’ and ‘determining the worth or significance of an activity, policy or program’ (see Section 1.1). In practice, M&E does not exist in a vacuum, and is frequently influenced by other organisational demands and broader political concerns. In fact, evaluations may be commissioned for various purposes – often simultaneously – and monitoring information and evaluation reports may be used and interpreted in a variety of ways.

Most notably, evaluation findings are often used by senior decision-makers to inform funding decisions – whether to make spending cuts, or to agree to continue funding a programme – or to demonstrate that their decision to fund a programme was justified. This can become an impetus for programmes to undertake M&E. For example, our Albania case study found that both studies of the Support to the Security Sector Reform Programme published in 2006 ‘shared a common purpose: to encourage prospective donors to contribute funds to the programme’.¹⁵ Whether this objective is overtly stated or not, if evaluators and/or project staff (and indeed, other interested stakeholders, such as partner government beneficiaries) understand that the findings will be used to ‘sell’ the project, this will have a profound impact on how information is gathered and presented. In such circumstances, it is impossible to see the evaluation as a ‘lessons learnt’ exercise, since many stakeholders will have a strong incentive not to admit mistakes, let alone reflect on what could have been done differently and more effectively.

At the other end of the scale, many evaluations are perceived as little more than ‘box-ticking exercises’: they are being done because they have to be done, but will lead to no visible changes (see ‘Ensuring take-up of M&E findings’ below). In such circumstances, many stakeholders may see no reason to participate in the evaluation any more than absolutely necessary, again reducing the chances that useful lessons will arise from the evaluation.

¹³ In 2004, the Joint Utstein Study of Peacebuilding for the Norwegian Ministry of Foreign Affairs concluded that it is virtually impossible to measure impact at the project or programme level for peace and security programmes, and recommended that the emphasis should shift to measuring the impact of overarching ‘peacebuilding strategies’, evaluating how different projects ‘add up’ and contribute to ‘peace writ large’. Smith D, *Towards a strategic framework for Peacebuilding: Getting their act together – Overview report of the Joint Utstein Study of Peacebuilding*, (Royal Norwegian Ministry of Foreign Affairs, 2004). Similarly, a recent report for Sightsavers International highlighted the challenges of conducting a meaningful impact assessment of single projects or organisations, and recommended that evaluators consider ‘us[ing] impact assessment to try to assess the relative impact of different approaches NOT to try to aggregate the overall impact of the organisation’s work’. Chapman J and Mancini A, *Impact assessment: Drivers, dilemmas and deliberations*, (Sightsavers International, 2008).

¹⁴ Nathan L, *Operationalising the Principle of Local Ownership in Security Sector Reform*, June 2006.

¹⁵ DeBlicek S, *Monitoring and evaluation arrangements for the Support to Security Sector Reform Programme in Albania: a case study*, (Saferworld, 2009), p 8.

Between these two extremes there may be several overlapping and competing purposes to evaluations: is the evaluation more interested in financial probity and 'efficiency' (normally measured over the short term) or effectiveness and long-term impact? Is the evaluation assessing how the programme fared according to its stated goals, or is it trying to 'backwards report' how the programme meets higher organisational objectives, even if these objectives were never directly referred to during its design? Are those who commissioned the evaluation doing so because they wish to learn from the programme, or because the issue has suddenly become politically high-profile and there are demands for 'answers' from politicians and the media?

In reality, it will never be possible to completely reconcile or avoid these competing visions of how the information generated by M&E processes should be used. As far as is practical, the users and owners of monitoring information, those commissioning the evaluations and the evaluators themselves should consciously recognise – even if only in private – that this information may be used in various ways. With this in mind, wherever possible evaluations should focus on identifying lessons learnt.

Ensuring take-up of M&E findings

A final, extremely common challenge for M&E is ensuring that the findings of M&E processes, particularly evaluations, are taken on board and that this results in changes to policy and practice. This can apply both at the national government level and within the donor organisation.

In the case studies undertaken for this research, all of which looked at donor-assisted SSR, take-up of review findings by donor organisations was patchy (though since these case studies looked at project/programme level M&E, some variation in the degree of action taken is to be expected). Several factors may influence the degree of take-up, including how much senior managers in the donor organisation perceive M&E to be a priority, how far managers and project staff believe that the evaluation process was appropriate and that its conclusions are valid, whether funding is available for follow-up activities, and what options exist at that point for changes to programme design and implementation. It also needs to be acknowledged that policy decisions are not always based only on available evidence, but are often influenced or even driven by other political concerns. Evaluations may thus be ignored or rejected because of wider political considerations. Given that SSR programmes are always politically sensitive and sometimes very high-profile, such considerations often come into play – though good-quality M&E can ensure that decision-making, while still attuned to these political factors, is rooted more deeply in a solid evidence base.

Most of the reviews and evaluations that were accessible to the research team were annual or mid-term reviews that are routinely held regardless of demand from programme stakeholders. While a standard requirement to undertake project reviews and evaluations during or immediately following a project may be positive, there is a risk that these may become mere 'box-ticking exercises'. (The UK Department for International Development (DFID)'s in-house guidance on reviews and evaluations recognises that this may happen with annual reviews and suggests remedial action.¹⁶) As far as the research could identify, none of the five surveyed donors required formal actions in response to evaluations of this kind.

Most donor organisations studied by the research team were perceived – both internally and externally – to need better systems and incentives to promote learning and action from evaluations. One donor official interviewed went further, adding that 'upward accountability for reviews and lessons learnt just isn't there', the implication being that senior management needs to be involved in the evaluation and follow-up processes. If senior decision-makers do not have the time and incentives to follow through on recommendations, a response is obviously far from guaranteed. Moreover,

¹⁶ Op cit DFID, *Guidance on Evaluation and Review*, 2005.

for project/programme-level reviews, the stakes are often not high enough to hold decision-makers' attention. By contrast, more strategic M&E exercises (e.g. periodic reviews of the UK Conflict Prevention Pools) are more likely to garner attention and obtain a high-level response (though even then, this may not actually lead to any notable action being taken).

Another difficulty is that evaluation reports may be comparatively long and complex, while the time to read and respond to them is often limited. For example, DFID annual reviews and Output-to-Purpose Reviews (OPRs) sometimes contain long lists of recommendations which are not often prioritised or broken down into action steps. Furthermore, many evaluation reports do not contain a good summary that decision-makers can easily absorb.

An OECD DAC publication on using evaluation feedback to improve lesson learning and accountability suggests a number of techniques that can be used to reach different audiences:¹⁷

- take steps to understand how learning happens within and outside the organisation, and identify where blockages occur;
- assess how the relevance and timeliness of evaluation feedback can be improved, and take steps to ensure this happens;
- develop a more strategic view of how feedback approaches can be tailored to the needs of different audiences;
- put much more effort into finding better ways of involving partner country stakeholders in evaluation work, including the feedback of evaluation lessons; and
- take steps to increase the space and incentives for learning within the organisation (both from evaluations and other sources).

Three conclusions can be drawn from this. Firstly, strategic level reviews, synthesis evaluations and subsequent policy briefs are important for generating interest in programmatic issues, particularly at more senior levels within donor organisations. Secondly, it is probably necessary to require or incentivise donor staff to consider, respond to and act on evaluation findings, provided this sits well with overall project management. The European Commission's (EC) new commitment to issuing responses to evaluation reports is a useful example of how donor staff can be spurred towards responding. Thirdly, succinct, action-oriented evaluation reports are more likely to win the attention of decision-makers, and evaluators should be tasked with identifying a limited number of priority actions and presenting their findings in an accessible format.

All of the above has looked specifically at donor take-up of the findings of M&E activities. An even more challenging question is how to ensure that relevant findings are taken up by the national government that is undertaking the SSR process in question. This is closely related to the question of local ownership (see 'Poor local ownership of M&E' below), since local actors are much less likely to respond positively to the findings of evaluations over which they feel they have no ownership or control. Joint reviews of evaluation findings are obviously part of the solution (and do already happen for some projects), but local take-up of findings is unlikely to occur if a sense of ownership has not been established much earlier in the process.

Beyond this, several of the points made above with regard to how reviews and evaluations are written and presented are perhaps even more important when addressing local actors. Reports that are long, complex and peppered with jargon are likely to be even more off-putting to local actors, especially since in the majority of cases English is unlikely to be their first language. Even if a translation is available, many terms related to M&E – and indeed to SSR – may not translate easily into the local language. Jargon may be alienating to civil society actors, particularly those from disadvantaged groups

¹⁷ OECD DAC, *Evaluation Feedback for Effective Learning and Accountability*, (OECD, 2001), p 45.

and minorities. Also, given that lack of capacity is likely to be an issue, decision-makers may have very little opportunity or desire to read a very detailed report with a long list of recommendations, let alone act on its findings. This further underlines the need for a succinct summary and prioritisation of a few key recommendations. In addition, it is often unclear to whom recommendations are directed – where a recommendation is for the national government, this should be made very clear in the evaluation report.

2.2 Sector-specific challenges

There are a number of challenges that are specific to the M&E of the security and justice sectors and SSR programmes, including:

- The complexity of the security and justice sectors
- The lack of agreed international security system models against which to measure outcomes
- Limited numbers of specialist SSR evaluators
- Cultures of secrecy and limited access to information
- Change happens over the long term, but evaluations are usually short-term

The complexity of the security and justice sectors

Security systems are complex in that they have many constituent parts and each sub-sector differs significantly in its purpose, functioning and orientation. For example, evaluating reform of justice institutions, where issues such as due process, responsiveness, legitimacy and access are paramount, may be quite unlike evaluating the reform of intelligence services.

Furthermore, what happens in one part of the security system is likely to be influenced by what happens in other sub-sectors, which can make it difficult not only to attribute change (see ‘Attribution challenges’ above) but also to understand exactly how an institution or programme is functioning. It is highly challenging to understand such complex, intertwined processes and exploring these issues in detail would require evaluators with a vast array of skills and knowledge (see ‘The implications of a multi-sectoral and integrated approach to SSR’ below).

The lack of agreed international security system models against which to measure outcomes

Another challenge is that while long-term goals for SSR are likely to be in line with international norms and values, there is often no international consensus on how to achieve these goals. For example, while all stakeholders may agree that justice reforms should lead to stronger rule of law, greater accountability and equal access to justice for all, there may be various models that lead towards these goals. The OECD DAC Handbook notes that ‘Most English-speaking countries practice common law, and most Francophone and Latin countries practice civil law. This leads to differences in the structure of the justice system and the relationships between judges, prosecutors and the police... In some countries, religious legal systems play an important role.’¹⁸ Evaluators thus risk judging reforms by the standards of models that are familiar to them; even if they acknowledge this risk and try to review the reforms on their own terms, the lack of analogous experiences makes it harder to decide whether any changes that have taken place are appropriate, efficient, sustainable, etc.

¹⁸ Op cit OECD DAC, *Handbook on Security System Reform*, 2007, p 31.

Limited numbers of specialist SSR evaluators

Given its highly specialised nature, SSR evaluations require particular skill sets. True specialists will tend to have deep knowledge of one sector (e.g. policing) or aspect (e.g. criminal investigation) but not others. As donor and programme staff interviewed for this research noted, most SSR issue specialists tend not to have much M&E knowledge. A common solution is to use mixed evaluation teams bringing together people with a range of skills and knowledge.

Another challenge is that there are very few evaluators who come from developing countries where donor-supported SSR is common. Evaluation teams are thus likely to lack valuable perspectives on how SSR programmes are perceived by local partners and beneficiaries.

When security clearance is required (see 'Cultures of secrecy and limited access to information'), this can result in a small pool of 'evaluators', some of whom are in fact not professional evaluators but issue specialists or former government employees. This may make fresh thinking harder to come by. Conflicts of interest may also arise if the same people are involved in programme design, implementation and evaluation.

The long-term solution must be to draw new staff into SSR evaluation, and indeed SSR more generally. The field would benefit from more issue specialists and more governance and development practitioners, who could where necessary be trained in the specifics of the M&E of SSR. However, the highest priority at the moment should be to attract more evaluators and staff from developing countries.

Various short-term measures could also improve the situation. One is to pay greater attention to the background of evaluators so as to avoid potential conflicts of interest. Another is to build more mixed evaluation teams that include staff from developing countries, security and justice specialists and experienced evaluators sourced from professional networks. Evaluation teams should be built around the competencies required rather than the profile of known individuals and favoured consultants. Where tendering is involved, required competencies can be included in Terms of Reference documents.

Cultures of secrecy and limited access to information

Many security system institutions across the world have well-entrenched cultures of secrecy. Even in developed countries, effective oversight and transparency in relation to security policy and practices is new and partial. Of course, complete freedom of information in the security system is not realistic, as some information will always need to be withheld, whether for reasons of national security, due process in law, the proper functioning of public administration or the privacy of individuals. Nonetheless, the amount and type of information that is restricted often goes beyond what is reasonable or legally necessary, and unreformed security institutions often use secrecy to conceal corruption and mismanagement. This tendency to restrict information may well be deeply rooted in governments which are implementing SSR with donor support, and this can make an evaluator's work particularly challenging. Such cultures of secrecy also limit the prospects for broadening participation in M&E processes to include more non-governmental voices.

A separate but related point is that donors who support SSR may also have reasons for limiting access to information. For example, SSR projects dealing with intelligence reforms are likely to involve members of donor country intelligence agencies, who may be uncomfortable with evaluators knowing what they are doing. Similarly, SSR programmes in stabilisation-type environments, where there is significant physical and political insecurity, are also likely to operate in circumstances where access to information is restricted. In such circumstances, evaluators may require security clearance, which may slow down the evaluation process. Alternatively, as in the UK, it may lead

to a more restricted pool of professional evaluators and issue specialists being used for sensitive evaluations.¹⁹

Long-term change, short-term evaluation

Many of the changes sought by SSR (e.g. cultures of accountability or increased trust in police services) are only likely to manifest themselves over the medium to long term, yet donor programmes and their reviews and evaluations continue to be short-term. Evaluations carried out during or immediately after a programme may not be appropriate for assessing whether these changes have taken place, yet rarely are evaluations carried out long after the programme ends to see whether planned long-term changes have been achieved and sustained.

While longer-term donor commitment is probably the most obvious solution to such problems, other steps that may help include: only measuring for those results that can realistically be expected to manifest in the time available and setting expectation levels appropriately, ensuring that national capacities to measure longer-term changes are built up, and working to incorporate questions around SSR processes into the more infrequent strategic-level evaluations that are becoming increasingly common among donors.

2.3 Challenges arising from policy

A third set of challenges is associated with the OECD DAC SSR framework and the SSR policies of donor institutions, including:

- The implications of a multi-sectoral and integrated approach to SSR
- Forging meaningful links to development frameworks
- Ensuring conflict sensitivity

The implications of a multi-sectoral and integrated approach to SSR

The OECD DAC Handbook advocates an integrated approach to security and justice programming, which makes the quality of interaction across government departments a topic for evaluation in itself. This applies at two levels. Firstly, at the level of the national government that is undertaking SSR, evaluations may consider the coherence of SSR programmes that stretch across various national agencies – or where there are several separate SSR-related activities occurring simultaneously, whether they link together strategically. A related question is whether various security system institutions are interacting more coherently and strategically as a result of the SSR programme. A further question might be how well SSR activities connect with the wider developmental context, e.g. through links to Poverty Reduction Strategy Papers (PRSPs). All this suggests that guidance for evaluators could include a standard criterion²⁰ or prompt question which looks at integrated approaches and programme linkages.

Secondly, at the donor level, evaluations may consider the degree of coherence that exists in the support offered by different departments of the same donor, and between different donor organisations. A further layer of complexity may be added by looking at how (in)coherence at the donor level has affected coherence at the partner government level.

In practice, there is usually neither the time nor the resources to look into such matters in any depth, particularly since those commissioning evaluations are usually donors who tend to present their objective in terms of finding out whether their assistance

¹⁹ The research team heard of one case where an intelligence specialist was drafted into an evaluation team, but with separate tasks and reporting procedures. Such approaches can hinder the evaluation team from preparing a joint analysis and may also pose logistical or ethical difficulties during fieldwork.

²⁰ Possible labels for this criterion include 'comprehensiveness', 'linkages', 'strategic approach', and 'co-ordination'.

programme ‘works’ and thus do not expect the evaluation to consider broader issues that at first glance are not directly related to their programme (see ‘M&E serves multiple, sometimes competing purposes’ above). It may also be impossible to build a sufficiently multi-disciplinary team to look at all sectors and issues (e.g. governance, rights, policy as well as individual institutions and their sub-functions) in any one evaluation. Expanded teams are still warranted, but ultimately, evaluators need to recognise that constraints on SSR programmes often means a comprehensive approach will either take considerable time to develop or cannot be easily realised. These include competition among donors, political considerations regarding partner governments and uncertainties over funding or limited entry points. For example, while a 2004 mid-term evaluation of support for the Justice Law and Order Sector (JLOS) in Uganda noted that the military was not considered by the JLOS programme despite its obvious influence on the law and order situation, our case study found that addressing these linkages would have a political dimension that some donors would find uncomfortable to manage.

Forging meaningful links to development frameworks

Given the interest in SSR policy literature around linkages between security and development, the case studies and the desk review examined what can or has been done in practice to link the M&E of SSR to national development strategies and processes. We found few cases where any real attempt had been made to establish a link, other than rhetorically. There could be many reasons for this, ranging from the different motivations and purposes that lie behind donor programmes to the time it takes for new policy prescriptions to filter down into programme design. Our PNG and Uganda case studies highlighted that where national development strategies have high political prominence, linking SSR or at least justice programming to these strategies can be very beneficial. In both cases, country-level M&E frameworks were constructed with an eye to generating evidence of the programmes’ contributions to national development. However, both these and other cases (such as the Sierra Leone Security Sector Reform Programme (SILSEP) in Sierra Leone) showed that SSR’s role in development and poverty reduction is likely to be indirect and therefore difficult to evidence. SSR outcomes should be viewed as *contributing to*, rather than producing, these higher-order impacts (the United Nations Development Programme (UNDP)’s outcome hierarchy is a useful attempt to capture this).

Ensuring conflict sensitivity

Evaluators have a duty to ensure that both the SSR programmes that they are evaluating, and the evaluations themselves, are conflict-sensitive, i.e. that at a minimum they do not do anything that would aggravate conflict dynamics, but instead attempt to address the causes of conflict.

This suggests the need for a prompt question asking evaluators to assess the conflict sensitivity of the programmes they are evaluating. In principle, SSR programmes conducted in conflict-affected environments should incorporate some form of conflict analysis (whether separately or as part of other assessments) and should use this to assess the programme’s impact on broader peace and conflict dynamics. This did not seem to have occurred in any of the case studies we examined nor, so far as we are aware, has it occurred anywhere else. Evaluators assessing SSR programmes could therefore usefully consider the programme’s conflict sensitivity and make recommendations for how this could be improved. If no conflict analysis has been conducted, evaluators should ideally conduct their own conflict analysis against which to measure programme impact. However, it may be difficult to find SSR and/or evaluation experts who also have the appropriate skill set for this task. Furthermore, where no initial conflict assessment has been made, there may be no baseline against which to judge the impact of the programme on conflict dynamics.

Evaluators also have an ethical duty to ensure that their evaluations are conflict-sensitive. Evaluations are often politically sensitive and there is a risk of an evaluation being ‘conflict-insensitive’ and actually exacerbating conflict risks. This requires evaluators to have a good understanding of the context in which they are working and to be aware of how their evaluation process (whom they consult or do not consult, whose perspectives they reflect in their evaluation, etc) and their findings might impact upon conflict dynamics. This does not mean that evaluators are obliged to modify their conclusions in order to avoid saying anything too sensitive – evaluations may legitimately identify problems, challenges or failings that could provoke strong reactions from some stakeholders. In such circumstances, however, evaluators must ensure that they present their findings as accurately and sensitively as possible, and they must highlight any concerns about the conflict sensitivity of their evaluation to the person/organisation that commissioned it. It may be necessary to publish an edited version of the evaluation for a wider audience, restricting access to the most sensitive findings to a smaller circle.

In practice, it is unclear what is done to ensure that evaluators acquire adequate background knowledge of the context in which they conduct the evaluation, particularly in complex and fast-changing fragile and conflict-prone environments. Again, we are not aware of any cases in which evaluators referred to a conflict analysis before or during the evaluation process.

2.4 Challenges generated by donor bureaucracy and politics

The phrase ‘donor politics’ is often used to refer to a range of factors which shape and constrain how donor assistance programmes work. A number of important background issues were identified during the research which shape the thinking and practice of most donor countries and whose effects are visible in several country case studies. These include:

- Incoherence and lack of co-ordination
- The politics of donor support and local ownership of SSR programmes
- Poor local ownership of M&E
- Staffing constraints
- Decentralised decision-making within aid bureaucracies
- Short-term funding cycles
- Staff turnover and workload
- Problems with vague and hidden objectives

With regard to all of these factors, an overarching challenge for evaluators is to identify where the challenges that arise could be avoided or overcome relatively easily by improving practice and policy, and where there are serious structural issues that cannot be easily overcome affecting how donors operate and interact with each other and with partner governments. These need to be accepted, understood and factored into the evaluation.

Incoherence and lack of co-ordination

The coherence and whole-of-government agenda is also relevant for the M&E of SSR. Donor surveys in the UK and the Netherlands demonstrated that different donor departments can have different ideas of what SSR is about, different policy objectives and different mandates. This gap has narrowed through joint working and joint funding arrangements (in the US and the UK), but it remains a challenge that can sometimes play out in evaluations (see for example the Sierra Leone case study). This need not be due to political disagreements. The complexities of linking together multi-budget, multi-strand programmes that stretch across various donor departments mean that some incoherence is almost inevitable; the evaluator may find it no easier

than programme staff to grasp the programme in its entirety.

Coherence, and more broadly the quality of donor conduct in playing a supportive role for SSR despite intra-governmental differences, are important areas for SSR evaluations. A corollary to this is that guidance on the M&E of SSR must be useful and comprehensible for different departments of donor governments, whether it privileges some perspectives over others or attempts somehow to combine them.

The politics of donor support and local ownership of SSR programmes

Over the last few years, it has increasingly been recognised that donor-supported SSR has tended to have poor local ownership.²¹ SSR interventions have often been more in line with the donors' needs and views than those of partner countries. Monitoring and more importantly evaluation should be seen in this context. Evaluators should be alerted to the issues surrounding local ownership and should be tasked with assessing donor conduct, respect for local priorities and pursuit of local ownership. This might be indicated by the extent and quality of consultations held on different aspects of programming and the extent to which national government institutions actually implement SSR-related decisions and policies.

Poor local ownership of M&E

However, the question of local ownership is not only a question that evaluators must ask about the SSR programme – it is also a question for the M&E process itself. Who decides which M&E activities to undertake and when? Who sets the scope, criteria and terms of reference for evaluations? Who conducts M&E and who is the subject of it? Who participates in M&E and who gets to access findings or use them for learning or decision-making?

Badly executed, M&E can also be seen as a process belonging solely to the donor. Partner countries need support and encouragement to build up their M&E procedures and ensure that the partner government itself has some control over the process and is able to elaborate its views on the effectiveness of the support it receives from donors. However, this does not usually occur, especially since donor programme staff may themselves lack the experience, training and resources to prioritise and carry out M&E effectively. This can mean that little monitoring is undertaken and little partner government involvement is sought in any monitoring or reporting that does take place, at least until the donor's own procedures require it to undertake an annual, mid-term or end-of-project review and evaluation. At this point evaluators – often foreign nationals – can overwhelm partner governments with requests for information, without much attention to ownership or capacity. This does little to encourage a culture whereby monitoring, reviewing and evaluating are routine and locally-driven within partner institutions.

Furthermore, few SSR programmes are designed with an understanding of the partner country's M&E capacities and management information systems, in the security system and beyond, and even fewer SSR programmes make building the capacity of partner government M&E systems a part of their mandate. A notable exception to this rule is the recently launched Security Sector Accountability and Police Programme in the Democratic Republic of the Congo (DRC), funded by DFID. This five-year programme includes a defined M&E component which looks not only at monitoring and evaluating the activities of the donor, the national government and the consultancy teams engaged in delivering the project, but also aims to build the M&E capacity of the Congolese government.²² Support to build the M&E capacity of partner governments

²¹ See for example Donais T (ed), *Local Ownership and Security Sector Reform*, (DCAF, 2008).

²² Some information about this programme can be found in the DFID Country Plan for DRC 2008–2010, available at: [www.dfid.gov.uk/Documents/publications/DRC-countryplan08-10\[1\].pdf](http://www.dfid.gov.uk/Documents/publications/DRC-countryplan08-10[1].pdf) and on DFID's online project database (projects.dfid.gov.uk – project component code 113961-105).

should be a core component of most SSR assistance programmes where the partner country lacks M&E capacity, particularly for programmes that are relatively large in scale.

Staffing constraints

In many cases, the staffing levels of donor aid institutions are fixed or decreasing while aid budgets are rising (e.g. the US, the Netherlands, and the UK). One implication of this is that donors will increasingly turn to external contractors to deliver, and perhaps also to design, monitor and evaluate work. Particularly in the US, there are concerns that contractors have been given too much leeway when designing indicators, commissioning evaluations and compiling reports to donors. These concerns need to be considered more seriously, given the potential conflicts of interest that can arise in such situations and also given the low level of M&E experience of many contractors. A related concern is that although donors may claim to adhere to the principle of independent evaluation, many evaluations are effectively internal, carried out by project staff and other key stakeholders. This is particularly the case with respect to lower-level project evaluations and reviews, as the US, UK (DFID) and the Netherlands case studies all show.

Decentralised decision-making within aid bureaucracies

The decentralisation of most aid donor bureaucracies (UNDP being a partial exception) presents challenges for M&E, since most donor offices do not have much M&E expertise. At the field-office level, monitoring and reporting can easily be seen as a burden, as several of our case studies demonstrated. Tolerance or appetite for tailored guidance and tools to support the M&E of SSR is mixed. On the one hand, many staff already think reporting (which is often mistakenly equated with monitoring) is burdensome, while evaluation is often perceived as something that might be required but is not all that useful, and at worst intrusive (as reported in the US case study). On the other hand, field staff are generally not issue specialists and may change jobs every couple of years, so targeted, concise and easy-to-use tools for monitoring and reporting are likely to be welcome. Since evaluations should be external and independent, less guidance is required for field-level staff in this regard (though they may still benefit from information about the purposes of evaluation and how field-level staff can best co-operate with evaluators). Instead, detailed guidance on SSR evaluation should be targeted more towards senior project managers and higher levels of management and decision-making.

Short-term funding cycles

Even though SSR often has medium and long-term goals, funding cycles remain short-term. This means that in order to attract funding, those designing SSR projects are often under pressure, consciously or unconsciously, to be overambitious about what can be achieved in a period of one to three years. Evaluators interviewed for this research reported that goals are indeed often vague and overly ambitious, and that programme designers tend to over-estimate what it is possible to achieve. Although this problem is not unique to SSR, it has great significance given the long timeframe for some of the outcomes sought by SSR and the difficult and uncertain environments in which SSR programmes usually operate.

Staff turnover and workload

In all of our case studies there were problems with monitoring linked to high staff turnover, a lack of guidance and heavy workloads. All of these factors appear to dissuade or prevent field staff from routinely collecting and analysing information on the progress of their SSR projects. Standard donor requirements for project documents

and logframes did not always appear to be sufficient to help staff identify the most salient information to collect and analyse.

Problems with vague and hidden objectives

Several high-level reviews of donor SSR projects have shown that their design can be wanting in various ways.²³ Programme goals are sometimes ambiguous. This may reflect a lack of knowledge of the context, the sector or the issues, but it may also be a way for a programme to find its feet over time without being pinned down to specific commitments. Alternatively, it may reflect the fact that a programme is being carried out for deliberately unstated political objectives, ranging from the political imperative to run a programme as a demonstration of support for partner governments through to overlaps with donor counter-terrorism work. Programme staff may not admit these objectives (they may not even fully be aware of them), and evaluators may struggle to pick up on them as they tend not to feature in project documents or monitoring reports.

Evaluators need to probe these sorts of issues to the extent that they are able (though such questioning is not always welcome and is often not part of their terms of reference). This will involve greater reliance on verbal exchanges with past and current programme staff. Evaluators may need to hold group discussions to avoid their analyses becoming hostage to the widely differing views of programme purpose that may be found among stakeholders.

Even if evaluators do manage to understand these issues, opportunities to share their findings publicly might be limited, since donors who obscure their reasons for supporting SSR programmes at the design stage may still wish to do so at the evaluation stage. This may be particularly true of sensitive programmes where evaluators require security clearance (see 'Cultures of secrecy and limited access to information' above). At the very least, however, it should be possible to share such findings with those who commissioned the research and with other relevant senior decision-makers; if donors are not prepared to be clear about the real objectives of a programme even in private, the evaluation is unlikely to provide much insight into how effectively these objectives are being met, and the evaluation risks becoming little more than a bureaucratic exercise.

2.5 Challenges arising from poor programme design and management

Some of the challenges surrounding the M&E of SSR programmes arise because of flaws in the initial design of the programme or the way in which it has been managed and resourced throughout the project cycle. Key issues include:

- Weak programme design
- Poor preparation for M&E
- M&E is under-resourced
- Local populations and marginalised groups have little participation in M&E
- Poor use of external data sources

Weak programme design

Many of the project documents examined for the desk review had vague and overly ambitious goals and purpose statements. In many cases, project designers did not

²³ For example, a review of the UK Government's Africa Conflict Prevention Pool funding in 2006 found that 'record-keeping in departments is lacking and when a review is conducted lessons are not passed on consistently even within country strategies and even more rarely between country strategies.' Ball N, Biesheuvel P, Hamilton-Baillie T and Olonisakin F, *Internal Review of ACP and DFID Security and Justice Sector Reform Programming in Africa: Inception Report*, (DFID, 2006), p 6. See also: Olonisakin F with Ikpe E and Badong P, *The Future of Security and Justice for the Poor: A 'Blue-Sky' Think Piece*, (King's College London, 2009); and Ismael O and Hendrickson D, *What is the case for a security and justice focus in development assistance programming? An assessment of existing literature and evidence*, (GFN-SSR, 2009).

appear to have articulated a coherent vision of what progress would look like or how it could be measured; for example, several case studies highlighted problems in setting specific, measurable indicators and/or accompanying targets. In the Sierra Leone case study, the challenge was deeper, as for a long time there was no clearly agreed strategy for reforms. The Sierra Leone case study also shows how important it is to agree strategy and define goals and objectives in a way that takes account of the different perspectives of the different donor ministries involved in SSR.

Poor preparation for M&E

As in the mainstream development sector, a big issue for the M&E of SSR is that it is often not a priority for project designers and thus M&E procedures are poorly developed from the start. Some of the projects reviewed for this research had made little or no preparation for M&E, while some had not dealt with the issue in project documents at all (e.g. the Albania case study).

M&E issues need to be treated seriously early on, with time and effort dedicated to agreeing what success looks like and then planning how, when and by whom it will be measured (e.g. through agreeing indicators or at least a monitoring plan) and ensuring adequate resources for M&E are built into budgets. SSR needs assessments should also consider the capacity of national institutions in relation to information collection and M&E. Donor project documents should in turn promote the task of supporting national systems, if necessary including indicators relating to how such systems will be strengthened over time. Subsequent M&E by donors should examine whether these activities have helped to strengthen the capacity of national systems for data collection.

Under-resourcing

Under-resourcing for M&E is often a problem. The resources committed to the M&E of SSR activities need to be proportional to the scope, duration and complexity of the work, so it is difficult to generalise about what resources are necessary. A common rule of thumb in other sectors is to allocate 5–10 percent of the project budget to M&E. In most of the case studies, considerably less than this figure had been spent on M&E. In fact, the M&E of some complex SSR programmes may actually require *more* than 10 percent of the budget to be effective; donor interviewees in Uganda suggested that sector-wide M&E approaches need additional resources, to the order of 10–15 percent of the overall budget.

Participation issues

The reviews and evaluations examined in our case studies had relatively little input from key groups of beneficiaries, such as the local population as a whole, but also marginalised groups, rural populations and so on. Most evaluations have focused predominantly on urban areas and changes at the national level (PNG and Uganda provide partial exceptions) and have mostly interviewed project staff, donor staff, other key government officials and a few other experts. There is thus an urgent need to broaden the range of voices heard in most evaluations so that they more accurately represent the views of the populations that are meant to be the ultimate beneficiaries of SSR programmes. Ideally, this range of voices should be broad enough to allow evaluators to disaggregate such input according to categories such as gender, age, income, etc, since different groups are likely to be affected by SSR in different ways.

However, the Uganda case study suggested that in order to strengthen civil society/public engagement in SSR, donors may need to provide vigorous support to civil society groups due to their disempowerment and poor understanding of security issues. The Uganda case also showed that surveys that seek input from a broad range of voices need to be refined over time as understanding grows among those working on SSR as to what information they really need and how best to elicit it.

Poor use of external data sources

In several of the case studies (e.g. the UK, the US and Albania surveys), there was a marked reluctance to seek information from external sources. Most case studies showed undue reliance on information sourced from security system institutions and partner government officials (as in Sierra Leone and PNG). In the two case studies that looked at sector-wide justice programmes (Uganda and PNG), considerable effort had gone into running consultations, focus groups and opinion polls over the years to track public perceptions – one good example of using information from external actors; such methods were also used in Albania and Bosnia but to a lesser extent and they were not clearly linked to decision-making. Sometimes, as in Albania, donors were also failing to make good use of national data collection systems.

2.6 Contextual challenges²⁴

M&E must take into account the societal and institutional context in which SSR is taking place. This is particularly true in what might be termed ‘fragile’ or ‘conflict-prone’ environments, where the majority of SSR programmes take place. Some of the most important contextual challenges are:

- The M&E framework must reflect the local context
- Difficulties in establishing baseline data
- Low national capacity for data collection

The M&E framework must reflect the local context

Academic research shows that security systems in developing, fragile and conflict-affected states operate differently to those in mature democracies.²⁵ Capacities are generally lower and responsiveness to public demand is more limited. A shared national view of safety, security and justice as public goods may not exist, and accountability may be low due to a lack of political space for critical voices and/or civil society participation. Moreover, in many cases there will be no sole locus of decision-making within institutions or across the system, but rather a web of competing power centres. In such circumstances, decision-making is more likely to take place through informal networks. There may also be a plurality of legal systems and/or a large role for non-state security and justice mechanisms.

All of this presents challenges for those designing, managing, monitoring and evaluating SSR activities. M&E can be significantly hampered without good analysis, particularly political analysis. Where there are multiple centres of power and the style of governance is primarily informal, this must be mirrored in M&E techniques. Using local analysts may help, as can informal data collection methods. In contexts where change often occurs through informal channels, the need to build trust and maintain relationships with interlocutors is even greater. It is also very important to triangulate information from different stakeholders in order to ensure the accuracy and balance of any data collected.

Difficulties in establishing baseline data

The case studies showed that it is rare for programme designers and managers to undertake set-piece baseline assessments that draw on primary research and external information sources at the start of the project (the PNG case was exceptional in that a

²⁴ There is a useful discussion of conditions in fragile states and the implications for M&E in Sartorius R and Carver C, *Monitoring, Evaluation and Learning for Fragile States and Peacebuilding Programs: Practical Tools for Improving Program Performance and Results*, (Social Impact, 2006).

²⁵ Conflict, Security and Development Group, *State Responsiveness to Public Security Needs: The Politics of Security Decision-Making – A Comparative Study of Nigeria, Sri Lanka and Uganda*, SSR Policy Briefing for HMG, (King's College London, 2008).

study was built into the design phase). Such gaps are often regretted by donor field staff later on in the project when they are looking to monitor and evaluate progress.

Project managers find it particularly hard to set baselines and targets in fragile and conflict-prone environments given heavy time pressure and the difficulties of gathering relevant information. Even when baselines have not been set from the beginning, regardless of whether this is because of poor design or contextual challenges, it is important to establish baseline data as soon as possible. This may include 'reconstructing' a baseline from secondary sources or stakeholder interviews, if this can be done in a plausible fashion.

During the research, an important distinction emerged between two different types of 'baseline' assessments. The first is the 'set-piece' study noted above, which involves primary research to gather a range of qualitative and quantitative information against which progress can be tracked. The other is simpler, what might be termed an 'analytical' baseline such as that included in a country strategy, project document or logframe. While set-piece studies are obviously more detailed and thus more useful in terms of getting to grips with the local context and setting indicators and objectives that are locally appropriate, these in turn usually stem from a good analytical baseline. Too often, it appears that insufficient effort has gone into creating this analytical baseline during project design and start-up; moreover, subsequent evaluations seldom refer back to this baseline.²⁶

Low national capacity for data collection

In most case studies, M&E had proved challenging because national systems and capacities for data collection, handling, storage and analysis within national security and justice institutions were weak (Uganda), absent (Sierra Leone) or difficult to adapt to (e.g. PNG with its oral culture). Albania presented a partial counter-example, with national data systems being under-used. (See also 'Poor local ownership of M&E' and 'Poor preparation for M&E' above.)

The obvious lesson is that national capacity for data collection must be taken into account and made a feature of programming. This will likely involve engaging a range of national institutions, both within the security and justice sectors (e.g. police research departments) and more generally (e.g. national statistics offices and national planning and development agencies), and also non-governmental bodies such as universities, non-governmental organisations (NGOs) and the media. Security system institutions may need support to improve routine collection of administrative data, much of which will be activity- and output-focused (e.g. recorded crime, personnel attendance rates, numbers of arrests and detentions, etc). National bodies such as planning ministries and audit offices may over time provide more qualitative, outcome-focused data. Civil society can provide an alternative source to official data, providing its own information and analysis or demanding data from institutions.

Ultimately, the goal should also be to strengthen the capacity of oversight actors, ranging from the executive to parliaments, the judiciary and civil society, all of whom can contribute their own evaluations and reports on the security system and on particular programmes. The scope and purpose of such reports may be framed to suit national rather than donor purposes and will probably be subject to problems arising from low capacity. Nonetheless, reports by these institutions will often carry greater political weight than the average donor-backed evaluation, especially if they are independent from both the programmes they are evaluating and the national security and justice system more generally.

In highly insecure environments, monitoring and data collection is likely to be even more problematic. Much data may simply be unavailable, and rapid changes to the

²⁶ DFID Output to Purpose reviews are a notable exception, since they are designed to measure progress in exactly this fashion.

context may make earlier information irrelevant. Programme staff may be unable to gather data themselves because of security concerns and logistical constraints that make it impossible to undertake research in conflict-affected areas. In such cases, programme managers and evaluators need to be as flexible and innovative as they can about collecting useful information and adapting it sensitively to the local context. This might include involving beneficiaries more directly in the M&E process by encouraging them to gather the information required; this also has other benefits in terms of building local capacity to hold the authorities to account.

3

Content and process: What should we be measuring and how?

ONCE THE NEED FOR M&E IS ACKNOWLEDGED, the fundamental questions are what exactly should be measured, and *how*? This section of the report focuses on these questions, exploring four aspects in detail:

- What are the appropriate foci in terms of level and unit of analysis for M&E of SSR?
- What are the main dimensions of change that SSR is supposed to produce (according to current policy thinking)?
- What are the key questions (and criteria) that those undertaking M&E need to be asking of SSR programmes?
- Who should drive, who should own and who should be involved in the M&E process?

3.1 Where should we focus our M&E efforts?

In the course of the research, important questions arose about the proper unit of analysis of M&E (sometimes referred to in M&E literature as ‘the evaluand’) and the most appropriate level at which to conduct it. In fact, we identified several distinct levels or foci:

Programme/institutional level

- Donor programming efforts, or aspects of them such as the performance of contractors or the conduct of the donor (e.g. in line with aid effectiveness standards such as ownership alignment, harmonisation, co-ordination or in line with stated SSR policy)
- National security and justice institutions (their performance being subject only in part to donor efforts where any support is provided)

Sector/System level

- Donor support to security and justice sectors/systems (e.g. through sector-wide approaches (SWAps) and sector frameworks, budget support and similar)
- National SSR ‘processes’ (the broader state and trajectory of security system performance and governance in a country, region or territory)

- The linkage of either of the above to national development or peacebuilding frameworks (e.g. PRSPs, transitional results matrices and associated surveys and dialogue processes)

Country/context level

- This may cover political and security issues, but might also touch on socio-economic aspects – this can be part of ongoing context, conflict or risk monitoring

Policy level

- General compliance of donors or national governments with relevant policy standards on SSR across a range of programmes and contexts
- Relevance, appropriateness and impact of such policies

Each level/focus is a legitimate and important focus for M&E. However, each has different political significance and its own set of technicalities. The **programme/institution level** is probably the instinctive choice for those running donor-funded SSR programmes, as they wish to know about the day-to-day performance of national security and justice institutions and possibly their contribution to that performance. In the early stages of SSR processes, this level of M&E is likely to make the most sense to different stakeholders. It may also be at this level that the donor's contribution to any change in performance is easiest to identify. Most reviews and evaluations examined during the research were project/programme-focused: their unit of analysis was the project at hand, as set out in project documents and logframes; and more often than not they measured progress in carrying out planned activities or delivering outputs (e.g. a new law or refurbishing a training academy).

Sector/system level M&E carries the potential to focus minds on development-security linkages, encourages thinking about cross-sectoral issues, and offers significant opportunities for joint monitoring among different donors or by donors and national government. Our case studies in Uganda and PNG showed that by developing agreed M&E frameworks over time at the sector level and by connecting reporting with budgeting processes, donors and national governments could increase interest in M&E, institutionalise data collection and analysis, and contribute to a franker exchange on priorities and progress. M&E at this level is more likely to be impact- as well as outcome-focused, concentrating more fully on the functioning of national security systems and the institutions which donors have supported and the consequences for the well-being of beneficiary populations (e.g. ability to access markets or travel freely on roads, objective and subjective measures of security and access to justice). It then works backwards to assess what contribution, if any, donors had made to observed changes. The M&E system examined in our Uganda case study was a partial example of this: while there are little in the way of formal mechanisms to review donor performance, there is a strong informal peer review approach between the Justice, Law and Order Sector and the donors.

Tracking progress at the **country/context level** allows data to be gathered on often neglected questions around drivers of change, political will and conflict dynamics. However, at this level it is much harder to attribute impact to individual programmes. Nevertheless, all SSR programmes should maintain some political and context monitoring, particularly in conflict-prone contexts.

Lastly, at the **policy level**, evaluation of donor performance or policy compliance is scarce. Two case study interviewees noted that such evaluation is valid, but very rare. They explained that SSR reviews tend to be initiated by donors rather than their partner governments, and focus on indicators and logframes more obviously owned by donors than by partner governments and security institutions. This demonstrates the links between programme M&E and programme ownership – agreement on what it is important to measure and what success looks like, and the degree of government participation in M&E processes are important tests for national ownership of SSR (and the M&E of SSR). However, in the two case studies where partner government owner-

ship of M&E could be rated as relatively high (PNG and Uganda), systems tended to be more complex, with added potential for confusion as to the appropriate roles and responsibilities of different actors (in both cases, the list included donors, national governments and their security and justice institutions, contractors, civil society, and national planning and monitoring institutions). In situations where direct budget support underpins reforms, the distinctions become even more blurred.²⁷

3.2 Defining the 'dimensions' of change

M&E specialists will sometimes talk of the 'domains' or 'dimensions' of change along which programmes are designed. Our research suggests that those involved in SSR work would find it helpful to have greater clarity and advice on the different dimensions within which SSR programmes can seek change. This was indicated both by explicit requests for more information from interviewees and from an analysis of numerous project documents, logframes and reviews which point in many cases to a lack of rigorous thought in programme design, or a failure by evaluators to sufficiently interrogate design and the theory of change behind it.

In our view, a valid and useful framework illustrating the main dimensions of change for SSR programming should reflect standard programme designs, but must go beyond them where they are known to be lacking, and must also reflect policy commitments where relevant. We found OECD DAC SSR policy and guidance to be fairly exhaustive (perhaps because it follows several years of experimentation during programming and was compiled with extensive 'practitioner' input). The most appropriate reference points are therefore:

- Key elements of the 'Policy Statement' in *Security System Reform and Governance: a DAC Reference Document*, principally the five 'basic working principles'.²⁸
- 'The Ministerial Statement on Key Policy and Operational Commitments from the Implementation Framework for Security System Reform (IF-SSR)', signed on behalf of OECD DAC Ministers and Heads of Agency in Paris on 4 April 2007.²⁹

These are summarised in the matrix below, with additional commentary on the possible implications for a standard M&E framework.

However, the OECD DAC guidance should not be taken as the final word on what is worth measuring when it comes to SSR. The OECD's approach has its detractors, some of whom find it overly ambitious and out of touch with developing, fragile and conflict-affected country realities. Others point to gaps in the framework. Time – and comprehensive impact and policy evaluation – will tell whether the OECD DAC approach to SSR is feasible and largely adequate. In the meantime, it is necessary to add questions such as gender, financial management and poverty reduction, on which DAC guidance is considered to be weak,³⁰ to the list of items below.

²⁷ In fact, in several of the case studies it was not entirely clear whether a review/evaluation was focused on the performance of the partner country and its institutions or the donor and its contribution to their performance (e.g. SILSEP III evaluation in Sierra Leone). One useful approach appears to be to give evaluators the scope, time and resources to consider these different aspects and draw them out in their reports, such as in the mid-term review of the Uganda JLOS.

²⁸ OECD DAC, *Security System Reform and Governance: a DAC Reference Document*, (OECD, 2005), pp 11–14.

²⁹ OECD DAC 'The Ministerial Statement on Key Policy and Operational Commitments from the Implementation Framework for Security System Reform (IF-SSR)', 2007, pp 10–12. Note: Some of these principles are then reprised in a summary of four 'overarching objectives' that international actors should provide support for: i) Establishment of effective governance, oversight and accountability in the security system; ii) Improved delivery of security and justice services; iii) Development of local leadership and ownership of the reform process; iv) Sustainability of justice and security service delivery.

³⁰ For example, in a recent think-piece for DFID, 'Funmi Olonisakin argues that despite the agency having at times propounded a 'pro-poor' approach to security and justice work, including SSR, its programming has generally failed to identify or target poor, vulnerable and marginalised groups. Similarly, no specific measures to track outcomes for these groups, however defined, have been developed. Op cit Olonisakin et al, *The Future of Security and Justice for the Poor*, 2009.

GUIDING PRINCIPLE

IMPLICATIONS FOR M&E FRAMEWORK

SOURCE: OECD DAC Ministerial Statement (2007)

<p>1 Donors should engage in SSR with three major overarching objectives: i) the improvement of basic security and justice service delivery; ii) the establishment of an effective governance, oversight and accountability system; and iii) the development of local leadership and ownership of a reform process to review the capacity and technical needs of the security system.</p>	<ul style="list-style-type: none"> ■ Identified dimensions of change should include improved service delivery, governance, oversight and accountability and ownership of reform processes
<p>2 Technical inputs to SSR should be delivered and co-ordinated with a clear understanding of the political nature of SSR and institutional opportunities and constraints.</p>	<ul style="list-style-type: none"> ■ Prompt question for evaluators as to whether programmes are founded on the basis of both assessments of institutional capacities and political/drivers of change analysis which is updated through context monitoring
<p>3 The political terrain needs to be prepared in partner countries and early investments made in appropriate analysis.</p>	<ul style="list-style-type: none"> ■ Political terrain as a valid dimension of change ■ Prompt question for evaluators around handling of political dialogue and analysis
<p>4 Assessment tools should inform the design of realistic, focused programmes, which can make significant contributions to supporting partner countries in addressing the security and justice needs of all citizens.</p>	<ul style="list-style-type: none"> ■ 'Impact' to be defined, at least in part, in terms of changes to the lives of beneficiary populations, with changes disaggregated by social difference (e.g. gender, age, income, etc) ■ Enhanced security and access to justice as valid dimensions of change
<p>5 Programmes need to be designed to help identify local drivers of reform and be flexible in supporting local ownership as it emerges.</p>	<ul style="list-style-type: none"> ■ Prompt question for evaluators as to whether programmes are founded on the basis of such political/drivers of change analysis, updated through context monitoring
<p>6 Donors must support partner countries to lead SSR processes as the starting point for sustainable reforms. But because ownership and leadership are never monolithic and not always easy to determine, opportunities to foster multi-stakeholder coalitions for change should be prioritised.</p>	<ul style="list-style-type: none"> ■ Prompt question to investigate support provided by donors to national leadership during process-focused evaluations ■ Ownership to be explicitly referenced as broad-based, involving multiple stakeholders ■ Evaluators should review what has been done to build the M&E capacity of the partner country
<p>7 Donors must work with partners to ensure that initiatives to support the delivery of security and justice are conflict-sensitive and sustainable financially, institutionally and culturally. Sustainability is a key issue in the design and delivery of support to security and justice service delivery.</p>	<ul style="list-style-type: none"> ■ Prompt question for evaluators on conflict-sensitivity ■ Conflict/context monitoring required in conflict-prone environments ■ Standard development evaluation criterion of relevance (or humanitarian aid criterion of appropriateness) valid for SSR evaluation ■ Standard development evaluation criterion of sustainability valid for SSR evaluation, with finance, institutional make-up and culture all areas of legitimate enquiry
<p>8 SSR programmes need to take a multi-layered or multi-stakeholder approach. This helps target donor assistance to those providers, state and non-state actors simultaneously, at the multiple points at which actual day-to-day service delivery occurs. A multi-layered strategy helps respond to the short-term needs of enhanced security and justice service delivery, while also building the medium-term needs of state capacity.</p>	<ul style="list-style-type: none"> ■ See points 1 above and 11 below
<p>9 The international community needs to move from ad hoc, often short-term, projects to more strategic engagement.</p>	<ul style="list-style-type: none"> ■ Security sector/system level a valid and important focus for M&E ■ Scope of SSR programmes, their strategic 'reach' and degree of linkages to other programmes and frameworks to be referenced under relevant evaluation criteria and possibly included as indicators for 'strategic engagement' ■ Possible prompt questions for evaluators on these topics
<p>10 Donors should strive to develop specific whole-of-government capacity to support SSR</p>	<ul style="list-style-type: none"> ■ Existence and functioning of donor mechanisms for whole-of-government support to SSR a valid concern for process-oriented M&E
<p>11 SSR objectives need to focus on the ultimate outcomes of basic security and justice services. Evidence suggesting that in sub-Saharan Africa at least 80% of justice services are delivered by non-state providers should guide donors to take a balanced approach to supporting state and non-state security and justice service provision.</p>	<ul style="list-style-type: none"> ■ The 'non-state' to be referenced as a sub-sector in any standard M&E framework ■ Possible prompt question for evaluations around whether non-state dimensions have been factored into assessment and design

GUIDING PRINCIPLE *continued*

12 The international community should use appropriate instruments and approaches for different contexts, and should build support across the justice and security system to ensure a more strategic approach to SSR.

13 The international community needs to align support to the dominant incentive frameworks and drivers for change.

14 SSR should be viewed as an integral part of the planning process for immediate post-conflict situations and peace support operations.

15 The Implementation Framework for Security System Reform (IF-SSR) should be used to help place sub-sector reforms in the context of system-wide needs.

16 The key principles agreed in the 2004 DAC SSR Guidelines need to be translated into evaluation indicators. A focus on programme outcomes requires an evaluation of strategic objectives, impact and not only project outputs.

IMPLICATIONS FOR M&E FRAMEWORK *continued*

- Criteria of relevance/appropriateness valid (see point 7 above)
- See also point 9 above

- See points 2–5 above

- Questions around linkages between SSR programmes, peace support operations and peacebuilding frameworks to be given as prompts under relevant evaluation criteria

- Sub-sector projects should be assessed in part for their overall contribution to meeting system-wide needs, including the real security challenges facing a country (relevance criterion)
- Prompt questions as to the degree to which sub-sectoral projects are designed to link across institutions and levels of the security and justice system and with other programmes and development strategies (linkage or co-ordination criterion)

- M&E framework must ensure due focus on outcomes and impacts, in turn implying a focus on strategic (sector/system) level results

OECD DAC Handbook Definition of SSR

17 The overall objective of international support to security system reform processes is to increase the ability of partner countries to meet the range of security and justice challenges they face, 'in a manner consistent with democratic norms, and sound principles of governance and the rule of law'...³¹

- Norms of good and democratic governance, human rights and rule of law to be referenced in M&E framework and/or evaluation criteria

Basic working principles for donor support to SSR processes (2005)

18 People-centred...

- See point 4 above

...locally owned...

- See point 6 above

...based on democratic norms and human rights principles and the rule of law...

- See point 17 above

...seeking to provide freedom from fear and measurable reductions in armed violence and crime.

- See point 4 above

19 Seen as a framework to structure thinking about how to address diverse security challenges facing states and their populations, through more integrated development and security policies and through greater civilian involvement and oversight...

- See point 15 above (applied this time to all projects, not just sub-sectoral ones)
- Enhanced civilian involvement and oversight of security systems as a valid dimension of change

20 Founded on activities with multi-sectoral strategies...

- See points 9 and 15 above

...based upon a broad assessment of the range of security and justice needs of the people and the state.

- See points 9 and 15 above
- In addition, process-oriented M&E should consider whether baseline assessments have been carried out on the security and justice needs of the people and the state

21 Developed adhering to basic governance principles such as transparency and accountability

- Process-oriented M&E to consider whether there is 'due process' in SSR programming

22 Implemented through clear processes and policies that aim to enhance the institutional and human capacity needed for security policy to function effectively and for justice to be delivered equitably.

- Institutional and human capacity as valid change dimensions
- Security policy as a valid change dimension
- Equitable access to justice a valid change dimension

³¹ Op cit OECD DAC, Handbook on Security System Reform, 2007, p 22.

Combining this guidance with logic models used by evaluators, it is possible to capture the key SSR dimensions of change in a single overarching framework that separates out different logical levels of results (impact, outcome, activity/process) and gives examples of outcomes and indicators for the key sub-sectors and institutions in which most SSR work takes place. In addition, outcomes and indicators that illustrate sector- or system-wide results need to be defined, and guidance must be given on cross-cutting themes such as gender, human rights and poverty reduction. This may require careful balancing of indicators or disaggregation of data sets.³² This would not supplant other frameworks listed in this report, but rather it would act as an *aide-memoire* or prompt for SSR programme designers, managers and evaluators, identifying key change trajectories and suggesting areas that might otherwise be overlooked. Its various sub-sections would direct the reader to any other useful guidance, including indicator sets. The matrix outline below, developed by Saferworld specifically in relation to policing, gives a basic illustration of how this might work:

IMPACT LEVEL

INTENDED RESULTS	SUCCESS INDICATORS
<p>Dimension 1 (direct): Physical well-being</p> <p>Includes factors such as freedom from fear, intentional harm and perceptions of physical security SSR example: Enhanced physical security for local people</p>	<ul style="list-style-type: none"> ■ Reduction in the number of people reporting they have been a victim of crime ■ Reduction in number and degree of severity of human rights violations, including those by security forces ■ Reduction in violent crime statistics
<p>Dimension 2 (direct): Empowerment/social changes</p> <p>Includes factors such as greater awareness and understanding of problems and solutions, participation in decision-making and public institutions, access to resources SSR example: People successfully claim rights to justice and security</p>	<ul style="list-style-type: none"> ■ Stronger civil society networks active on relevant issues ■ Indicators of participation, etc
Dimension 3: etc	■ etc

OUTCOME LEVEL

INTENDED RESULTS (for police sub-sector)	SUCCESS INDICATORS
<p>Dimension 1: Accountability and oversight</p> <p>Example: Police services established on the basis of a clear constitutional and legal framework; legal framework provides an effective basis for regulation of policing and establishes lines of accountability</p>	<ul style="list-style-type: none"> ■ Existence and clarity of key legal texts, e.g. Constitution, Police Act and supporting regulations to basic standards, adherence to them, provision for accountability measures therein
<p>Dimension 2: Capacity</p> <p>Example: Adequate ratio of police personnel to population</p>	<ul style="list-style-type: none"> ■ Increased number of police personnel relative to population ■ Appropriate redeployment of existing personnel
Dimension 3: etc	■ etc

PROCESS/ACTIVITY LEVEL

INTENDED RESULTS (for police sub-sector)	SUCCESS INDICATORS
Dimension 1: New and improved Police Act drafted and enacted.	■ Improved Police Act drafted, reviewed and passed by parliament
Dimension 2: Recruitment and redeployment of police personnel in pursuit of appropriate police-public staff ratio	<ul style="list-style-type: none"> ■ Improved recruitment procedures instituted ■ Greater intake of appropriate candidates
Dimension 3: etc	■ etc

³² The Vera Institute's 'Global Guide' (see below) looks in some detail at perverse indicators in the justice sector and the need to combine indicators into 'baskets' so as to safeguard people's legal rights.

3.3 Exposing theories of change

Interrogating programme logic and underlying theories of change is a major ‘content’ issue for the M&E of SSR. Given the poor evidence base for much policy and guidance and the apparent flaws in the design of many project documents examined for this research, encouraging evaluators to expose the programme’s intervention logic could offer important lessons. Many evaluations do not interrogate programme logic because of their limited scope. Donor reviews – including all those examined in our case studies – tend to concentrate on tracking progress towards programme goals and objectives (in DAC terminology, effectiveness). This is a valid and important focus, particularly during a project’s lifetime, but it can leave the most important questions unanswered, such as whether the programme design is appropriate to the context or whether the intended target groups have benefited in terms of enhanced security. For example, in the case of the SILSEP reviews which we studied, while due attention was paid to whether the project was successfully contributing to improved oversight of security institutions at the executive level as envisaged in programme design, successive reviews did not question the project’s design or question whether it should also have addressed Sierra Leone’s (extremely weak) external oversight bodies.

In fact, there are at least two distinct aspects of successfully reviewing a programme’s intervention logic. Firstly, it is necessary to determine the intervention logic and the accompanying theories of change. The Swedish International Development Agency (SIDA) Evaluation Manual argues that ‘without a preliminary analysis of goals and objectives, indicators and assumptions, etc, it is difficult to pose the right questions. A preliminary analysis of the intervention logic will also be useful to the evaluators. Otherwise, they may waste plenty of time in figuring out what they actually should be evaluating.’³³ However, this may in itself be a challenge. Often, the intervention logic and the accompanying theories of change are not explicitly stated in programme documents. Evaluators usually need to talk at length with programme staff (and possibly their predecessors) in order to unravel the thinking behind the project. Such conversations are time well spent, especially in a field where hidden objectives and indistinct programme design are common. Advice on how to identify the intervention logic, together with information on theories of change found in related fields, may prove useful in developing tailored guidance on SSR. Work on underlying theories of change in conflict prevention and peacebuilding could serve as a useful starting point.³⁴

Once the intervention logic is understood, the second stage is for evaluators to review whether this logic (and its underpinning theory of change) was indeed appropriate and whether it was likely to be effective. This is of crucial importance, since many unsuccessful programmes can be traced back to initial design flaws based on incorrect assumptions about how to effect meaningful change. Many of the prompt questions in the table below (‘Criteria and prompt questions for the evaluation of SSR programmes’) should be seen in this light.

3.4 Tailoring evaluation criteria

On the question of evaluation criteria, we found a fair degree of support for the five standard OECD DAC development evaluation criteria (relevance, efficiency, effectiveness, impact and sustainability), in that a small majority of reviews and evaluations examined had used them as a reference point. Several interviewees suggested that these criteria have ‘raised the bar’ for evaluations, acting as a prompt for those designing evaluation questions and ensuring that key issues are given due attention. However, a significant minority of interviewees, particularly among those running projects, questioned the ‘off-the-shelf’ use of the criteria (e.g. PNG and Uganda, also UK Government interviewees), arguing that while they are useful as a prompt, they are

³³ Molund S and Schill G, *Looking Back, Moving Forward: SIDA Evaluation Manual*, (SIDA, 2007), p 66.

³⁴ See for example op cit Anderson M et al, *Encouraging Effective Evaluation*, 2007.

a blunt tool.³⁵ In our judgement, the five DAC criteria must be tailored and elaborated in order to address aspects of OECD DAC-compliant SSR such as cross-sectoral working and linkages between sectors and programmes, questions about internal donor coherence, and the need to take a broad view of security and justice needs in-country.

There are of course other, additional evaluation criteria on offer. The humanitarian aid sector has adopted criteria such as coherence, coverage, linkages/connectedness and consistency with values,³⁶ and each criterion may be interpreted in several ways (see for example the different donor handbooks on M&E listed below, some of which offer different definitions or prompt questions under each of the five standard development criteria). The forthcoming OECD DAC guidance on evaluating conflict prevention and peacebuilding activities takes a combined approach, suggesting that all of the above criteria can be used and reinterpreted as needed.³⁷

In view of the complexities of SSR, the scope of OECD DAC-oriented SSR and the whole-of-government agenda, we find this combined approach the most promising. Using many of the existing guides to development evaluation listed here as our starting point,³⁸ but taking into account the specifics of SSR and SSR policy elaborated above, in the table below we suggest the following criteria and accompanying prompt questions for SSR evaluations. These questions and criteria should not be seen as fixed or indivisible, but rather as a menu from which those designing or carrying out evaluations can select as appropriate.

Criteria and prompt questions for evaluation of SSR programmes

Relevance/ appropriateness	<ul style="list-style-type: none"> ■ Is the intervention consistent with the justice and security concerns facing the state and its population? How urgent is it from the point of view of different target groups, particularly the poor and vulnerable? ■ Is the intervention based on an up-to-date context assessment, covering (as appropriate) conflict drivers, security and justice needs, institutional needs and capacities and drivers of change and political will? ■ Is the intervention in tune with the policies and administrative systems of the country and/or relevant international counterparts in the areas of development, security and peacebuilding? ■ Is the intervention a technically adequate solution to the security and justice problems facing the country and its population? Over time, will it address the main security and justice problems facing the country and its population? ■ Does the intervention balance considerations of long-term capacity-building for the state with more immediate service delivery, including through non-state mechanisms? ■ Has the intervention responded to changing circumstances over time? ■ Is the intervention consistent with donor policies and priorities?
Effectiveness	<ul style="list-style-type: none"> ■ To what extent do changes in the intervention's area of coverage match the intended outputs, purpose and goal? ■ To what extent are observed changes the result of the intervention rather than other factors? ■ What are the reasons for the delivery or non-delivery of the intervention's specified objectives? ■ What can be done to make the intervention more effective?

³⁵ The existence of these standard criteria does not guarantee that they are followed or appropriately interpreted during evaluations. For example, despite wide acceptance of the relevance criterion, in several of our case studies it could be argued that evaluations had not given adequate consideration to country context: the SILSEP II evaluation in Sierra Leone makes no real reference to the country's dire socio-economic conditions; one AusAID evaluation in PNG pays little attention to the dominance of oral traditions in PNG; and the JLOS evaluation in Uganda does not look seriously beyond the JLOS sector to question how programming addresses the broader security challenges facing Ugandans. The DAC criteria are also not without their critics. See for example Chianca T, 'The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement', *The Journal of MultiDisciplinary Evaluation*, vol 5 no. 9, 2008.

³⁶ See for example OECD DAC, *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies*, (OECD, 1999).

³⁷ OECD DAC, *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities: Working Document for Application Period*, (OECD, 2008), p 33.

³⁸ With particular thanks to the Swedish International Development Agency (SIDA), whose in-house evaluation manual we have followed closely here.

Efficiency	<ul style="list-style-type: none"> ■ Has the intervention been managed with reasonable regard for efficiency? ■ What measures were taken during planning and implementation to ensure that resources are efficiently used? ■ Could the intervention have been implemented with fewer resources without reducing the quality and quantity of the results? ■ Could more of the same result have been produced with the same resources? ■ Could an altogether different type of intervention have solved the same development problem but at a lower cost? ■ Was the intervention economically worthwhile, given possible alternative uses of the available resources?
Impact	<ul style="list-style-type: none"> ■ What are the intended and unintended, positive and negative effects of the intervention on people and institutions? How has the intervention affected the well-being of different groups of stakeholders? ■ What do beneficiaries and other stakeholders affected by the intervention perceive to be the effects of the intervention on themselves? ■ To what extent has the intervention contributed to the strengthening (including capacity and accountability) of institutions? To what extent has the intervention led to the development and improvement of relevant policies? ■ To what extent can changes that have occurred during the life span of the intervention or the period covered by the evaluation be identified and measured? ■ To what extent can identified changes be attributed to the intervention? What would have occurred without the intervention?
Sustainability and ownership	<ul style="list-style-type: none"> ■ What steps have been taken to create processes, structures and institutions through which the population can access justice and security over the long term? Has human as well as institutional capacity been built up? ■ Is the intervention consistent with partners' priorities and effective demand? Is it supported by local institutions and well integrated with local social and cultural conditions? ■ Has the intervention sought to build effective management and leadership of reforms? Did partner country stakeholders (including civil society and oversight actors) participate in the planning and implementation of the intervention? ■ Were the goods, services and technologies provided during the intervention to partner institutions appropriate to the economic, educational and cultural conditions in the partner country? ■ Do partners have the financial capacity to maintain the benefits from the intervention when donor support has been withdrawn? Is a credible exit strategy envisaged or in place?
Coherence	<ul style="list-style-type: none"> ■ Are different departments within individual donor governments co-operating sufficiently according to an agreed strategy and policy agenda? Are mechanisms in place for 'whole-of-government' support to SSR? ■ To what extent, if any, are donor concerns with 'hard' security issues (e.g. counter-terrorism) in conflict with development-style SSR objectives?
Co-ordination/linkages	<ul style="list-style-type: none"> ■ What steps have been taken to forge strategic engagement across the security and justice system, working across the different levels and institutions that make up the system? ■ Where possible, has the intervention forged links with other relevant programmes and frameworks, including (as relevant) peace support operations, post-conflict recovery and peacebuilding strategies and frameworks, and national development frameworks such as Poverty Reduction Strategy Papers (PRSPs)? ■ Is the intervention consistent and complementary with activities supported by different donor organisations (if present)?
Consistency with values	<ul style="list-style-type: none"> ■ Does the intervention promote norms of good and democratic governance, respect for human rights and the rule of law? ■ Is the intervention designed and carried out in accordance with basic governance principles of transparency and accountability? ■ Does the intervention promote equitable access to justice and security for populations, including the poor and vulnerable?

Beyond these criteria, another overarching question for any evaluation should be: *what has been learnt* throughout the life of this programme? This emphasises that it is not only by doing evaluations that lessons are learnt, and evaluators are not the only people capable of learning lessons. In fact, every programme and every activity will provide lessons in itself, and project staff (in both the partner government and the donor organisation) are likely to have learnt a great deal along the way. Drawing out and documenting these lessons can be just as important a task for evaluators as reaching their own conclusions – and indeed, the final evaluation is likely to be more thorough and useful if it builds on such lessons.

3.5 The M&E process: who should own, drive, and participate?

How M&E is carried out is just as important as *what* is done, particularly if we wish to ensure that the findings of M&E activities are valuable (rather than a ‘box-ticking exercise’) and that they are taken up by the appropriate actors. The previous chapter identified several issues regarding local ownership and participation in the M&E process. This short section recaps and builds upon some of these questions.

Regarding the question of who ‘drives’ M&E, at the stage when an SSR programme is first proposed there may be little appetite for an M&E framework among local actors. If so, the donors that are providing assistance to the SSR programme must start by raising the issue of M&E and exploring the attitudes of local actors towards it. This should be done early in the process of designing the SSR programme, and should seek to build understanding and agreement among key local stakeholders (including non-governmental actors) not only about *what* kind of M&E should be undertaken, but *why*. It may not be necessary to reach full agreement – greater consensus may develop naturally as the process develops and all actors ‘learn by doing’ – but donors should take care from the start to present M&E in a way that does not feel like an external imposition. As noted in the paragraphs on ‘Agreeing how to measure change’ above, stakeholders should work together during programme design to develop and agree on how change will be measured, and this may be an important step towards establishing local ownership at an early stage.

It has already been argued that donor-led SSR assessments should also analyse the capacity of the national government to undertake M&E. Since in most situations M&E capacity is likely to be low, there is a strong argument in favour of making capacity-building of national M&E systems (at the national/sectoral/sub-sectoral levels, as appropriate) an integral part of SSR programme design. If this is done correctly, it should lead to a gradual transfer of ownership of the M&E of SSR (and as a consequence, of the SSR programme itself) to the national government as local capacity is strengthened.

Ultimately, an SSR programme that has achieved sustainability would be expected to establish full local ownership of the M&E structures at the end of the programme. *For the duration of the programme*, however, what is needed is in fact *joint ownership* of the M&E structures between donor and national government. There can be no formula for what this joint ownership looks like in practice, since it must be appropriate to the context, but aspects to consider include:

- Ensuring that all relevant local and donor stakeholders (including non-governmental actors) are included in the management and reviewing structures that will oversee the M&E system and act on the information generated through M&E;
- Agreement about how information will be shared between local and donor actors (including non-governmental actors where it is appropriate to share such information), including agreed procedures for information sharing; and

- Identification of which M&E processes and structures most require improvement (at the national level, but possibly also at the donor level) and appropriate steps being taken by both donor and national government actors to make such improvements.

Another important point to emphasise is that all aspects of the M&E process should encourage the greatest possible degree of *participation* from different stakeholders. This means that not only should all relevant national state actors be involved in M&E (which will include several institutions beyond the security system, such as other relevant ministries and agencies, and also members of the legislature – parliamentarians, committee representatives, opposition politicians, etc), but also that there is broader civil society participation and ownership of the M&E process. Particularly during reviews and evaluations, care should be taken to consult with a wide variety of non-governmental voices representing different social categories and constituencies, since different groups are likely to have different perspectives on SSR. This helps both to provide a broader range of views on the SSR process and to understand how SSR has affected different categories of people. In most cases, local representatives should also be included in the evaluation team, since their insight into the local situation should help to ensure that the evaluation fully understands the local context.

Lastly, steps should be taken to increase the chances that the evaluation findings are taken up at the local level. Several actions were suggested in the section on ‘Ensuring take-up of M&E findings’ above, including joint reviews of evaluation findings, clarity about where the evaluations relate to donor assistance and where they relate to the national level, direct recommendations to national government actors where change is required at that level, and a succinct, jargon-free format which is accessible for people who may not be native English speakers.

4

Available resources for the M&E of SSR

THIS PAPER does not attempt to promote particular tools for the M&E of SSR over others. In such a complex area, it makes more sense to first present a general illustrative scheme setting out what needs to be measured, which can then be used alongside various context-specific tools. The desk research identified a number of products that may prove useful in designing or carrying out the M&E of SSR. Given that M&E as a discipline is well developed, the number of publications, websites and similar products that offer general guidance is too numerous to count. We have cited only a few examples which we found to be directly relevant. Material dedicated specifically to SSR/rule of law programming is much more limited, and it thus makes sense to reference much of this material in any guidance notes.³⁹ We also refer to key resources and consider their relevance for our own work.

4.1 SSR/rule of law-specific materials

Perhaps the earliest example of an attempt to give any specific guidance on evaluating SSR is a text by Heiko Borchert from 2003, *Security Sector Reform Initiative (SSRI): How to advance security sector reforms with the help of a new assessment and development framework*,⁴⁰ which proposed that SSR reviews be carried out along three dimensions:

- Democratic governance (including sector management, guidelines, goals and legislation and the role of civil society)
- Capability provision (including mission and role of sector actors, their quantitative description, capabilities, procurement and capacity-building)
- Co-operability (standards, joint security bodies, planning and development)

The *OECD DAC Handbook on SSR*, though it does not provide much in the way of detailed or specific guidance on M&E, is a useful reference point. Some of the advice it offers on programme management is relevant for M&E from a process point of view. Its advice on programme design and sector-specific assessments (e.g. justice, defence)

³⁹ Papers by Gordon Peake and by Dr Ann Fitz-Gerald and Dr Sylvie Jackson provide useful summaries of most SSR-specific material: Peake G, *Background Paper on Evaluating Security System Reform for the OECD DAC Workshop on Evaluating Security System Reform*, (University of Birmingham, 2006); Fitz-Gerald A and Jackson S, 'Developing a Performance Measurement System for Security Sector Interventions', *Journal of Security Sector Management*, vol 6 no. 1, 2008.

⁴⁰ Borchert H, *Security Sector Reform Initiative (SSRI): How to advance security sector reforms with the help of a new assessment and development framework*, paper prepared for the Annual Conference of the Working Group 'Security Sector Reform', Partnership for Peace Consortium of Defence Academies and Security Studies Institutes, Berlin, 15–17 June 2003.

illustrates some of the key areas in which SSR programmes seek change (the following categories are suggested for assessments: accountability and oversight, capacity, management, co-ordination with other parts of the system, and engagement with the international community).⁴¹ As well as offering loose definitions of M&E terms, sections 3.5–3.9 of the Handbook also give advice on using and setting indicators, including against the four overarching goals for SSR identified early on in the publication (local ownership, democratic governance, service delivery, management and sustainability). Section 3.9 offers the five standard OECD DAC development evaluation criteria as suggested criteria for SSR evaluations. However, while this guidance is informative, much is either too cursory (the indicator framework) or is not specific to SSR (the DAC criteria).

More recently, in *Developing a Performance Measurement System for Security Sector Interventions*, Dr Ann Fitz-Gerald and Dr Sylvie Jackson championed the use of a ‘balanced scorecard’ tool covering the following areas:⁴²

- **The societal perspective**, which includes rule of law, access to justice, legal norms, Millennium Development Goals, freedom from fear and want, PRSPs and access to daily subsistence
- **Enabling mechanisms**, including defence reform, justice reform, police reform, intelligence reform, free and fair elections and provision of basic services
- **The resources perspective**, which includes equipment, people, infrastructure and donor funding
- **Future perspectives**, including training, education, increasing awareness and exposure and sector-specific plans

Fitz-Gerald and Jackson’s categories cover both security and development issues and present a forward-looking perspective. The scorecard itself is a potentially interesting tool because it provides a kind of strategy map,⁴³ a visual depiction of the actions needed to achieve stated goals.⁴⁴

Looking more specifically at justice reforms, various methodologies also exist that seek to monitor and evaluate the ‘rule of law’, and the effectiveness of rule of law interventions. However, since there is very little international consensus about what actually constitutes the rule of law, its fundamental elements or the importance of one element relative to any other, different methodologies may look at very different issues and factors as part of their assessment. A paper by Maria Dakolias, *Methods for Monitoring and Evaluating the Rule of Law*, reviews many of the different tools that are available in this field, identifying a range of methodologies from highly quantitative indicator sets to strongly qualitative assessment tools.⁴⁵ Many of these are particularly focused on the intersection between the rule of law and economic development. Those that are most relevant to SSR and security and justice programming include:

- The American Bar Association’s Judicial Reform Index,⁴⁶ a set of 30 indicators and factors which establish standards in the areas of: quality, education and diversity of judges; judicial powers; financial resources; structural safeguards; transparency and judicial efficiency. The ABA Rule of Law Initiative uses this tool as the basis for detailed evaluations of judicial reform in countries around the world.

⁴¹ Op cit OECD DAC, *Handbook on Security System Reform*, 2007, p 126.

⁴² Op cit Fitz-Gerald A and Jackson S, ‘Developing a Performance Measurement System for Security Sector Interventions’, 2008.

⁴³ Ibid, p 11.

⁴⁴ Research by Cranfield University (ibid, p 14) suggests that the scorecard may have particular uses at the strategic and institutional levels. At the higher strategic level, where questions of donor coherence and donor-partner harmonisation loom large, a tool to aid discussion about higher-level objectives is indeed attractive. While donor programmes will almost certainly continue to use conventional measurement tools such as logframes (underpinned by logic models) and indicators, developing country justice and security institutions may fare better with a mixed approach: measuring progress against a few priority indicators for internal performance measurement, and using a scorecard tool to aggregate, summarise and communicate various aspects of performance.

⁴⁵ Dakolias M, ‘Methods for Monitoring and Evaluating the Rule of Law’ in Centre for International Legal Cooperation, *Applying the “Sectoral Approach” to the Legal and Judicial Domain*, (CILC, 2005).

⁴⁶ An overview of the Judicial Reform Index can be found on the American Bar Association website at: www.abanet.org/rol/publications/judicial_reform_index.shtml, accessed 23 September 2009.

- The World Bank's Worldwide Governance Indicators⁴⁷ are based around six dimensions of governance, one of which is the rule of law. There is one overall indicator for the quality of the rule of law which is calculated by aggregating information from other sources, most of which are perception-based studies by other organisations. The World Bank website also has a series of links to detailed resources on 'performance evaluation' for law and justice institutions.⁴⁸

While not a methodology as such, a good example of how to undertake a detailed evaluation of donor-assisted judicial reform at the sector/system and country level can be found in the US General Accounting Office (GAO) evaluation of US Rule of Law Assistance to former Soviet Union countries. The methodology – used in its broader sense to include the process of setting scope and objectives – was highly qualitative, based on extended field visits and large numbers of stakeholder interviews both in the US and in partner countries.⁴⁹

The Vera Institute's *Global Guide to the Design of Performance Indicators across the Justice Sector*⁵⁰ offers standard indicators for key institutions or aspects of the justice sector (police, prosecution and defence, judiciary, non-custodial sentencing, prisons and accountability mechanisms, non-state institutions). Unlike most other materials found on the subject, it also offers 'strategic level' (i.e. sector level) indicators for both safety and security and access to justice. In each case, indicators are matched to desired outcomes and guidance is given on data sources and the pros and cons of using particular indicators.

*The PRIME System: Measuring the Success of Post-Conflict Police Reform*⁵¹ is also instructive. It is a framework for measuring the success of post-conflict police reform developed by Princeton University and field-tested in collaboration with UNPOL. Consciously developed to go beyond traditional police reform measurement systems which have tended to be output-focussed, and specifically targeted at post-conflict police services and their would-be reformers, the system essentially provides a set of generic indicators along four dimensions (pillars) that correspond with commonly accepted good policing practice:

- **Performance Effectiveness:** issues such as capacity (e.g. adequate manpower), authority and reach, handling of crime statistics and external co-ordination with the justice system
- **Management and oversight:** personnel issues, organisational procedures, planning and monitoring and internal and external oversight and accountability mechanisms (e.g. the existence and effectiveness of an Ombudsperson's office)
- **Community relations:** issues from human rights to public acceptance
- **Sustainability:** indicators for budgeting, training and equipment, political independence and staff salaries and benefits⁵²

The authors advise that the indicator set should be used as a menu and tailored to specific projects and contexts depending on mission objectives and other factors.

Similarly, in their text *The Police That We Want*, developed to cater for the effective reform and oversight of the post-apartheid South African police, David Bruce and Rachel Nield suggested five dimensions of change and provided example indicators for each:

- Protecting democratic political life
- Governance accountability and transparency

⁴⁷ Available online at: <http://info.worldbank.org/governance/wgi/index.asp>, accessed 23 September 2009.

⁴⁸ Available online at: <http://go.worldbank.org/LRFA0Q06E1>, accessed 23 September 2009.

⁴⁹ United States General Accounting Office, *Former Soviet Union: U.S. Rule of Law Assistance Has Limited Impact*, (US GAO, 2001).

⁵⁰ Op cit Foglesong et al, *Measuring Progress toward Safety and Justice*, 2003.

⁵¹ Bajraktari Y et al, *The PRIME System: Measuring the Success of Post-Conflict Police Reform*, (Woodrow Wilson School of Public and International Affairs, Princeton University, 2006).

⁵² Ibid, p 8.

- Service delivery for safety, security and access to justice
- Proper police conduct
- Police as citizens

There is also an in-house indicator framework developed by UNDP's Bureau for Crisis Prevention and Recovery (BCPR) for 'Justice and Security Sector (JSSR)' work (see the UN Survey paper for more details).⁵³

In parallel with this type of work, a vibrant and contested literature and body of practice exists concerning the performance measurement of security and (more typically) justice institutions in developed countries, which may be partially applicable to the measurement of SSR. The most advanced examples include the US COMPSTAT and UK Policing Performance Assessment Framework (PPAF) systems for policing,⁵⁴ and any number of strategic planning and customer accountability plans and frameworks in use by institutions ranging from the UK's Ministry of Defence and Border Agency to the US Office of Border Patrol.⁵⁵ Over the last twenty years or so, justice systems from Latin America to Russia have also begun to develop performance measurement systems of varying degrees of detail and sophistication.

Most of these materials are in some way useful for our purposes. All offer a perspective on the key dimensions of change sought by SSR/rule of law work. The more elaborate systems (e.g. the Vera Institute's Global Guide, the PRIME system) deserve to be referenced as a possible source of performance indicators. Yet while inspiration can be derived from such sources, many do not sit well with the realities of less developed, fragile and conflict-affected states. Many either neglect or do not adequately address dimensions like public demand, empowerment and satisfaction. These dimensions should be measured as a way of triangulating information from different sources, putting security institutions under closer scrutiny, and building a civil society voice on security and justice issues.

4.2 Guidance on related issues

Alongside guidance specific to areas such as SSR/rule of law, we also identified a number of products related to M&E in conflict-affected or fragile environments. One of the most relevant is the OECD DAC's draft *Guidance on evaluating conflict prevention and peacebuilding*,⁵⁶ which contains detailed discussion of matters such as evaluation criteria, indicators and the applicability of notions such as 'impact' in these types of environments.

We also found more recent guidance on *Monitoring, Evaluation and Learning for Fragile States and Peacebuilding Programs*, produced by the consultancy Social Impact for USAID, to be useful. This has illustrative indicators for programming related to SSR and intelligently discusses some of the more innovative M&E tools.⁵⁷

The OECD DAC's recent policy paper on *Whole of Government Approaches to Fragile States*⁵⁸ identifies issues around whole-of-government working that are increasingly acknowledged as important to the process side of SSR.

Lastly, a set of standard indicators designed by the United States Institute of Peace (USIP) for the US State Department to guide results-based management in conflict contexts, *Measuring Progress in Conflict Environments (MPICE): Metrics Framework*

⁵³ Dalrymple S, *Survey of the United Nations Organisation's arrangements for monitoring and evaluating support to security sector reform*, (Saferworld, 2009).

⁵⁴ Policing Performance Assessment Framework (PPAF), available at: <http://police.homeoffice.gov.uk/performance-and-measurement/managing-police-performance/?view=Standard>, accessed 28 June 2009.

⁵⁵ US Office of Border Patrol and Office of Policy and Planning, *Border Patrol Strategy: Office of Border Patrol* (US Customs and Border Protection, 2004).

⁵⁶ Op cit OECD DAC, *Guidance on Evaluating Conflict Prevention*, 2008.

⁵⁷ Op cit Sartorius and Carver, *Monitoring, Evaluation and Learning*, 2006.

⁵⁸ OECD DAC, *Whole of Government Approaches to Fragile States*, (OECD, 2006).

for *Assessing Conflict Transformation and Stabilization*, is also instructive,⁵⁹ though we would strongly counsel against using standard indicators that are not context-specific. USAID's handbook of governance indicators⁶⁰ attempts a similar exercise for democratisation and good governance work.

4.3 General guidance on M&E

As stated above, there is no shortage of general guidance material on M&E. The most apposite material includes the OECD DAC document *Evaluating Development Co-operation: Summary of Key Norms and Standards*,⁶¹ which offers useful checklists and prompt questions to help assess the quality of an evaluation, and in-house donor guides on M&E produced by DFID,⁶² SIDA,⁶³ the World Bank⁶⁴ and UNDP,⁶⁵ which together give many useful insights on both the substance of M&E and the procedural or bureaucratic side of things. Taken together, they offer a wealth of advice on everything from evaluation criteria and prompt questions for evaluators to writing TORs or forming steering groups to guide evaluations. A draft practice paper by DFID on M&E in fragile and conflict-affected contexts⁶⁶ and a World Bank guide to 'real world' evaluations are also useful, since each covers obvious pitfalls in these difficult contexts, advising for example on the use of mixed methods and data sets to overcome data shortages and bias.⁶⁷

⁵⁹ Dzedzic M, Sotirin B, and Agoglia J, *Measuring Progress in Conflict Environments (MPICE) – A Metrics Framework for Assessing Conflict Transformation and Stabilization*, (United States Institute for Peace, 2008).

⁶⁰ USAID Center for Democracy and Governance, *Handbook of Democracy and Governance Program Indicators*, (USAID, 1998).

⁶¹ Op cit OECD DAC Network on Development Cooperation, *Evaluating Development Cooperation*, 2008.

⁶² Op cit DFID, *Guidance on Evaluation and Review*, 2005.

⁶³ Op cit Molund and Schill, *Looking Back, Moving Forward*, 2007.

⁶⁴ World Bank, *Monitoring & Evaluation: Some Tools, Methods and Approaches*, (World Bank, 2004) provides an overview and also links to more detailed tools such as the World Bank's *Key Performance Indicator Handbook and Logframe Handbook*.

⁶⁵ United Nations Development Programme Evaluation Office, *Handbook on Monitoring and Evaluating for Results*, (UNDP, 2002).

⁶⁶ This paper is currently unpublished.

⁶⁷ Bamberger M, *Designing quality impact evaluations under budget, time and data constraints*, (World Bank, 2005).

5

The demand for guidance and next steps

5.1 Demand factors

AS NOTED ABOVE, while guidance on some aspects of the M&E of SSR has been developed, nothing is available that takes the latest policy thinking as its departure point and comprehensively covers process *and* substance, monitoring and evaluation. There is thus a need for tailored guidance on M&E. During our research we were keen to identify the requirements of specific potential users of guidance, the factors that drove demand, and any obvious upcoming opportunities to incorporate new guidance into different organisations' procedures and workplans.

One obvious factor driving demand for guidance among donor organisations is the aid effectiveness agenda and the pressure to manage for results. Acceptance of this thinking now appears to reach beyond aid institutions into other arms of government and multilateral activity (e.g. the UN Department of Peacekeeping Operations (DPKO) as well as UNDP, Ministry of Defence (MOD) as well as DFID). This in turn means that there is high interest across most donor countries, and across individual departments within countries, in improving M&E practice for SSR work. However, most of our country case studies and donor surveys suggest that there has been limited concrete progress in measuring and managing for results.

A second factor, discussed earlier, is the growing awareness among donors and developing country governments of the sheer difficulty of reforming security sectors and the patchy record of donor SSR programmes to date. Many who are aware of these difficulties see better M&E as part of the solution, though as a couple of cases in our research suggest, donors are likely to be more bullish about this than partner government personnel where local ownership over programmes – and M&E – is lacking.

5.2 Potential users of any guidance

In the course of our research, various users and targets were identified across different donor governments and multilateral institutions. Although we found fairly consistent demand for guidance (and in particular indicator sets) among all surveyed donors and their individual departments, it was perhaps highest within the UK and the US, with the EU and UN also expressing strong interest. Some have already begun developing policies on SSR, which provides an entry point for further work, or specific tools and guidance with which to link. For example:

Forthcoming policies or strategies	<ul style="list-style-type: none"> ■ The UK, the Netherlands and the US are making moves to develop cross-governmental SSR policies, which creates obvious opportunities for discussion around supporting implementation. ■ DFID's evaluation department has been liaising with evaluation specialists in the MOD and Foreign and Commonwealth Office (FCO) to try and agree a common approach to evaluation in fragile and conflict-affected states.
Forthcoming evaluations	<ul style="list-style-type: none"> ■ The Netherlands Ministry of Foreign Affairs (MFA) has proposed a strategic level thematic evaluation of SSR work for 2010. ■ SSR may receive similar treatment as part of an evaluation of DFID's conflict prevention or fragile states work at around the same time (linked with development of a new evaluation policy for DFID). ■ A similar proposal is under consideration by DPKO's Office of Internal Oversight Mechanism. Also within the UN, DPKO's best practices unit has been tasked with looking at an M&E framework for SSR.
Programme opportunities	<ul style="list-style-type: none"> ■ DFID has also let a contract for the M&E of a Security Sector and Police Accountability Programme in DRC to a consultancy, GRM International, who have expressed an interest in testing any guidance produced by Saferworld. ■ AusAid has expressed interest in testing any guidance developed by Saferworld on the M&E of SSR in support to some of its justice sector work in the Asia-Pacific region.
Requests for guidance	<ul style="list-style-type: none"> ■ Over and above this, the OECD DAC International Network on Conflict and Fragility (INCAF) has committed to adding an additional chapter to its handbook on Security and Justice on the subject of M&E and is amenable to adding SSR-related examples to its draft guidance on evaluating conflict prevention and peacebuilding work.

There is also considerable potential for any guidance to be used by the International Security Sector Advisory Team (ISSAT), a multi-donor initiative developed at the Geneva Centre for the Democratic Control of Armed Forces (DCAF). ISSAT is planning to utilise and build on existing tools to develop operational guidance on M&E mechanisms for its donor members.

In short, there are numerous potential targets and a number of obvious opportunities on offer, each of which is worth addressing.

5.3 Towards specific guidance

Previous sections of this report have discussed key issues relating to the M&E of SSR, including: what to measure during attempts to reform security systems; how those designing M&E systems or carrying out evaluations can be supported through tailored frameworks, guiding questions and criteria; what resources can be drawn upon in this effort; and who the most immediate users of specific guidance on this subject might be and what their specific needs are.

Together, the answers to these questions suggest a way forward in developing tailored guidance as a contribution to more robust and appropriate M&E in this area, and as a precursor to any further activities. The most immediate follow-on questions to address, are around the exact content, format and ownership of any guidance that Saferworld might produce. Having identified a good level of demand and a range of opportunities and potential users, it would be easy to conclude that one size will not fit all. However, we feel that donor governments have a sufficiently shared understanding of the issues (as demonstrated by commitment to OECD DAC SSR policy, similar individual policy statements on SSR, and concerns to meet Paris and Accra commitments), and share enough institutional characteristics (e.g. use of logframes), to warrant the production of a single guidance product on this subject. This would be slanted towards donor programme staff who bear most responsibility for designing programmes and commissioning evaluations. It could later be drawn on to produce more targeted material for individual governments and agencies, evaluators, field-level SSR project staff (including non-governmental and private contractors), or partner governments.

The box below sets out some essential points to include in any guidance:

Essential points to include in any guidance on M&E of SSR

Introductory points

- 1 Definitions and purposes of M&E
- 2 Importance of M&E for effective SSR (and donor assistance)
- 3 Challenges and particularities of M&E in this area (e.g. secrecy within the sector, both on donor and partner government sides; the comprehensive and norm-based nature of OECD SSR; challenges of data collection)

'Content' issues

- 4 Different levels and focuses for M&E (e.g. donor performance or partner government institution performance; security system or institutional level)
- 5 Focusing on both process and results
- 6 Dimensions of change and suggestive indicators (at system, institutional level and for key themes)
- 7 Identifying theories of change
- 8 Evaluation criteria

Programmatic/process issues

- 9 Building legitimate and skilled evaluation teams
- 10 Generating broad-based participation and ownership of M&E (including the roles of HQ; field teams; partners; beneficiaries) and how to involve them
- 11 Aiding conflict-sensitivity
- 12 Building capacity for M&E among partner governments who are undertaking SSR
- 13 Building M&E in project management arrangements (an M&E 'framework')
- 14 Moving from M&E to lesson learning

In addition, it may also prove helpful to produce a number of supporting tools, for example:

- A generic terms of reference (TOR) for SSR evaluations
- A checklist for assessing the content of evaluations
- Illustrative indicators covering the system level and key institutions and themes
- References to other guidance products and indicator sets (as listed earlier)
- Summary of different data collection methods

The case for including most of these items has been made in earlier parts of this report, but some items require additional commentary. For example, on the basic question of M&E terminology, while our research showed that all donor government institutions involved in SSR, from defence to foreign ministries, are involved in tracking the progress of their work and reporting on it, there is still much confusion and nervousness around the language of M&E.

We found strong appetite in many quarters for indicator sets illustrating the key dimensions of change sought by SSR programmes, probably because this sits easily with what are now fairly standardised approaches to planning, reporting and M&E, including the use of logframes. Indicator sets that are badly designed and poorly used can be stultifying and can deaden the imagination of programme designers. However, they can be very helpful if they are used merely to illustrate what is critical for successful programme design and M&E, and formatted clearly. If derived from clear policy and guidance such as the OECD DAC SSR Handbook, such frameworks can also encourage donor staff to comply with espoused norms and key principles and work together with partners to include these standards in programmes. Moreover, so long as indicator frameworks are disaggregated by different levels of the results chain (outputs, outcomes, etc) and along the different dimensions of change that SSR processes seek (e.g. increased capacity, better management, governance accountability, public empowerment), they are less likely to be limited to merely repackaging actual programme designs in a new format. Given that many donor programmes lack explicit and credible intervention logics, this is another potential advantage of developing tailored M&E guidance.

Our research showed a real need to clarify and better illustrate the different dimensions along which changes should occur through SSR. Based on past experience,⁶⁸ we believe it is possible to design such a framework. Yet given the scope and technical complexity of OECD DAC-type SSR, any such framework may need to be constructed so as to capture a number of cross-cutting and component aspects of SSR. There is a case for including example indicators that relate to overarching or cross-cutting themes (e.g. ownership, gender-sensitivity, human rights). Institutional or sectoral indicators were much in demand from our interviewees (i.e. defence reform, justice, corrections, etc), while the system-wide requirements of current policy thinking (e.g. around cross-sectoral working) should also feature. We would however counsel strongly against the use of any illustrative indicators as standard or generic indicators that could be taken 'off the shelf' and transplanted directly into programme logframes, as this is the antithesis of good programme design and respect for local ownership.

Lastly, it is worth re-emphasising a concern that has been repeated throughout the paper regarding the ownership of the M&E process itself and the M&E capacity of partner governments. Any guidance on the M&E of SSR needs to indicate how local ownership of the evaluation process can best be secured. This is likely to focus both on how to engage a variety of local actors more effectively throughout the process of designing and implementing M&E procedures, and also to recommend that in many cases donors should include capacity-building of partner government M&E practices in programme design.

5.4 Next steps

Based on the findings of this research, Saferworld intends to develop a detailed guidance product that will be of practical use to a range of users and can easily be tailored to the needs of specific users (such as national governments engaged in SSR, field-office donor staff, more senior decision-makers in donor organisations, and non-governmental actors who are engaged in or interested in security and justice programming), taking account of their different perspectives and requirements regarding SSR processes. This will be tested with key stakeholders, including not only key donor governments and multilateral bodies, but also partner governments and other relevant local actors, in order to ensure that it is relevant and useful. Saferworld hopes that this will contribute to improved SSR programmes and in turn to better security and justice outcomes for their intended beneficiaries.

At the same time, OECD DAC's International Network on Conflict and Fragility (INCAF) is planning to develop a practical toolkit on the M&E of security and justice programming. It is expected that Saferworld will play a key role in this process, using this paper and the Saferworld guidance product to inform the OECD DAC toolkit.

In time, we also hope that more rigorous M&E of donor support to SSR at the programme level will provide a basis for a broader policy evaluation of donor efforts in this area, as presaged by the OECD DAC Ministerial statement of 2004.⁶⁹ We found that although most donors have in-house guidance on M&E, in many cases this had not translated into results-focused monitoring, reviewing and evaluation.⁷⁰ Instead of true impact evaluations of SSR work, we found instead that the country case studies were framed mainly around mid-term and end-of-project reviews.⁷¹ While such

⁶⁸ In an unpublished document written for SIDA in 2008, Saferworld elaborated a matrix of objectives and indicators specific to SSR that might be used to guide M&E. Unlike most other frameworks, this carefully distinguishes both the main dimensions along which change is intended and the impact, outcome and output levels of a standard programme logic. In addition, potential programme impacts are designated either 'direct' (e.g. enhanced physical security) or 'indirect' (greater material wealth). This tool was well received.

⁶⁹ '...The key principles agreed in the 2004 DAC SSR Guidelines need to be translated into evaluation indicators.' Op cit OECD DAC, 'Ministerial Statement', 4 April 2007, pp 10–12.

⁷⁰ Notable exceptions were the impact-focused M&E frameworks championed by donors in our two justice-focused case studies (Uganda and PNG), which may be due to the sector-wide approach used in both cases.

⁷¹ Partial exceptions were the UK Government reviews of its Conflict Prevention Pools. However, while these contain many insights and allow comparisons to be made between different activities, projects and programmes, they are understandably focused on process issues, such as inter-departmental coherence, planning, financing, etc.

studies generate some lessons about programmes, they necessarily focus more on questions of efficiency and effectiveness than on beneficiary impact, relevance to country context, or the wisdom behind programme design or donor policy. We are therefore considering a follow-on piece of work around (a) supporting the development of a credible evidence base regarding the impact of donor SSR programming, and (b) encouraging greater accountability towards stated policy commitments, in part through policy-level evaluation.

Bibliography

- Albrecht P, *Monitoring and evaluation arrangements for the Sierra Leone Security Sector Reform Programme: a case study*, (Saferworld, 2009)
- Alvarez A et al, *United Nations Development Programme Haiti: Outcome Evaluation: Outcome 7: Rule of Law, Justice and Human Rights (Draft, 4 July 2003)*, (UNDP, 2003, unpublished)
- Anderson M et al, *Encouraging Effective Evaluation of Conflict Prevention and Peacebuilding Activities: Towards DAC Guidance*, (OECD DAC, 2007)
- Armytage L, *Monitoring Performance of Legal and Judicial Reform in International Development Assistance: Early Lessons from Port Moresby and Phnom Penh*, paper presented at International Bar Association Chicago Showcase on 'Judicial Reform: Economic Development and the Rule of Law', 2006, available at: <http://siteresources.worldbank.org/INTLAWJUSTINST/Resources/MonitoringImpact.pdf>, accessed 27 June 2009
- Asiimwe A et al, *Justice, Law and Order Sector Strategic Investment Plan Mid-term Evaluation 2001/2–2005/6, Volumes one and two, Final Report December 2004*, (International Human Rights Network, 2004)
- Asiimwe A, Fox J and Mukasa S, *Justice Law and Order Sector Technical Note on Developing Sector-Wide Output and Impact Indicators*, (2005, unpublished)
- AusAID, *PNG Law and Justice Sector Review and Contribution Analysis, December 2006*, (AusAid, 2006, unpublished).
- AusAID and Government of Papua New Guinea, *Law and Justice Sector Program Final Design Document*, (AusAID and Government of PNG, 2005, unpublished)
- AusAID and Government of Papua New Guinea, *Papua New Guinea – Australia Law and Justice Partnership (May 2008)*, Programme Design Document, (AusAID and Government of PNG, 2008, unpublished)
- Babaud S, *Survey of the European Union's arrangements for monitoring and evaluating support to security sector reform*, (Saferworld, 2009)
- Bajraktari Y et al, *The PRIME System: Measuring the Success of Post-Conflict Police Reform*, (Woodrow Wilson School of Public and International Affairs, Princeton University, 2006)
- Ball N et al, *Evaluation of the Conflict Prevention Pools: The Security Sector Reform Strategy*, (DFID, 2004)
- Ball N et al, *From Project to Program: Effective Programming for Security and Justice* (Conflict Research Unit, Netherlands Institute of International Relations Clingendael, 2007)
- Ball N and van de Goor L, *Promoting Conflict Prevention through Security Sector Reform: Review of Spending on Security Sector Reform through the Global Conflict Prevention Pool*, (PricewaterhouseCoopers, 2008)
- Ball N, Biesheuvel P, Hamilton-Baillie T and Olonisakin F, *Internal Review of ACPP and DFID Security and Justice Sector Reform Programming in Africa: Inception Report*, (DFID, 2006)
- Ball N, Biesheuvel P, Hamilton-Baillie T and Olonisakin F, *Security and Justice Sector Reform Programming in Africa*, (DFID, 2007)
- Bamberger M, *Designing quality impact evaluations under budget, time and data constraints*, (World Bank, 2005)
- Bastick M and Valasek K (eds), *Gender & Security Sector Reform Toolkit*, (DCAF/OSCE/ODIHR/UN-INSTRAW, 2008)
- Bocco R, de Martino L and Luethold A, *Palestinian Public Perceptions of Security Sector Governance: Summary report*, (DCAF, 2005)
- Biekart K et al, *Evaluation of the Dutch 'Governance and Human Rights Programme' in Guatemala 1997–2003*, (Policy and Operations Evaluation Department, Dutch Ministry of Foreign Affairs, 2004)
- Biesheuvel P, *Analysing Options for Implementing an Inter-Departmental Security Sector Reform Approach in the Netherlands*, (2008, unpublished)
- Biesheuvel P, Hamilton-Baillie T and Wilson P, *Sierra Leone Security Sector Reform Programme, Output to Purpose Review*, (Security Sector Development Advisory Team, 2007, unpublished)
- Biesheuvel P et al, *Sierra Leone Security Sector Reform Programme III, Output to Purpose Review*, (Security Sector Development Advisory Team, 2005, unpublished)

- Biesheuvel P and Stapleton A, *UK Justice Sector Programme Sudan, Output to Purpose Review, May 2007* (DFID, 2007, unpublished)
- Björkman M and Svensson J, 'Power to the People: Evidence from a Randomized Field Experiment of a Community-based Monitoring Project in Uganda', *The Quarterly Journal of Economics*, vol 124 issue 2, 2009
- Borchert H, *Security Sector Reform Initiative (SSRI): How to advance security sector reforms with the help of a new assessment and development framework*, paper prepared for the Annual Conference of the Working Group 'Security Sector Reform', Partnership for Peace Consortium of Defence Academies and Security Studies Institutes, Berlin, 15–17 June 2003, available at: http://www.borchert.ch/paper/SSR_Paper.pdf, accessed 27 June 2009
- Brambilla P, *Gender and Monitoring: A Review of Practical Experiences*, (BRIDGE, Institute of Development Studies, University of Sussex, 2001)
- Bruce D and Nield R, *The Police That We Want: A Handbook for Oversight of Police in South Africa*, (Centre for the Study of Violence and Reconciliation, 2005)
- Brzoska M, *Development Donors and the Concept of Security Sector Reform*, (DCAF, 2003)
- Buvinić M, Alda E and Lamas J, *Emphasizing Prevention in Citizen Security: The Inter-American Development Bank's Contribution to Reducing Violence in Latin America and the Caribbean*, (Inter-American Development Bank, 2005)
- Caputo E et al, *Methodology for Evaluations of Budget Support Operations at Country Level: Issue Paper*, (EuropeAid Evaluation Unit, 2008)
- Cavelier B et al, *Evaluation Guide 2007*, (Direction générale de la coopération internationale et du développement, Ministère des Affaires étrangères et européennes (France), 2007)
- Chanaa J, *Security Sector Reform: Issues, Challenges and Prospects*, International Institute for Strategic Studies, Adelphi Paper 344, (Oxford University Press, 2002)
- Chapman J and Mancini A, *Impact assessment: Drivers, dilemmas and deliberations*, (Sightsavers International, 2008)
- Chianca T, 'The OECD DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement', *The Journal of MultiDisciplinary Evaluation*, vol 5 no. 9, 2008, available at http://survey.ate.wmich.edu/jmde/index.php/jmde_1/article/view/167/183, accessed 26 June 2009
- Chukwuma I, 'Designing Indicators of Safety and Justice: Lessons from the CLEEN Foundation's National Crime Victims Surveys in Nigeria', *CLEEN Foundation Newsletter*, July 2008, available at: http://www.cleen.org/newsletter/E__cleen-web-site_newsletter_newsletter_paperpresentatio.pdf, accessed 27 June 2009
- Church C and Shouldice J, *The Evaluation of Conflict Resolution Interventions: Framing the State of Play*, (INCORE/United Nations University, 2002)
- Cohen C, *Measuring Progress in Stabilization and Reconstruction*, (United States Institute of Peace, March 2006)
- Commins S, *Community Participation in Service Delivery and Accountability*, (UCLA, 2007)
- Conflict Research Unit of the Netherlands Institute of International Relations Clingendael, *Towards a Whole-of-Government Approach to Security System Reform, Conference background paper: The Hague 9–10 April 2008*, (Netherlands Institute of International Relations Clingendael, March 2008)
- Conflict, Security and Development Group, *State Responsiveness to Public Security Needs: The Politics of Security Decision-Making – A Comparative Study of Nigeria, Sri Lanka and Uganda, SSR Policy Briefing for HMG*, (King's College London, 2008)
- Crisis Management Initiative, *Conference proceedings from the Crisis Management Initiative meeting 'Workshop on evaluation of crisis management operations: Methodologies and Practices'*, 7 November 2007, Brussels, Belgium
- Dakolias M, 'Methods for Monitoring and Evaluating the Rule of Law' in Centre for International Legal Cooperation, *Applying the "Sectoral Approach" to the Legal and Judicial Domain*, (CILC, 2005), available at: www.cilc.nl/Conference_publication_2005.pdf, accessed 23 September 2009
- Dalrymple S, *Survey of the United Nations Organisation's arrangements for monitoring and evaluating support to security sector reform*, (Saferworld, 2009)
- Dalrymple S and Rynn S, *Monitoring and Evaluation arrangements for the Law and Justice Sector in Papua New Guinea: a case study*, (Saferworld, 2009)
- Davies R and Dart J, *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*, (MandE/Clear Horizon, 2005)

- DeBlieck S, *The Critical Link: Community Policing Practices in Southeastern Europe*, (UNDP Albania/SSSR Programme, 2007)
- DeBlieck S, *Monitoring and evaluation arrangements for the Support to Security Sector Reform Programme in Albania: a case study*, (Saferworld, 2009)
- DeBlieck S, *Representation, Relevance and Interest: An Assessment of the SSSR Programme's Community Problem Solving Groups*, (UNDP Albania/SSSR Programme, 2007)
- Department for International Development (UK), *Building the State and Securing the Peace*, (DFID, 2009)
- Department for International Development (UK), *Guidance on Evaluation and Review for DFID Staff*, (DFID Evaluation Department, 2005)
- Department for International Development (UK), *Independent Evaluation in DFID, Annual Report 2007/08*, (DFID Evaluation Department, 2008)
- Department for International Development (UK), *Malawi Safety, Security and Access to Justice Programme, Output-to-Purpose Review, September/October 2003*, (DFID, 2003)
- Department for International Development (UK), *SJG (Nigeria) Annual Performance Review 2007*, (DFID, 2007)
- Donais T (ed), *Local Ownership and Security Sector Reform*, (DCAF, 2008)
- Duignan P, *Principles of Outcome Hierarchies: Contribution Towards a General Analytical Framework for Outcomes Systems (Outcomes Theory): Strategic Evaluation Working Paper, Strategic Evaluation*, available at: www.strategievaluation.info/se/documents/122f.html, accessed 26 June 2009
- Dziedzic M, Sotirin B, and Agoglia J, *Measuring Progress in Conflict Environments (MPICE) – A Metrics Framework for Assessing Conflict Transformation and Stabilization*, (United States Institute for Peace, 2008)
- Engel P, Land T and Keijzer N, *A balanced approach to monitoring and evaluating capacity and performance: A preliminary proposal*, (European Centre for Development Policy Management, 2006)
- Fitz-Gerald A and Jackson S, 'Developing a Performance Measurement System for Security Sector Interventions', *Journal of Security Sector Management*, vol 6 no. 1, 2008
- Flew C and Rynn S, *Monitoring and evaluation arrangements for the Justice Law and Order Sector in Uganda*, (Saferworld, 2009)
- Foglesong T et al, *Measuring Progress toward Safety and Justice: A Global Guide to the Design of Performance Indicators across the Justice Sector*, (Vera Institute of Justice, 2003)
- Gannon M et al, *Criminal Justice Indicators 2005*, (Statistics Canada, Canadian Centre for Justice Studies, 2005)
- Germann W, *Evaluation of Security Sector Reform and Criteria for Success: Practical Needs and Methodological Problems*, (DCAF, 2002)
- Government of Guinea-Bissau and United Nations Development Programme, *Document de Projet: Appui au Renforcer l'État de Droit en Guinée Bissau à travers la réforme de la justice et la sécurité*, (Government of Guinea-Bissau/UNDP, 2008)
- Government of Jamaica, *Jamaican Justice System Reform: Project Strategy and Operational Plan*, (Government of Jamaica, 2006), available at: www.jis.gov.jm/Project_Strategy_and_Operational_Plan.pdf, accessed 27 June 2009
- Government of Uganda, *Justice, Law and Order Sector, Strategic Investment Plan II, 2006/7–2010/11, Consolidating Gains and Enhancing Impact*, (Republic of Uganda, 2006, unpublished)
- Government of the United Kingdom, *PSA Delivery Agreement 30: Reduce the impact of conflict through enhanced UK and international efforts*, (Her Majesty's Government, 2007)
- Haden P, *Objectives and Indicators for the Malawian Justice Sector – the Process of Development, (Draft) Report*, (International Organisation Development, 2003)
- Haden P, *Objectives and Indicators for the Malawian Justice Sector – the Current Scenario, (Final) Report*, (International Organisation Development, 2003, unpublished)
- Higgins K, *The Global Conflict Prevention Pool Caribbean Strategy 2004–2007: An Evaluation*, (DFID, 2007, unpublished)
- Hvidemose D and Mellon J, *Monitoring and evaluation arrangements for the implementation of community policing in Bosnia and Herzegovina: a case study*, (Saferworld, 2009)

- International Centre for Migration Policy Development and TC Team Consult, *Financial, Organisational and Administrative Assessment of the BiH Police Forces and the State Border Service*, (ICMPD and TC Team Consult, 2004)
- Ismael O and Hendrickson D, *What is the case for a security and justice focus in development assistance programming? An assessment of existing literature and evidence*, (GFN-SSR, 2009)
- ITAD Ltd and RuralNet Associates, *Participatory Monitoring and Evaluation: Documentation to Support Generic Framework*, (ITAD, 2004), available at: www.itad.com/newweb/pubs/PM&E%20Framework%20AUGo4.pdf, accessed 26 June 2009
- Jones N, Jones H, Steer L and Datta A, *Improving Impact Evaluation and Use*, (Overseas Development Institute, 2009)
- Jones R and Young V, *CIDA Evaluation Guide*, (Evaluation Division, Performance & Knowledge Management Branch, Canadian International Development Agency, 2004)
- Jones S et al, *Establishing Law and Order After Conflict*, (RAND Corporation, 2005)
- Kelly L, David R and Roche C, *Guidance on M&E for Civil Society Programs*, (Demand for Better Governance Unit, AusAid, 2008)
- Kennedy School of Government at Harvard University, *Indicators of Safety and Justice: Their Design, Implementation and Use in Developing Countries: Summary of a Workshop Held at Harvard University, 13–15 March 2008*, (Harvard, 2008), available at: www.hks.harvard.edu/criminaljustice/publications/justice_indicators_workshop_2008.pdf, accessed 26 June 2009
- Kenway J, *AusAID Assistance to Papua New Guinea's Law and Justice Sector (2003–2007): Monitoring and Evaluation*, (AusAID, 2007, unpublished)
- Kusek J Z and Rist R C, *Ten Steps to a Results-Based Monitoring and Evaluation System*, (The International Bank for Reconstruction and Development/The World Bank, 2004)
- Lanz D et al, *Evaluating Peace Mediation*, (Crisis Management Initiative, Center for Peace Mediation and Swiss Peace, 2008)
- Loquai C and Le Bay S, 'Building capacities for monitoring and evaluating decentralisation and local Governance: Experiences, challenges, perspectives', *In Brief No.19, December 2007*, (European Centre for Development Policy Management, 2007)
- Massuanganhe I J, *Participatory Monitoring & Evaluation at District and Local Level: Mechanisms, Evidences and Practices (MEP)*, (UNDP/UNCDF, 2005)
- Molund S and Schill G, *Looking Back, Moving Forward: SIDA Evaluation Manual*, (SIDA, 2007)
- Nathan L, *Operationalising the Principle of Local Ownership in Security Sector Reform*, (2006)
- Olonisakin F with Ikpe E and Badong P, *The Future of Security and Justice for the Poor: A 'Blue-Sky' Think Piece*, (King's College London, 2009)
- Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) Network on Development Cooperation, *Evaluating Development Cooperation, Summary of Key Norms and Standards*, (OECD, 2008)
- Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC), *Evaluation Feedback for Effective Learning and Accountability*, (OECD DAC, 2001)
- Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC), *Glossary of Key Terms in Evaluation and Results Based Management*, (OECD DAC, 2002)
- Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC), *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities: Working Document for Application Period*, (OECD DAC, 2008)
- Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC), *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies*, (OECD DAC, 1999)
- Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC), *Guidance for Managing Joint Evaluations*, (OECD DAC, 2006)
- Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC), *OECD DAC Handbook on Security System Reform: Supporting Security and Justice*, (OECD DAC, 2007)
- Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC), *Security System Reform and Governance: a DAC Reference Document*, (OECD DAC, 2005)

- Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC), *Whole of Government Approaches to Fragile States*, (OECD DAC, 2006)
- Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC), 'Ministerial Statement on Key Policy and Operational Commitments from the Implementation Framework for Security System Reform (IF-SSR)', signed on behalf of OECD DAC Ministers and Heads of Agency in Paris, 4 April 2007
- Osinbajo O, *Justice Indicators for State and National Use: The Lagos State Experience*, paper presented at a workshop on Indicators for Safety and Justice at the Kennedy School of Government, Harvard University, March 13–15, 2008, available at: www.hks.harvard.edu/criminaljustice/research/justice_indicators.htm, accessed 27 June 2009
- Oxford Policy Management and the IDL Group, *Evaluation of the Implementation of the Paris Declaration: Thematic Study – The applicability of the Paris Declaration in fragile and conflict-affected situations*, (OPM/IDL, 2008)
- Peake G, *Background Paper on Evaluating Security System Reform for the OECD DAC Workshop on Evaluating Security System Reform*, (University of Birmingham, 2006)
- Peake G, Hills A, and Scheye E (eds), *Managing Insecurity: Field Experiences of Security Sector Reform*, (Taylor and Francis, 2006)
- Price N and Pokharel D, 2005, 'Using Key Informant Monitoring in Safe Motherhood Programming in Nepal', *Development in Practice*, vol 15 no. 2, 2005
- Ramkumar V, *Case Study, Part 1: The Right to Know Movement in India*, (Just Associates, 2004), available at: www.justassociates.org/MKSS%20Case%20study%20Section%20I.pdf, accessed 27 June 2009
- Rauch J, *Assessment of ICTJ Liberia SSR Activities*, (ICTJ), 2007, unpublished)
- Roche C, *Impact Assessment for Development Agencies: Learning to Value Change*, (Oxfam GB, 1999)
- Ryan B, *An Evaluation of the UNDP Support to Security Sector Reform in the Republic of Albania*, (UNDP Bureau for Crisis Prevention and Recovery, 2006)
- Rynn S, *Survey of key donors and multilateral organisations on monitoring and evaluation of security sector reform: United Kingdom case study*, (Saferworld, 2009)
- Saferworld/CICS (Bradford University), *Concept paper on measurable results and indicators for peace and security*, (Saferworld/CICS, 2008, unpublished)
- Saferworld, *The role of external contractors in the provision of security and justice reform programmes: Enhancing service delivery and accountability*, (Saferworld, 2009, forthcoming)
- Sartorius R and Carver C, *Monitoring, Evaluation and Learning for Fragile States and Peacebuilding Programs: Practical Tools for Improving Program Performance and Results*, (Social Impact, 2006)
- Scheye E and McLean A, *Enhancing the Delivery of Justice and Security in Fragile States*, (OECD DAC, 2006)
- Schorer P, *Implementing Community Policing in Bosnia-Herzegovina 01.10.2007–31.12.2010: Project Document*, (St Gallen, 2007)
- Sherman J, *Survey of the United States Government's arrangements for monitoring and evaluating support to security sector reform*, (Saferworld, 2009)
- Singh J and Shah P, *Community Score Card Process: A Short Note on the General Methodology for Implementation*, (Social Development Department, World Bank, 2003)
- Smith D, *Towards a strategic framework for Peace building: Getting their act together – Overview report of the Joint Utstein Study of Peacebuilding*, (Royal Norwegian Ministry of Foreign Affairs, 2004)
- Smith J, *Dutch government policy and practice on the monitoring and evaluation of security sector reform programmes*, (Saferworld, 2009, unpublished)
- Smutylo T, *Crouching Impact, Hidden Attribution: Overcoming Threats to Learning in Development Programs*, (International Development Research Centre, 2001)
- Staats J L, Bowler S and Hiskey J T, 'Measuring Judicial Performance in Latin America' in *Latin American Politics and Society*, vol 47 no. 4, 2005
- Tautz S, *Monitoring and Evaluation of projects to combat commercial sexual exploitation of children (CSEC)*, Health, Education and Social Protection Division Convention Project 'Protection of Minors against Sexual Exploitation', (Gesellschaft für Technische Zusammenarbeit (GTZ), 2007)

- United Nations Development Programme, *Handbook on Planning, Monitoring and Evaluating for Development Results*, (UNDP, 2009)
- United Nations Development Programme, *Guidelines for Planning, Monitoring and Evaluation in Conflict Prevention and Recovery Settings*, (UNDP, 2009)
- United Nations Development Programme, *Strengthening the Rule of Law in Conflict- and Post-Conflict Situations: A Global Programme for Justice and Security 2008–2011*, (UNDP, 2008)
- United Nations Development Programme Evaluation Office, *Handbook on Monitoring and Evaluating for Results*, (UNDP, 2002)
- United Nations Development Programme Liberia, *Rule of Law Results Framework*, (UNDP, 2008, unpublished)
- United Nations Development Programme Liberia, *Strengthening the Rule of Law and Administration of Justice in Liberia, Project Document, 2008–2010* (UNDP, 2008, unpublished)
- United Nations Development Programme Kosovo, *Support to Security Sector Development (3SD) Project, Development of M&E and Impact Assessment System, Consultancy Report*, December 2007–April 2008, (UNDP 2008, unpublished)
- United Nations Development Programme Somalia, *Rule of Law Programme (ROLS) Mid-year Report, 2007*, (UNDP Somalia, 2007, unpublished)
- US Agency for International Development Center for Democracy and Governance, *Handbook of Democracy and Governance Program Indicators*, (USAID, 1998)
- US General Accounting Office, *Former Soviet Union: U.S. Rule of Law Assistance Has Limited Impact*, (US GAO, 2001)
- US Office of Border Patrol and Office of Policy and Planning, *Border Patrol Strategy: Office of Border Patrol* (US Customs and Border Protection, 2004)
- Uster H, *External evaluation of Phase 10 of the regional programme: Supporting police reforms in South Eastern Europe*, (Swiss Agency for Development and Co-operation, 2007)
- Van de Goor L et al, *The Stability Assessment Framework, Designing Integrated Responses for Security, Governance and Development*, (Dutch Ministry of Foreign Affairs/Netherlands Institute of International Relations Clingendael, 2005)
- Waglé S, Singh J and Shah P, *Citizen Report Card Surveys – A Note on the Concept and Methodology*, Social Development Note No. 91, (Social Development Department, World Bank, 2004)
- The World Bank, *Monitoring & Evaluation: Some Tools, Methods and Approaches*, (World Bank, 2004)

Saferworld works to prevent and reduce violent conflict and promote co-operative approaches to security. We work with governments, international organisations and civil society to encourage and support effective policies and practices through advocacy, research and policy development and through supporting the actions of others.

COVER PHOTO: UNMIT assists police training in Timor-Leste.

© UN PHOTO/MARTINE PERRET



Saferworld
The Grayston Centre
28 Charles Square
London N1 6HT
UK

Phone: +44 (0)20 7324 4646

Fax: +44 (0)20 7324 4647

Email: general@saferworld.org.uk

Web: www.saferworld.org.uk

Registered charity no. 1043843
A company limited by guarantee no. 3015948

ISBN 1-904833-43-8