

OECD Handbook on Security System Reform:

Supporting Security and Justice

Section 10: Monitoring and Evaluation

Acknowledgements

This tenth section of the OECD DAC's *Handbook on Security System Reform: Supporting Security and Justice* was prepared for the International Network on Conflict and Fragility (INCAF). It was researched and drafted by Duncan Hiscock and Simon Rynn of Saferworld. Special recognition is owed to the following individuals for their input: Rory Keane, previously of INCAF, and Erwin van Veen for their support in the planning and development of this chapter; Bryn Hughes, Katarina Kotoglou and Katrin von der Mosel for their detailed comments on an earlier draft of this chapter; Macha Farrant and other staff and partners of the UK's Department for International Development (DFID) in Sierra Leone and Yemen, whose collaboration on monitoring and evaluation of security and justice programmes informed later drafts; and all members of Saferworld's Expert Advisory Group on Monitoring and Evaluation, whose work for a 2008/2009 project on monitoring and evaluation helped to provide the foundations for this chapter. The OECD would also like to thank the Geneva Centre for the Democratic Control of Armed Forces, and especially Megan Bastick, for help on the gender dimension of this section. The publication was prepared by Erwin van Veen within the OECD secretariat. Stephanie Coic contributed to the graphic design.

Objectives of this section

This section of the toolkit provides practical guidance on monitoring and evaluation (M&E) of security system reform programmes. It builds on the introduction to monitoring and evaluation in Sections 4 and 8 of the OECD DAC *Handbook on Security System Reform* (OECD, 2007).

Who is the toolkit for?

The toolkit is intended primarily for international actors who provide assistance to security system reform (SSR) programmes. This includes various categories of people (note that more than one category may apply):

- Anyone responsible for designing an SSR programme.
- (International) SSR programme managers, whether in the field or headquarters.
- Programme staff with responsibility for day-to-day monitoring activities.
- People with significant experience of M&E (including evaluators), who wish to know specifically about the monitoring and evaluating security system reform.
- SSR experts who may not have much experience of M&E and thus require more basic guidance.

Much of the toolkit will also be relevant for other people who wish to learn more about how to assess the effectiveness of SSR programmes, including partner country government officials, parliamentarians, academics and civil society organisations.

What does the toolkit cover?

The toolkit is divided into six chapters:

1. What is specific about monitoring and evaluating SSR?
2. Definitions, purposes and principles for monitoring and evaluation.
3. Building monitoring and evaluation into programme design.
4. Monitoring: collection, ownership and use of monitoring information.
5. Evaluation: designing, managing and responding to evaluations.
6. Common challenges and possible solutions.

When should I refer to the toolkit?

As well as providing a general introduction to how to monitor and evaluate SSR programmes, the toolkit can be used at any point in the project cycle:

Programme design	<ul style="list-style-type: none">• Chapter 3 identifies key issues that should be addressed during programme design to ensure that the SSR programme has an appropriate M&E system
Implementation	<ul style="list-style-type: none">• Chapter 4 is about how to collect and use monitoring information. It should be read in conjunction with Chapter 3, since what is monitored depends on how the key changes to be measured have been defined during programme design• Chapter 6 provides tips on how to deal with common problems.
Evaluation	<ul style="list-style-type: none">• Chapter 5 outlines the main issues to consider before, during and after an evaluation.• Chapter 6 also gives advice on problems concerning evaluations.

Key features of this toolkit

This toolkit contains several features to make it as accessible and suitable as possible for meeting the needs of its different users:

- **Questions to consider:** At the end of some sections there is a brief list of key questions summarising what programme designers/managers should consider.
- **Checklists:** At the end of Chapters 3, 4 and 5 there are checklists summarising the main M&E issues that must be addressed during the design and implementation of SSR programmes.
- **Action points:** Some sections highlight institutional or programmatic issues that go beyond monitoring and evaluating specific SSR programmes. These are highlighted as action points to be addressed by international actors supporting SSR programmes.
- **Other sources of information:** As Chapter 1 argues, despite the specific nature of SSR, many generic development-appropriate M&E tools and approaches can be used when monitoring and evaluating SSR programmes. Where possible the toolkit indicates where readers can find more detailed information on M&E tools and methodologies from other sources.

A note on indicators

One of the most pressing issues for many programme designers and managers is finding suitable indicators for monitoring their security and justice programmes. Some readers may therefore be disappointed to find that this toolkit does not contain lists of standard indicators. The reason is that standard indicators are inadvisable because they are not context-specific, are unlikely to be locally owned, and because SSR is too complex to be measured with a simple list of indicators (Chapter 3).

Instead, the toolkit suggests that indicators should be *inspired* by other indicators that have already been used in similar programmes. With this in mind, it proposes establishing a database of indicators for SSR programmes (including for sub-sectors such as police, defence, court reforms), that would include information about:

- What changes the indicator attempts to demonstrate.
- How and where the indicator has been used before.
- The strengths and weaknesses of each indicator.
- Any unexpected or perverse incentives arising if the indicator was used to set a target.

Rather than being a standard, unchanging list of indicators, this database would be much more detailed and flexible. Database users would be able to contribute whenever they had something to offer, not only when indicators sets are being drafted. The indicators in the database would also improve over time as lessons were learned and indicators were updated and revised.

OECD DAC member states may decide to establish and contribute to such a database, but at the time of publication this is still only a suggestion.

1. What is specific about monitoring and evaluating SSR?

Those familiar with monitoring and evaluation (M&E) in other fields often ask: “What, if anything, is specific about monitoring and evaluation security system reform?”

Compared to other spheres of international assistance, SSR programmes have so far been less rigorously and less effectively monitored, and evaluations have been less frequent and less informative. The first question to explore is thus not what is specific about the monitoring and evaluation security system reform, but why do SSR programmes have such difficulty with M&E?

- **Key SSR actors may not be familiar or comfortable with monitoring and evaluation** processes that are commonplace for development institutions, or may have different expectations of how to define and measure results. These actors might include the military and police staff, lawyers, judges and intelligence officers, as well as parliamentarians and civil society organisations. Such issues may be particularly acute in partner countries where security sector officials are not familiar with such M&E processes or are sceptical or openly hostile towards certain forms of monitoring and evaluation. In such circumstances, it can be very hard to reach mutual agreement on what to measure and how.
- **Different actors still have different understandings of what SSR is about.** Despite improvements in recent years, different donor departments and multilateral agencies still have different mandates and perspectives on SSR. This extends to monitoring and evaluation, as they define success in different ways and have different approaches to collecting and analysing information. This makes it hard to establish an integrated M&E framework that is useful for all parties. Once partner country actors are included in the M&E system, this becomes even more difficult, given that they will often have a very different approach to monitoring and evaluation from those providing international assistance.
- **SSR is highly politically sensitive.** SSR addresses fundamental issues about how security and justice is provided, who exercises control, and how. It can therefore be highly politically sensitive. Because of this, the objectives of SSR programmes are often presented in vague terms and more sensitive objectives may not be openly stated. This can complicate M&E: should the M&E system measure changes according to official objectives, or according to unstated but possibly more important goals? Also, strongly embedded cultures of secrecy (both within partner country institutions and donors) are a major issue for monitoring and evaluation as they often restrict access to relevant information.
- **SSR is a complex process.** 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. This is particularly the case for integrated SSR programmes that promote “whole-of-government” approaches to reform.
- **SSR is often undertaken in fragile and conflict-affected environments.** M&E is particularly challenging in such circumstances because “standard” approaches may not be appropriate in a rapidly-changing and insecure situation. In particular, it can be very difficult to collect relevant information systematically. In such circumstances it becomes even more important to ensure that M&E processes are conflict-sensitive and do not unnecessarily put people at risk.
- **Security sector actors often need to be pushed to prioritise gender issues.** Women, men, girls and boys all have very different security needs and perceptions, and thus different expectations of security providers. However, most security sector institutions are heavily male-dominated – as are most SSR programmes. Therefore, without specific measures to prioritise gender issues, they are often ignored. This extends to monitoring and evaluation: sufficient attention must be given to gender questions to ensure that relevant information is routinely captured by monitoring

mechanisms, that evaluations consider how SSR programmes have addressed gender issues, and that M&E mechanisms are themselves implemented in a gender-sensitive manner.

This does not mean, however, that M&E tools, guidelines and systems that are used in other developmental contexts are not relevant to security system reform. Rather, they may need to be adapted for use with SSR programmes. To use an analogy, there is no need to reinvent the wheel, but it may need off-road tyres to work in more difficult environments. This toolkit gives numerous examples of standard M&E tools and processes applied to SSR contexts; where such tools need to be adapted for SSR programmes, the toolkit suggests how this could be done.

However, the effectiveness of these tools will depend primarily on the capacity and skills of those who use them. SSR programme managers who are not familiar with development-style M&E processes may first find it helpful to learn more about standard M&E tools and procedures, and this toolkit suggests further reading if required. Even if programme managers feel confident about M&E, they should be aware that other key stakeholders (various local actors, but quite possibly also other international actors) may lack skills and capacity. Building local capacity to monitor and evaluate SSR should thus be a core element of many SSR programmes, and extra training/support may also be required for international actors.

Lastly, international actors who support security system reform may also find it necessary to adapt their overarching M&E system to SSR contexts. There is often little thought given to the suitability of institutional M&E frameworks for SSR programmes, especially in fragile and conflict-affected environments. The toolkit makes no precise recommendations on how to adapt standard frameworks for SSR programmes, since it will depend on what currently exists within each agency; it does however identify key areas where greater flexibility may be required and where M&E tools may need to be adjusted.

2. Monitoring and evaluation: definitions, purposes and principles

What are monitoring and evaluation?

Taken together, monitoring and evaluation refer to a process of measuring changes in programmes or policies and assessing their impact. By providing management and key stakeholders with such information, M&E can inform current and future programme planning and delivery.

Though often grouped together as “M&E”, monitoring and evaluation are two distinct, but related, functions (Box 10.1). *Monitoring* is the continuous process of gathering and interpreting information. Its main purpose is to maintain up-to-date data to allow managers and other interested parties to track progress against stated objectives and commitments. It also includes the structures and mechanisms by which stakeholders review information and make decisions about possible changes.

Box 10.1. Definitions

Monitoring: The continual and systematic collection of data on specified indicators to show the managers and main stakeholders how a development intervention is progressing and whether objectives are being achieved in using 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 emphasise 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, which looks at 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 programme. It might involve an assessment, which is as systematic and objective as possible, of a planned, on-going or completed development intervention. (Note: evaluation in some instances involves defining appropriate standards, examining performance against those standards, assessing actual and expected results and identifying relevant lessons).

Source: Based on OECD (2002), *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Paris.

Monitoring is often confused with reporting. *Reporting* refers only to the compilation and delivery of information (for example by programme staff to managers or the public), not to the process of collecting, interpreting and storing information – all of which are crucial elements of monitoring. While good monitoring can make reporting less onerous and more informative, it is not the same thing.

Evaluation is not a process, but a specific activity. It is an assessment of the performance and impact of a programme or policy, either in its entirety or focusing on specific elements and issues. Evaluations may be undertaken during the life-time of the programme (often known as reviews), but are most frequently carried out at the end of a programme to assess its impact and to learn lessons for similar programmes in future.

Purposes and uses of monitoring and evaluation

The definitions above indicate that the main purpose of monitoring and evaluation is to measure and assess the progress of SSR programmes in order to improve their performance and achieve greater results. However, there are different ways to do this, and monitoring and evaluation can be used for at least three key purposes:

1. **A management tool to drive change.** Most international actors now subscribe to a “results-based management” approach. In simple terms, this approach claims that the best way to achieve change is to set clear objectives and targets and then to measure progress against them. This is only possible with an effective M&E system.
2. **An accountability tool.** International actors are concerned with discovering whether programmes “work”, what impact (if any) they have, and whether they provide value for money. Good monitoring and evaluation can hold SSR programmes accountable: before partner governments and local beneficiaries, before senior management and before taxpayers who contribute to development budgets, etc.
3. **A lessons learning tool.** Monitoring and evaluation can be used to provide lessons about how programmes can be run more effectively and achieve greater impact, either now or in the future.

Guiding principles for monitoring and evaluating SSR programmes

The monitoring and evaluation of SSR programmes must respect two sets of principles: (1) good practice M&E in any sector; and (2) the principles underpinning SSR programmes. Table 10.1 shows how these principles come together for monitoring and evaluating SSR. It lists seven key principles, shows the implications of each principle for how M&E systems should be structured and managed, and lists the implications for choosing what to measure. It also refers back to key principles of SSR and evaluation as stated in previous OECD DAC guidance.

Table 10.1. Guiding principles for monitoring and evaluation of security system reform

Key principle for M&E of SSR	Implications for M&E system	Implications for what to measure	References to key principles for SSR/evaluation
People-centred	M&E system should be designed to respond not only to institutions but to beneficiaries	M&E should treat individuals and communities as the ultimate beneficiaries	Service delivery to communities. Improve the delivery of professional security and justice services to local communities ^a
Local ownership and participation	M&E system should have sufficient local ownership and should strengthen local ownership over time M&E system should involve participation of a wide range of actors, including non-state beneficiaries where possible	M&E should measure extent of local ownership of and breadth/depth of participation in the SSR programme	Local ownership. Enhance local ownership of SSR processes and assistance programmes through increasing government leadership and widespread civil society participation ^a Participation of donors and recipients. Partnership with recipients and donor co-operation in aid evaluation are both essential; they are an important aspect of recipient institution-building and of aid co-ordination and may reduce administrative burdens on recipients

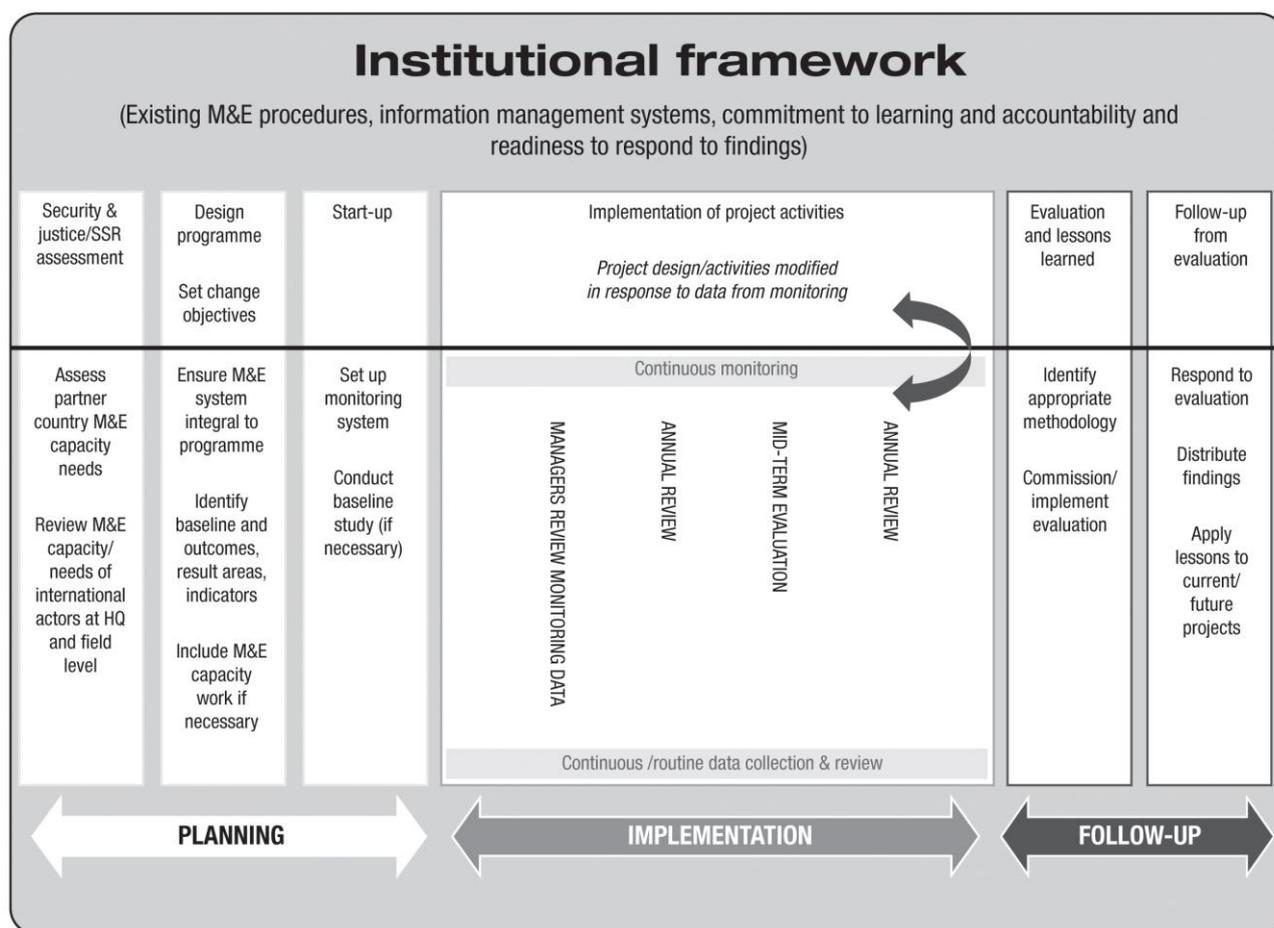
<p>Accountability, transparency and governance</p>	<p>M&E system and data should be impartial, independent, credible and transparent</p>	<p>M&E should measure whether SSR programmes and security systems demonstrate/promote good governance, are as transparent as possible, and allow and promote accountability to the public</p>	<p>Governance and accountability. Strengthen the governance and oversight of justice and security institutions to ensure that service providers are accountable, human rights are respected, and the rule of law is upheld^a</p> <p>Impartiality and independence. The evaluation process should be impartial and independent of the policy-making process, and the delivery and management of development assistance^b</p> <p>Credibility and transparency. The evaluation process must be as open as possible with the results made widely available^b</p>
<p>Integration and coherence</p>	<p>M&E system for SSR programme should use and strengthen existing national and international M&E mechanisms as far as possible</p>	<p>M&E should review whether SSR programme and support link appropriately across relevant sectors and are coherent with other reforms</p>	<p>Integration. SSR should be seen as a framework for structuring thinking about how to address diverse security challenges facing states and their populations through more integrated development and security policies^a</p> <p>System-wide. Holistic and comprehensive reform – or a system-wide approach – is SSR’s overarching objective, and may be readily pursued in some cases. But in many places, such goals may (at least initially) be politically unrealistic^a</p>
<p>Sustainability</p>	<p>International actors should ensure that partner country M&E systems are sustainable beyond the life of the programme</p>	<p>M&E should identify and promote sustainability of SSR programmes</p>	<p>Sustainability. Increase the sustainability of justice and security service delivery through developing human capacity and strengthening budgetary processes and financial management^a</p>
<p>Conflict sensitivity</p>	<p>M&E system and data should be designed and used in a conflict-sensitive manner</p>	<p>M&E should assess the conflict sensitivity of SSR programmes</p>	<p>Conflict sensitivity. As SSR programmes are often in countries affected by, or at risk of, violent conflict, it is important to understand the conflict dynamics and root causes as well as the perspectives of relevant stakeholders in order to develop effective programmes of support^a</p>
<p>Gender awareness and equality</p>	<p>M&E system should be based on full and equal participation of men and women</p>	<p>M&E should assess how far SSR programme promotes gender equity</p> <p>M&E data should recognise the different security needs of men, women, boys and girls, and be disaggregated by sex and age</p>	<p>Gender awareness and equality. Integrating gender issues into SSR means acknowledging and responding to the different security needs of men, women, boys and girls [and] ensuring the full and equal participation of men and women within security system decision-making and institutions^c</p>

Sources: a) OECD (2007), *The OECD DAC Handbook on Security Sector Reform: Supporting Security and Justice*; b) OECD (1991), *DAC Principles for Evaluation of Development Assistance*; c) OECD (2009), *The OECD DAC Handbook on Security Sector Reform: Supporting Security and Justice: Section 9: Integrating Gender Awareness and Equality* (slightly adapted to clarify the “four sets of needs”).

3. Building monitoring and evaluation into programme design

Figure 10.1 illustrates how monitoring and evaluation are relevant to each stage of an SSR programme, as well as the specific inputs required at each stage. Monitoring and evaluation for each separate programme are influenced by the institutional M&E framework at both donor and partner country level. This is represented by the grey oval that surrounds the boxes. Each box then represents a stage in the project cycle, grouped together into three phases: planning (assessment and design); implementation (start-up and project activities); and follow-up (evaluation and follow-up). The area below the line in each box indicates what needs to be done regarding M&E at each stage in the programme. This diagram can thus be used both to prompt specific activities and to ensure that all activities are properly planned and resourced during programme design.

Figure 10.1. How monitoring and evaluation fit within a programme



Designers of SSR programmes (Box 10.2) frequently do not address monitoring and evaluation in detail, and sometimes not at all. Often, monitoring and evaluation are only seriously considered once institutional requirements within the donor organisation trigger a demand for reports or annual reviews. At this late stage, a variety of problems may arise (see 'Common challenges and possible solutions' below). Many of these problems could be avoided or minimised if a systematic approach to M&E had been adopted from the start.

The best way to ensure a systematic approach to monitoring and evaluation is to assign responsibility for the programme's M&E system during design. Terms of reference for the design process should require expertise on various M&E issues, such as: building logical frameworks (logframes); choosing appropriate indicators; assessing local capacity for and attitudes to M&E; budgeting for M&E, etc. Ideally, the design

team should include a specialist who would take the lead on M&E issues. Even then, addressing M&E appropriately during programme design should be the responsibility of the whole team.

Box 10.2. Who are programme designers?

This section of the toolkit is targeted primarily at programme designers – any person or group of people who are responsible for undertaking an initial security system reform assessment and/or designing an SSR programme. Best practice, particularly with regard to local ownership, means that programmes should be designed jointly between local stakeholders (including non-government stakeholders) and those international actors that will support the programme. At a minimum, the design team should include local representation and the programme should be developed in consultation with both government and non-government stakeholders. In practice, however, SSR programmes are often designed primarily by international actors, with relatively little local input.

The guidance for programme designers in this section is intended to be useful to *any* design team, whether the design process is locally led or driven by international actors. However, in recognition of the reality that the main design team often does not include sufficient local participation, this toolkit provides advice mainly to international programme designers on building local ownership of the programme's M&E system.

Is the SSR programme's M&E system in line with the international donor's institutional M&E framework?

The M&E system comprises the people and structures that manage and implement monitoring and evaluation, and the processes and actions they use to collect and analyse information. For each SSR programme, this system needs to be suited to the programme's key stakeholder institutions, partners and beneficiaries. This M&E system does not exist in a vacuum: it must be linked to existing structures and mechanisms at both donor and country level and adapted to the context and needs of the programme and its stakeholders.

Most international actors that support SSR already have their own institutional framework which defines how and when monitoring and evaluation take place. Some agencies have very detailed M&E guidelines, often with standard requirements for regular reporting against logical frameworks (logframes), annual reviews and obligatory evaluations for all programmes above a certain financial threshold. They may also have specific staff with responsibility for monitoring and evaluation. Others have less developed M&E systems and structures.

These systems generally do not take account of the specific and challenging nature of SSR programmes. It is thus important for international agencies to find appropriate ways of adjusting their M&E frameworks for SSR programmes.

The questions in Box 10.3 are about the institution's overall M&E framework as it relates to SSR. As such, they are systemic issues that should ideally be addressed at an institutional level, rather than during the design phase of a specific SSR programme (although in practice, it may only be at this stage that these issues become apparent). Another issue that is likely to arise is when the SSR programme follows a whole-of-government approach involving several donor government departments. These departments may have different institutional M&E frameworks, and it may be difficult to establish a harmonised M&E system for the SSR programme that suits the approaches and requirements of all of these departments.

Box 10.3. Questions for donors to consider

- **What are the institution's existing M&E requirements?** Is there an evaluation policy? Are there any obligations to undertake reviews or evaluations (e.g. for programmes above a certain financial threshold)? Is a logframe obligatory? Is it necessary to define indicators and set targets and baselines? What information do all programmes need to collect? Is specific reporting required on gender dimensions?
- **How flexible are these requirements?** Can the standard M&E framework be adapted for SSR programmes and/or fragile contexts? Are there any guidelines or advice on how to undertake monitoring and evaluation in difficult environments? Can the donor institution's M&E system be adapted to the partner country's capacity and needs?
- **What structures and resources exist within the donor institution?** Is there a specific department responsible for co-ordinating monitoring and/or evaluation? Are there specific staff members who have responsibility for statistics or evaluation? Do programme staff have the skills and confidence to undertake monitoring and evaluation effectively? Does the institution have a roster of experts that have experience of both evaluation and SSR? Does the institution have expertise on addressing gender issues during monitoring and evaluation, either generally or specifically for SSR?
- **Who has the authority to do what?** Who can or must review monitoring information? Are evaluations planned and managed centrally or by field level staff? Are senior managers required to respond formally to evaluation findings?

Action points

- **Review recent SSR programmes' experiences of M&E.** It may be useful to undertake an internal review of whether recent SSR programmes supported by the donor institution faced any problems with standard M&E frameworks, and if so, how these could be overcome. This should also look at the incentives, training and support available to staff, and whether these are relevant and adequate for the monitoring and evaluation of SSR programmes.
- **Agree cross-governmental approaches to monitoring and evaluation of SSR programmes.** Major donor governments should consider establishing a cross-government working group to explore a joint M&E framework for SSR programmes.

Other sources of information

- Saferworld's "Evaluating for Security and Justice" project analysed five major donors' M&E systems from an SSR perspective. These case studies can be found online at: www.saferworld.org.uk/downloads/pubdocs/Evaluating%20for%20security%20and%20justice.pdf
- In March 2010, the UK's Department for International Development (DFID) released a set of practice papers on working effectively in conflict-affected and fragile situations, including a paper on M&E. While not specifically focused on SSR programmes, it is highly relevant for SSR in many contexts: www.dfid.gov.uk/Documents/publications1/governance/building-peaceful-states-1.pdf

Collecting information in the pre-design period

Where possible, efforts should be made to collect as much relevant information as possible before the SSR programme is designed. This could include:

- Official statistics on relevant security and justice sector institutions.
- Data on use of non or quasi-state security and justice sector providers, as relevant
- Perception surveys: public perceptions of security and justice; public attitudes towards security and justice sector providers.

- Disaggregated data on how security needs differ between different groups, for example according to sex, age, ethnicity, religious affiliation, income level, etc.
- Monitoring information collected and any evaluations conducted for similar programmes in the past.
- Government and NGO reports to international and regional human rights monitoring bodies, such as the United Nations Human Rights Committee, the Committee Against Torture, and the Committee on the Elimination of All Forms of Discrimination against Women. These may have considerable information on the performance of security sector institutions.
- Any relevant academic and policy research (*e.g.* from think tanks, universities, NGOs).

In collecting this information, attention should be paid to groups that are marginalised and may be invisible in existing data, such as displaced persons or non-citizens. Strategies can then be developed to reach out to and involve such groups in the M&E system, as relevant.

This is obviously useful for programme designers, since the more they know about the current situation, the more likely they are to design an effective SSR programme which responds to genuine needs in the right way. From an M&E perspective, it may also identify or generate information that can be used to set baselines for monitoring the programme against. This will reduce the need for separate baseline research later on.

Much of this information may be collected as a matter of course during an SSR assessment¹. However, the SSR assessment may also identify knowledge gaps. Depending on the timescale and what resources are available, it may be appropriate to commission research in this pre-design phase. This is preferable to commissioning baseline research after the programme has been designed, because any information collected cannot easily be used to inform design. Where possible, this research should be done in a way that is likely to provide useful baseline data.

Reviewing partner country M&E capacity and needs

M&E systems for SSR programmes are often designed by the donor for the donor, with little attention paid to the partner country's capacity or needs. This undermines local ownership and is unlikely to create a sustainable M&E system that can secure the programme's achievements once it has ended. Donor M&E requirements can also be overwhelming for partner countries. For example, Sierra Leone complained formally to the United Nations about the burden imposed by multiple donor visits, reporting requirements and evaluations.

Programme designers therefore need to understand the partner country's approach to monitoring and evaluation, and its capacity to effectively monitor and evaluate security and justice institutions and programmes. Box 10.4 suggests some of the key questions that should be asked. If little information is available, it may be necessary to undertake specific research. This could be done as part of the SSR assessment or in parallel to it. The assessment should look at the following institutions:

- State security and justice institutions.
- Non/quasi-state security and justice institutions (where these are widespread).
- Other government agencies that will have relevant information on security and justice, such as state gender and equality institutions and any independent human rights mechanisms.

- Existing central executive M&E functions.
- Other government agencies for whom collection and management of information is a core function, such as national auditors and national statistics offices.
- Parliament.
- Civil society organisations.

Box 10.4. Questions for exploring partner country M&E capacity

- **Is analysis of the partner country's M&E system already available?** Have any assessments of its M&E system and capacity been undertaken, formally or informally? (These may relate to the state as a whole, rather than security sector institutions).
- **Are any reforms of the partner country's M&E system already taking place?** Are there any existing initiatives to improve data collection, analysis, management of statistics, etc?
- **How does the partner country currently assess the effectiveness of programmes and learn lessons?** What systems or practices – formal or informal – are used to assess whether programmes are effective? Are these systems collecting the right information as efficiently and effectively as possible, and analysing it in useful ways? How do these systems compare to the formal M&E systems used by major international actors? What role do parliamentarians, the media and civil society play in monitoring and evaluating programmes in the security sector?
- **What relevant data and analysis are available?** What information that is relevant to the SSR programme is already available from existing sources, including previous evaluations? Which information gaps are most problematic for the planned SSR programme?
- **What is the capacity of the partner country's M&E system (as it relates to SSR)?** What systems exist to capture and analyse information? What relevant skills do staff have? How sustainable is the existing system? How much demand is there within the partner government to improve its M&E system?

If this analysis concludes that the partner country would appreciate support to improve its M&E system, a specific strand or module within the SSR programme could be included to implement this (see also Section 5 of this handbook). This might involve various activities, such as:

- Training for a range of partner country actors (ministry officials, security sector professionals, parliamentarians, civil society organisations, etc.) on why M&E is important and how to monitor and evaluate security and justice. In addition, cultural and institutional attitudes to, and understanding of, how to measure change are important; therefore it will be vital for the training to also deal with *why* it is important to invest in M&E.
- Providing resources and training to improve information collection, storage and analysis.
- Developing a joint strategy to improve how different agencies collect and share information.
- Training on independent oversight and accountability, particularly for parliamentarians and civil society organisations.
- Training on monitoring and evaluating gender issues, if systems are currently not disaggregating data and are insensitive to the different security needs and perceptions of men, women, boys and girls and how women and men participate in the security sector.

The objective should be to build a jointly-owned M&E system that is strong enough to carry out M&E independently and that will be sustainable beyond the life of the SSR programme. The next step will be to gradually transfer full management and responsibility of the M&E system from international actors to the partner country. As such, it is important that these actions are integrated into the partner country's overall M&E mechanisms and are co-ordinated with any other ongoing reforms of these mechanisms.

Agreeing *what* changes to measure

Many SSR programmes lack a coherent vision of what success looks like, *i.e.* what changes the programme hopes to achieve. Project documents usually state one or more overarching aims, presented as goals, purposes or objectives (different agencies use these terms in different ways). However, these are often expressed using vague language and do not correspond to specific changes that the programme will bring about or contribute towards. This makes it difficult to establish an appropriate M&E system, since it is difficult to decide *how* to measure progress if it is not sure *what* changes need to be measured. Even when a programme is clear about the changes it wishes to achieve, it is often not expressed in a way that is easy to monitor and evaluate.

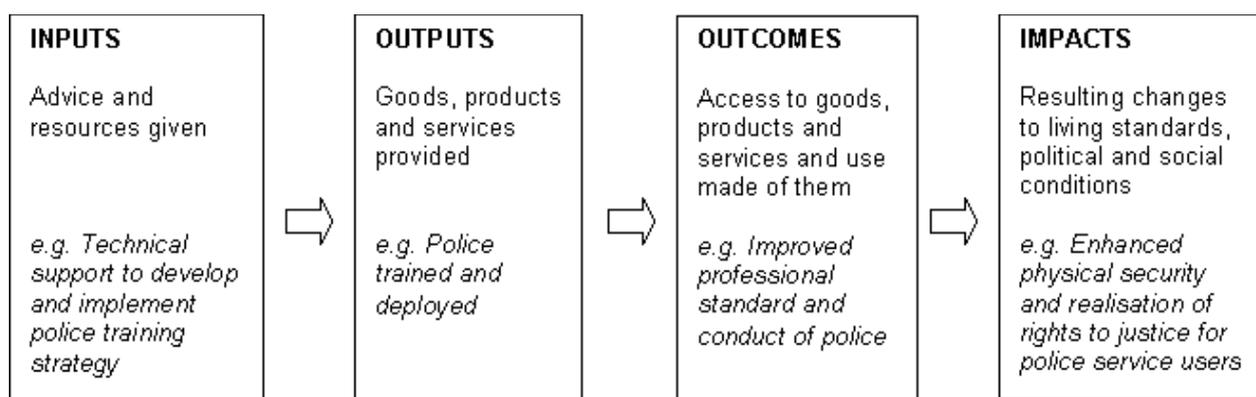
Therefore, it is highly beneficial to consider what and how to monitor (and to a lesser extent, what to evaluate) during programme design. In terms of *what* to measure, there are three key factors to consider:

1. Defining feasible and measurable impacts.
2. Identifying which “dimensions of change” the programme will address and deciding which are most important to measure.
3. Finding agreement among stakeholders on the changes the programme is seeking (as far as is possible in SSR contexts).

Defining feasible and measurable results

Impacts can occur at different levels and are defined using different terms – it is helpful to illustrate these different result levels using an example of a results chain from an SSR programme (Figure 10.2).

Figure 10.2. A results chain for an SSR programme



Inputs are the resources, such as money, goods, human resources or technology, which go into carrying out the planned activities. An example might be technical support to a Ministry of Internal Affairs to develop a new strategy to train policemen.

Outputs are the immediate results of activities, but say little about the effect of these activities. For example, if the project involves training extra police officers, the output may be that “500 police officers have been trained”. But this says nothing about the quality of that training, or the changes this has for the local population.

Outcomes are intermediate results. For example, if 500 more police officers have been trained, the outcome might be that 10 000 people in the area who previously did not have (or avoided) contact with the police now have a community police officer whom they see regularly.

Lastly, the *impact* is the long-term societal change towards which the activity is intended to contribute. In this example, the impact would be a reduction in perceptions of insecurity and a reduction in crime statistics in the area where the new police officers have been introduced. Impacts can be divided into both intended impacts (those which the programme designers did expect to happen) and unintended impacts (which happened as a result of the programme but were not foreseen at the design phase). Unintended impacts can be both positive and negative.

Note that each level in the chain does not guarantee the next. A strategy to train police officers does not mean that they will be trained. Training police officers does not automatically mean that their skills have improved or that they will be deployed properly. And deploying the police officers may not improve security if there are other reasons why the public still feel insecure or they are still unable to prevent crime in the area.

It is possible to monitor and evaluate changes at the output, outcome and impact levels. Generally, monitoring tends to focus more on output and outcome levels, while evaluations should look more at higher-order outcomes and impacts (although in practice, many evaluations also only look at outputs and outcomes and do not consider long-term impact). Programme designers need to understand how programmes can contribute to change at each level. This should ensure that there are realistic expectations of what changes a programme can achieve and measure.

Two trends occur as one moves up the results chain from inputs towards impacts. On the one hand, the changes that are sought become more significant. On the other hand, the capacity of the programme and its managers to control or influence what happens reduces (since success at one level of the results chain does not guarantee success at the next level up). At the impact level, it is rarely possible for a programme to do more than *contribute* to the overall goal (Box 10.5).

A failure to understand this has meant that many SSR programmes have had two weaknesses. Firstly, monitoring systems have focused too heavily on outputs, *i.e.* checking that the planned activities did take place. The programme tends to be labelled as a success if they did take place (for example, by confirming that a partner country's armed forces have received training and equipment, without assessing whether this has in fact improved security). Secondly, the expectation is often that SSR programmes will have impacts that are far beyond what the programme can realistically achieve. This displays little understanding that it can be difficult to attribute higher-level impacts to any one programme.

Box 10.5. Attribution versus contribution

It is only possible to attribute to an activity or programme any changes that happen if evaluators (or others) can demonstrate a direct causal link between the programme's action and the results. This is often easy to do at output level, and plausible at outcome level. It is rarely possible to do this at impact level.

For example, a desired impact might be "A 5% reduction in the number of people who report that they feel very or somewhat insecure". A cross-cutting SSR programme is launched that seeks to improve the capacity and behaviour of the armed forces and the police. After three years, there has been a 7% reduction in people saying they feel insecure. But can this be directly attributed to the programme? There may be many other reasons why this has occurred, such as the signing and maintenance of a peace agreement, an economic upswing that has made people feel more confident generally, a new Interior Minister coming to power who throws his/her weight behind police reform, a surge of vigilantism where "community patrols" deal roughly with suspected criminals, and so on. In fact, there are likely to be hundreds of major and minor factors that have influenced perceptions of insecurity.

In such circumstances, the best that can be said is that the SSR programme has probably contributed to the changes that have occurred. Demonstrating how this contribution has occurred can be a complex undertaking in itself (Chapter 5).

For this reason, security and justice programmes should be seeking to *contribute* towards positive change at the impact level, but should not expect the M&E system to *attribute* such changes to the programme.

Programme designers thus need to define and measure results at each of these three levels, but with the following considerations in mind:

- **Outputs** are usually easy to define and measure. Monitoring systems should track outputs to ensure that programme activities are proceeding as planned, but less time should be spent on defining and measuring outputs than is currently the case.
- At the **outcome** level, significant changes can be measured which are relatively easy to attribute to the performance of the programme. Programme designers should spend more effort on defining precise outcomes and establishing a monitoring system to track progress at this level.
- Regarding **impact**, realism is needed about what the programme itself can achieve, how the programme contributes to wider goals and the context in which these goals might be achieved. Methods such as impact assessments, contribution analysis and participatory evaluation may help to evaluate whether SSR programmes are in fact making a positive contribution to long-term goals such as peace, security, and democratic governance.

Identifying the most important dimensions of change

SSR programmes have many potential “dimensions of change”. This means that SSR programmes can address various different elements and areas of policy and practice relating to the security sector. The following list, which is not exhaustive, suggests some of the most common dimensions that can be improved through reform:

- Public security (incidence of crime, insecurity and injustice).
- Public perceptions of safety and security.
- The security needs of different groups (sex, age, ethnicity, religion, income, disability, etc.).
- Access to justice.
- Service delivery (*i.e.* delivery of security and justice as basic services).
- Institutional and human capacity.
- Security policy.
- Democratic governance and rule of law.
- Oversight and accountability.
- Ownership of reform processes.
- Civilian involvement in and oversight of security systems.
- Participation of women in security sector institutions.
- Non-state security and justice provision.
- Conflict sensitivity.
- Political dynamics.
- Sustainability.
- Financial and resources management.
- Cross-cutting issues, such as gender, human rights, and poverty reduction.

These dimensions are not mutually exclusive, and many programmes may seek to influence several dimensions at once. In fact, many SSR programmes affect many dimensions of change that were not the primary purpose of the programme. While it is theoretically possible to monitor and evaluate the performance of the programme along many dimensions, in practice this is rarely feasible. Programme designers must thus decide which dimensions of change are most important to measure and monitor, which changes to assess within these dimensions, and at which result level (output, outcome, impact). This requires striking a balance: if the focus is too narrow, the monitoring system may ignore or miss issues that are highly relevant to the programme; if the focus is too broad, the monitoring system will be unwieldy and unmanageable. Programme designers thus need to use their judgement to identify which changes are most important. Box 10.6 lists some key questions to ask when deciding on the most important dimensions of change.

Agreeing what to measure with key stakeholders

The Paris Declaration on Aid Effectiveness emphasises that development programmes must have local ownership². This obviously includes the design, monitoring and evaluation of SSR programmes. Paragraphs 14 and 15 of the Paris Declaration state the following:

14. Partner countries commit to:

- Exercise leadership in developing and implementing their national development strategies through broad consultative processes.
- Translate these national development strategies into prioritised results-oriented operational programmes [...]

15. Donors commit to respect partner country leadership and help strengthen their capacity to exercise it.

This means that the partner country should have ownership of the process of designing a programme, identifying what changes to measure, monitor and evaluate. The country should also have ownership of the M&E system for the programme (see also Section 4.1 below: “Establishing a monitoring system”). The Paris Declaration also emphasises that strategies and programmes should be developed through broad consultation that encourages the participation of as many stakeholders and beneficiaries as possible.

In fact, jointly designing the programme’s M&E system can not only promote local ownership of the M&E system, but also of the programme more generally. In particular, structured discussions about which indicators to monitor, and why, can stimulate much deeper thinking by local actors about what changes are really most important, whether the current programme design is the best way to achieve these changes, and if not, how things could be done better.

In practice, however, international actors may have to balance this commitment to local ownership with an awareness of the sensitive nature of SSR. Particularly in post-conflict and democratising contexts, SSR programmes are likely to challenge the interests of some key stakeholders. Therefore it may not be possible to reach agreement between all stakeholders. In some circumstances it may be necessary to develop two sets of anticipated changes: (1) those that the programme ultimately seeks to achieve; and (2) those that it is politically wise to present during the early stages of programme design and implementation. However, programme designers should be as open as possible about the changes the programme wishes to achieve, and as far as possible these changes should be agreed with the main stakeholders, including parliamentarians and civil society groups. This is in keeping with Principle 7 of the *OECD DAC Principles for Good International Engagement in Fragile States and Situations*: “Align with local priorities in different ways in different contexts³”.

Where governments demonstrate political will to foster development, but lack capacity, international actors should seek to align assistance behind government strategies...Where alignment behind government-led strategies is not possible due to politically weak governance or violent conflict, international actors should consult with a wide range of national stakeholders in the partner country.

It is also useful at this stage to define key terms in programme documents in a way that is acceptable to all key stakeholders. This will avoid later problems of stakeholders disagreeing about what “security” or “training” actually entail.

Box 10.6. Questions to consider when deciding what changes to measure

- **Does the programme clearly define the changes it wishes to achieve?** Have key changes been identified at the levels of outputs, outcomes and impacts? Have these changes been precisely defined, and could they be stated more clearly and accessibly?
- **Are these changes feasible?** Are the outcomes expected by the programme realistic? Does the programme design demonstrate an understanding that the programme can contribute to achieving long-term impacts, but cannot guarantee these impacts on its own?
- **How practical is it to measure these desired changes?** Have changes been expressed in a way that can realistically be measured (given the context, existing data and available resources)? Could programme objectives be stated in a way that makes these changes easier to monitor and evaluate? Which changes should be measured at which result level(s)?
- **Which are the most important dimensions of change to measure?** Which dimensions of change is the programme addressing directly? Which other dimensions might also be highly relevant to the programme? Has the right balance been struck between measuring the most important changes and keeping things simple?
- **Do all key stakeholders agree what changes the programme wishes to achieve?** Is it feasible to agree changes with all the main stakeholders, or does the SSR programme threaten the interests of major stakeholders? To what extent is there local ownership of the programme? Have programme designers consulted with a wide range of stakeholders and beneficiaries, including parliamentarians and civil society organisations?

Other sources of information

- The UNDP *Handbook on Planning, Monitoring and Evaluating for Development Results* includes a useful discussion of different result levels (pages 101-104): www.undp.org/evaluation/handbook/

3.5 Agreeing *how* to measure change: Indicators, targets and baselines

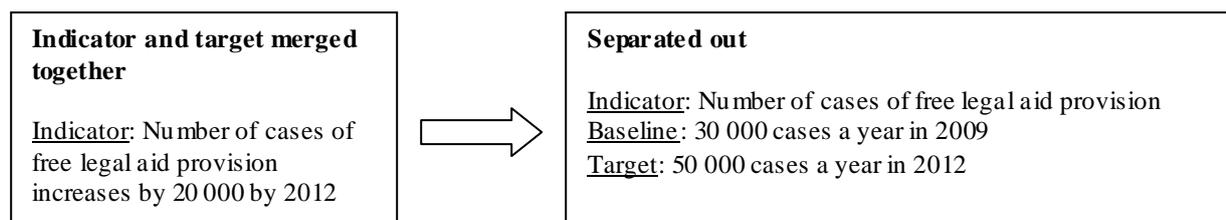
The next step after deciding *what* changes to measure is to agree *how* to measure these changes. Most major donors use logical frameworks (logframes) to guide their programme design, and these logframes specify indicators and targets. As in many other spheres of international assistance, SSR programme managers have often struggled to design and use indicators effectively. This section therefore looks primarily at how to set appropriate indicators and targets.

Before discussing indicators in detail, however, it should be noted that donor use of logframes can present certain challenges. In particular, donors may focus exclusively on their logframes at the expense of partner country M&E frameworks. Programme designers may find themselves caught between the local system and the institutional requirements of the international actors providing assistance. This is a familiar challenge for monitoring and evaluation in all spheres of aid and development, but it is particularly pronounced in SSR because of the issues and actors involved. Many security sectors have entrenched cultures of secrecy which can make it hard to access information and even to establish formal indicators and targets. As discussed above, international programme designers should seek as far as possible to reach agreement with all main stakeholders on what and how to measure, but in practice it can be necessary to define some indicators and targets without the agreement of all local stakeholders.

Indicators and targets

Indicators are often confused with targets. Targets are the change(s) that the programme wishes to achieve; indicators are pieces of information that are used to measure change and performance, and can thus indicate whether this target has been reached. This distinction is often not well understood, and it is common to see indicators presented in ways that also include the target. DFID's revised logical framework (2009)⁴ is among the first to separate out indicators and targets, and is a useful format to adopt if possible. Figure 10.3 shows the distinction between the two terms.

Figure 10.3. The importance of separating indicators and targets: an example



This distinction is important because indicators do not automatically need to be linked to specific targets, and in some cases it may not be appropriate for a programme to set a specific target during programme design. For example, this may be the case if further information needs to be gathered (such as baseline data) or if it is not sensible to set a specific target until the programme has had time to develop. In such circumstances, it is possible to identify the indicator (*i.e.* what information is required), and possibly also a “direction of travel” (*i.e.* in what way you expect the indicator to change), without setting precise targets.

Indicators can be set at the three result levels identified above: outputs, outcomes and impacts. Table 10.2 gives examples of indicators and targets for a (fictional) SSR programme that includes training for the intelligence services in counter-terrorism surveillance methods.

Table 10.2. Sample indicators and targets for a typical SSR programme

Result level	Example of change	Example indicators	Example target
Impact	Reduction in incidences of terrorist activity	Number of terrorist attacks that were not prevented	No terrorist attacks occur in 2011
	Improved perceptions of security among local population	Percentage of men, women, boys and girls saying they do not feel threatened by terrorism	75% report they do not feel threatened by 2012
Outcome	Intelligence services improve capacity to monitor actions of suspected terrorists	Ratio of individuals with suspected links to terrorist organisations under surveillance <i>versus</i> those not under surveillance (based on expert estimation)	Intelligence agency tracks activities of all individuals (100%) with suspected links to terrorist organisations
Output	Training of intelligence services in counter-terrorism surveillance methods	Number of intelligence officers trained	300 officers receive training by 2012
		Number of training workshops held	150 workshops held on 15 topics by March 2012

Perhaps the most common problem with indicators and targets is that they are often chosen by programme designers with relatively little thought, either because they are done in a hurry or because staff lack the experience and support to make more informed choices. Frequent problems arise as a result.

These include over-emphasising output indicators at the expense of outcome indicators; failing to predict when targets will create perverse incentives; choosing indicators for which information is difficult, expensive or impossible to collect; failing to signal the need for disaggregated data; and over-relying on quantitative indicators because they are easier to measure. These problems can be minimised by following the advice on good practice in using indicators and targets in Boxes 10.7 and Table 10.4.

Box 10.7. Good practice in using indicators and targets

- **Invest time in the process of choosing indicators and targets.** Reflect on all the options available to measure each result and refine targets and indicator sets over time as the programme, the understanding of partners, and the availability of information change.
- **Identify appropriate indicators at outcome level.** Ensure that the programme does not only monitor outputs and that there is sufficient emphasis on changes at outcome level.
- **Minimise perverse incentives.** Remember that “what gets measured gets done”. Choosing to measure one indicator may mean that the programme de-prioritises other important actions and results. Routine measurement of certain indicators can have perverse results. For example, measuring the time taken to process court cases can create an incentive for courts to work faster, but at the cost of due process.
- **Use multiple indicators or “baskets” of indicators to measure results at higher-level outcome and impact levels.** A balanced set of indicators that measure different aspects and that may combine quantitative and qualitative measures is more likely to cancel out biases.
- **Use a mix of quantitative and qualitative methods to measure indicators.** Quantitative indicators are often easier to collect and measure. However, quantitative indicators often do not give the full picture, and not every change that is important can easily be expressed in numerical format. Do not be afraid to use qualitative indicators where these are more appropriate.
- **Ensure that indicators and targets can reflect the needs and participation of various groups.** Consider how to measure changes that are relevant to the poor and the vulnerable, especially by disaggregating data and checking for measurement biases for/against certain groups.
- **Make your indicators gender-sensitive.** Measure whether men and women are equally participating in the programme activities, and insist on sex- and age-disaggregated data whenever feasible. Think about whether you need specific indicators to address the different security and justice needs of women, men, boys and girls (for example, looking at the types of human rights violations to which each group is most vulnerable).
- **Promote partnership, inclusion and ownership in setting and using indicators and targets.** Wherever possible, indicators and targets should be agreed jointly between the partner government and the international supporting organisations, and ideally with the participation of other local stakeholders and beneficiaries (this may include organisations that represent specific communities, such as women’s organisations, religious leaders, disability rights groups, etc).
- **Choose indicators that can be measured!** When identifying indicators, consider whether this information is already available, and if not, how easy it will be to collect it given the context and the resources that are available.
- **Test indicators.** Test indicators to make sure they are valid and appropriate measures of the result you want to achieve.
- **Keep it simple.** Try to measure what is most important and do it as simply and cheaply as possible. Wherever possible, use information that is already available and that is routinely collected. Build on existing information systems, particularly those of national institutions.

Putting these principles into practice

- Despite the importance of indicators, baselines and targets, SSR programme designers often find it difficult to set them appropriately. With this in mind, Table 10.3 gives some practical examples of how these principles can be put into action.

Given the challenges that programme designers face in setting indicators, some international actors have called for the development of standard indicator sets for SSR. Unfortunately, this is not possible for at least four reasons:

1. **Indicators must be context-specific.** Indicators must be sensitive to the local context, *i.e.* designed for, or adapted to it. This cannot be achieved by the “off-the-shelf” use of standard indicators.
2. **Standard indicators are unlikely to promote local ownership.** Above it is argued that the process of developing and choosing indicators jointly between local and international stakeholders can ensure genuine local ownership both of the programme and its M&E system. This is unlikely to be achieved when standard indicators are adopted (or worse, imposed), since there is little incentive to discuss whether they are useful and appropriate.
3. **SSR programmes are too diverse.** SSR programmes are so diverse in their design, scope, target institutions and approaches (including dimensions and theories of change), that it is simply unrealistic to develop either an overarching indicator set for all SSR programmes or even standard indicators for different types of SSR programmes.
4. **Better indicators are most needed at outcome level.** So far, attempts to develop indicator sets have largely looked at impact indicators, and there are a number of indices that assess themes such as the rule of law, fragility or good governance. However, while these provide clues about whether the programme is contributing to change at the highest level, it is at the outcome level – changes more directly attributable to the programme – where better indicators are most urgently required. The diversity and context-specific nature of SSR programmes means standard outcome indicators are unfeasible, if not impossible.

Further guidance on developing indicators

Although it is not advisable to adopt indicators unchanged from other sources, programme designers can still learn from indicators that have been used elsewhere. The best way to choose appropriate outcome indicators is to learn from the experience of similar programmes. With this in mind, it would be useful to maintain a database of outcome indicators that have been used in SSR programmes. This has been proposed to OECD DAC members as an action point for the future (Box 10.8).

In the meantime, programme designers may benefit from the following documents which provide indicator sets and/or advice on developing indicators in various sub-sectors of the security and justice system (Table 10.3⁵).

Table 10.3. Resources for indicator development in security system reform

Sub-sector	Resource
Justice	<p>Foglesong, T. et al. (2003), <i>Measuring Progress toward Safety and Justice: A Global Guide to the Design of Performance Indicators across the Justice Sector</i>, Vera Institute of Justice, New York.</p> <p>Offers guidance and standard indicators for justice programmes covering key institutions or aspects of the sector (police, prosecution and defence, judiciary, non-custodial sentencing, prisons and accountability mechanisms, non-state institutions).</p>
	<p>The American Bar Association’s Judicial Reform Index www.abanet.org/rol/publications/judicial_reform_index.shtml</p> <p>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.</p>

	<p>The World Bank Law and Justice Institutions website, http://go.worldbank.org/LRFA0Q06E1 World Bank indicators for justice reform projects and related guidance on performance evaluation for law and justice institutions.</p>
Policing	<p>Bajraktari Y. et al. (2006), <i>The PRIME System: Measuring the Success of Post-Conflict Police Reform</i>, Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey An outcome-oriented indicator framework for measuring progress in police reform.</p> <p>Bruce D. and R. Nield (2005), <i>The Police That We Want: A Handbook for Oversight of Police in South Africa</i>, Centre for the Study of Violence and Reconciliation, Johannesburg. Outcome-oriented framework for police reform in a context of societal transition.</p>
Rule of Law	<p>The UN Rule of Law website: www.unrol.org/Default.aspx Presents a set of indicators alongside a range of other tools to support implementation</p> <p>The World Bank's Worldwide Governance Indicators website: http://info.worldbank.org/governance/wgi/index.asp</p> <p>USAID's Democracy and Governance website: www.usaid.gov/our_work/democracy_and_governance/publications/</p>
Oversight and accountability	<p>Kinzelbach K. and E. Cole (eds.) (2007), <i>Monitoring and Investigating the Security Sector: Recommendations for Ombudsman Institutions to Promote and Protect Human Rights for Public Security</i>, DCAF, Geneva. This book focuses on the role of ombudsman institutions in monitoring the security sector, but is also useful more generally on how to hold security sector institutions to account.</p>
Gender and security	<p>OECD DAC (2009), "Section 9: Integrating Gender and Awareness and Equality", <i>OECD DAC Handbook on Security System Reform</i>, OECD DAC, Paris, www.oecd.org/dataoecd/4/52/42168607.pdf This section was published as an additional chapter to the SSR Handbook and provides an overview of gender issues in SSR, including how to address gender issues through M&E.</p> <p>Popovic, N. (2008), "Security Sector Reform Assessment, Monitoring and Evaluation and Gender", in M. Bastick and K. Valasek (eds.), <i>Gender and Security Sector Reform Toolkit</i>, DCAF and OSCE/ODIHR and UN_INSTRAW, Geneva, www.dcaf.ch/Publications/Publication-Detail?lng=en&id=47416 Includes discussion and examples of gender-sensitive SSR and justice reform indicators. Available in English, French, Indonesian and Arabic.</p> <p>Johannsen, A.M. (2009), "Training Resources on Security Sector Reform Assessment, Monitoring and Evaluation and Gender", in M. Bastick and K. Valasek (eds.), <i>Gender and Security Sector Reform Training Resource Package</i>, DCAF, Geneva, www.gssrtraining.ch/index.php?option=com_content&view=article&id=145&Itemid=33&lang=en Training resources on incorporating gender into assessment and M&E of SSR, including exercises on gender-sensitive indicators.</p> <p>United Nations Security Council (2010), <i>Women, Peace and Security: Report of the Secretary-General</i>, S/2010/498, UN, New York. Details a comprehensive set of indicators to measure implementation of UN Security Council Resolution 1325 on women, peace and security. Some indicators are to be used by states, others by UN agencies.</p>

Table 10.4. Using principles of good practice to improve the quality of indicators

Good practice principles	Example of weak practice	How it could be improved
Identify appropriate indicators at outcome level	A project aims to increase access to justice through a media campaign. It states an outcome (e.g. “Greater participation in law implementation and legal and judicial reforms”), but indicators are all at output level (number of articles published during media campaign, publication of Supreme Court decisions, number of street dramas on human rights organised, etc.)	<ol style="list-style-type: none"> 1) Revise outcome to clarify outcome of media campaign, e.g. increased knowledge of how to access justice services among target community; greater confidence in justice services 2) Establish baseline: study/estimate existing knowledge of and confidence in justice services 3) Then set indicators, baselines, and targets, e.g.: <ul style="list-style-type: none"> • <u>Indicator</u>: % of men, women, boys and girls in the target community who say they would turn to the legal system to get justice • <u>Baseline</u>: 20% in 2009 • <u>Target</u>: 45% by 2012
Avoid perverse incentives	Aware that not enough criminals are brought to justice, programme designers set a target to increase the number or percentage of arrested suspects who are sentenced. However, to meet this target, more people are prosecuted for minor crimes that previously would only have received a caution, while the police and the courts actually spend less time on serious crimes for which convictions may be difficult to secure.	<p>Consider the ultimate outcome to be achieved by the measure: the aim is not simply to sentence more people, but to remove serious criminals from the streets, to reduce the perception that criminals can operate with impunity, and to boost public confidence in the criminal justice system. Goal/purpose statements and impact indicators should focus on this, not on sentencing patterns.</p> <p>Perverse incentives can sometimes be avoided by using more precise language, e.g. the number of people sentenced for <i>serious</i> crime. However, even then this may create incentives for the courts to sentence more people without following due process. Whenever targets are set, programme designers need to think carefully about the potential for the target to be met in a way that is not in line with the programme’s intended overall goal and impact, and establish safeguards if possible (e.g. by evaluating random cases to see whether due process was followed).</p> <p>Another way to reduce the risk of perverse incentives is to set targets that are linked to a basket of indicators (see below) rather than only one particular change.</p> <p>It may not be appropriate to set a target at all. Conviction rates could be used as an indicator – to provide information for analysis – without the programme setting a specific target which might give managers in the criminal justice system an incentive to behave in perverse ways.</p>
Use baskets of indicators	A police reform programme uses the reported crime rate as an indicator of police effectiveness. But what does an increase in reported crimes actually mean? It could mean that there is more crime overall. But it could also mean that the public have more trust in the police and are therefore more willing to	<p>Select indicator sets that together provide a detailed picture. Use a mix of quantitative and qualitative indicators as appropriate. Use indicators from different sources.</p> <p>For example, if the goal is to increase public willingness to report crime to the police, indicators could be set for crime reported to the police and for crime victimisation rates as reported in an independent survey. If baseline figures show that most crimes are not reported to the police (e.g. 90% of property crime goes unreported), but that this figure has fallen to 70% after three years, this</p>

	report crime than in the past.	would suggest that people are more willing to report crime. This could be further explored, <i>e.g.</i> by holding a set of focus group discussions with crime-affected communities at the start of the programme and then at yearly intervals.
Use both quantitative and qualitative indicators	Continuing the example above, a police reform programme only uses reported crime and crime victimisation statistics to measure the programme's effectiveness in encouraging the public to report crimes to the police. However, these statistics only give a shallow picture of how attitudes are changing and may fail to spot significant issues during implementation.	Indicators do not have to be numerical. It is legitimate to set an indicator such as "Public willingness to report crime to the police", and then use a mixture of quantitative and qualitative methods to measure it. The two combined indicators above would be one way of doing this. Another quantifiable way would be to run public opinion surveys, for example asking whether people would be prepared to report personal property thefts to the police. This could be compared with the ratio of reported to unreported crime. It could also be tracked over time to see whether there is an increase in the number of people who say they would be prepared to report such crimes to the police. Other methods are non-quantifiable, such as focus group discussions. It is not possible to use focus groups to provide a numerical assessment of how attitudes are changing, but they are immensely useful as a way of exploring attitudes in much greater detail. They can highlight important issues that numerical indicators are unlikely to discover. For example, many focus group participants may say, "I would report this crime to the police, but the police station is not open when I finish work". If programme managers receive this information as part of their monitoring system, they can explore this issue further and if necessary adapt their programme accordingly.
Make your indicators gender-sensitive	Two key indicators are chosen for the success of an SSR programme: the percentage of people reporting that they feel safe and secure; and the percentage of people who are satisfied with the performance of key security providers. No gender-specific indicators are required, and thus the programme fails even to recognise that domestic and sexual violence are severe problems and that state security providers refuse to take such issues seriously.	Gender-sensitive information should be collected from the assessment stage of the project to identify any major differences in how security and justice are perceived by men, women, boys and girls – and programming should proceed from this basis. This requires at a minimum that all data such as statistics and perception surveys can be disaggregated by sex. However, it is not enough simply to disaggregate existing data – new attempts to collect relevant information are also needed. This might include, for example, asking specific questions on domestic and sexual violence in household surveys (perhaps indirect questions if these are more likely to elicit useful data). Another important step would be to carry out focus group discussions specifically with women (and with a female moderator) so that participants feel able to discuss such questions more freely than in a mixed group. Those responsible for planning M&E should also consider collecting qualitative and quantitative information from other relevant sources, such as women's shelters and organisations, human rights organisations, and government ministries responsible for gender/women and children etc. In this example, one option is to state clearly that the indicator requires sex-disaggregated information – this would provide more detail on whether, for example, women and girls also feel secure and whether they are happy with the performance of security providers. Another option might be to add indicators that are more specifically related to sexual and gender-based violence, such as indicators on the (estimated) number of women who report different types of sexual violence (<i>e.g.</i> domestic violence, rape, sexual harassment) to the police, and the percentage of those who do who are satisfied with security providers' responses.
Reflect the needs and	A programme to increase access to	Ensure that information provided for this indicator can be disaggregated into relevant categories, for

<p>participation of different groups</p>	<p>alternative dispute resolution (ADR) sets two indicators: “Percentage of those who have suffered a dispute in the past two years who have used ADR to resolve the dispute” and “Percentage of people who say they would be able to use ADR to resolve disputes if they needed to”. This says nothing about specific groups.</p>	<p>example: sex, age, ethnic group, income level, religion, rural/urban location, level of education, etc. This may require instructing statisticians and sociological research companies to collect this data and to disaggregate it as required. For example, disaggregating the first indicator may reveal that while 53% of those who have suffered a dispute have used ADR, only 15% are women, or that 60% of those who have used ADR are in urban locations despite the fact that 80% of the population lives in rural areas.</p> <p>Information on different groups can also be collected qualitatively, for example by holding focus groups with specific categories of people and exploring their opinions in detail.</p>
<p>Use indicators that are possible to measure</p>	<p>In order to make the police more representative of the population it serves, a reform programme sets a target to bring the percentage of police officers from each ethnic group in line with the national average. However, no census has been carried out for 20 years and the Ministry of Internal Affairs does not have reliable data about how many police officers it has, let alone how many are from which ethnic group.</p>	<p>The programme cannot conduct a national census! Is it possible to estimate the ethnic breakdown of the population in a way that is not politically controversial? Equally, is there a programme component for overhauling the personnel management system and collecting data on ethnicity?</p> <p>If the answer to these questions is no, it is unwise to set this target as it will not be possible to measure the indicator. It would be better to find an indicator and target that achieves similar goals but can be measured more easily. For example, a target of “25% of new recruits, including at more senior levels, are from ethnic group xxx” is less far-reaching but more likely to be achieved and much easier to measure. The programme can then ensure that the ministry records the ethnic background of all new recruits.</p>

Baselines

It is not possible to get a full picture of how indicators are changing over the course of the programme unless it is clear what the situation was at the start. This is referred to as a “baseline”. Wherever possible, programme designers should seek to use existing information for their baselines, not only because this is cheaper, but also because repeated use and refinement of existing material is more likely to be sustainable than undertaking stand-alone studies for each programme. As Section 3.2 suggests, potential baseline data can often be collected in the pre-design phase, even though at that point it will not be certain which indicators will be chosen.

However, establishing baselines can be problematic for SSR programmes, since they often deal with difficult contexts where little reliable information is available. This is exacerbated by cultures of secrecy within the security sector. In such circumstances, programme designers have three options.

1. They may propose revising the indicator or choosing a new indicator so that it is based on information that is already available. In some countries where government statistics may be unavailable or unreliable there may be competent civil society organisations or academic groups who are collecting and analysing similar information. For example, a government is unable to break down crime statistics according to ethnicity, but a prominent human rights organisation is recording incidences of crime against minority groups.
2. Leave the baseline (and target) empty during programme design, and then collect more information during the programme’s start-up phase. If information is required for various indicators, it is sometimes useful to conduct a separate baseline study. The design of this baseline should be strongly tailored to suit the M&E system so that it remains a useful point of comparison throughout the programme. In practice, however, such baseline assessments are rare. This is often because no resources were allocated to this during programme design, but it may also be because of the urgency of launching an SSR programme and a perception that undertaking detailed research requires too much time and resources and that is unlikely to provide useful information in a rapidly changing environment. Programme designers thus need to consider what form of baseline it is possible to establish, whether a stand-alone study is the best way to do this, and what much time and resources should be allocated to this.
3. Rely on the “analytical baseline” that is generated during the process of programme design. Programme designers usually refer to and/or generate considerable information about the conditions that exist when the programme is being planned, even if this is not systematically captured. Such information has its weaknesses. It is often based primarily on the perceptions and experience of the programme designers and is usually mostly qualitative. Nonetheless, in difficult contexts this analytical baseline may be the most that it is possible to gather, and it may be good enough for the needs of the programme.

Box 10.8. Questions to consider when deciding how to measure change

- **Have indicators been identified at each level (output/outcome/impact)?** Have appropriate indicators been chosen at each level – *i.e.* are they in line with what the programme can realistically achieve or influence at each result level? Has the programme struck the right balance between the indicators to monitor at each level?
- **Have targets been chosen appropriately?** Is there a clear distinction between targets and indicators? Is it necessary for every indicator to be attached to a precise target? What is the risk that this target may give incentives for perverse behaviour?
- **Have indicators been chosen according to the good practice guidelines in Table 3.2?** Have multiple indicators been used to gain a broader picture? Has a good balance of qualitative and quantitative indicators been chosen? Are indicators gender-sensitive? Will it be possible to show how the programme affects different groups and categories of people?
- **Has the programme established baseline data?** If not, has the necessary information been identified, and have resources been allocated to collect this data?
- **Are the indicators measurable?** Can the anticipated changes be measured to a “good enough” level, *i.e.* will it be feasible to monitor these changes using the available resources in the local context, and to evaluate results later on?

Action point

- **Compile a list/database of SSR outcome indicators.** International actors should consider compiling a list or database of outcome-level indicators that have been used in previous SSR programmes. This database should include: analysis (from programme staff) of the strengths and weaknesses of each indicator; any perverse incentives they caused; any challenges faced when collecting information for this indicator; which indicators work well in combination; and how to use each indicator most effectively.

Other sources of information

- *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.* This report by the Harvard Kennedy School of Government provides a review of recent experience of using indicator sets for security and justice reforms in developing countries. Available at: www.hks.harvard.edu/var/ezp_site/storage/fckeditor/file/pdfs/centers-programs/programs/criminal-justice/justice_indicators_workshop_2008.pdf.

Refining programme design: reviewing the programme’s theory of change

The underlying logic behind a programme’s intended results (outcomes and impacts) and the means employed to achieve these (inputs, activities and outputs) are variously referred to as the ‘theory of change’, the “programme logic” or the “intervention logic” (Table 10.5). Theories of change are seldom made explicit during the planning or implementation stages of a project, and are usually only explored at the evaluation stage, if at all. This is a missed opportunity. In many cases, SSR programmes are designed more in line with the ideas and models of international specialists than local needs and realities. Consciously reviewing the programme’s theory of change during the design phase can help to refine programme design, since it often reveals untested assumptions and/or a lack of creative thinking, and encourages more context-specific approaches.

Table 10.5. Theories of change found within SSR programmes⁶

Theory of change	Example of methods	Comments
Capacity building: National security and justice institutions will do their job better if they have greater skills and resources	Substantial focus on providing training and equipment, often in the absence of institutional reform, safeguards or reference to accountability and governance.	Hitherto the most commonly employed theory underpinning donor SSR work. Known limitations and sometimes unintended negative results.
Professionalisation: Improving standards of professional conduct, self-image and <i>esprit de corps</i> will lead to better behaviour by staff and enhance service delivery through a sense of pride and public service that in time breeds accountability.	Strong focus on codes, training, conditions of service, exchanges and learning from outsiders; surprisingly little on monitoring actual behaviour through methods such as observation, surveys and user groups.	Favoured by both progressive senior personnel in unreformed institutions and foreign advisers. Has strong personal appeal for staff interested in career progression and meritocracy. Often informed, for better or worse, by specific views on the suitability of systems and principles applied more consistently elsewhere.
Good governance and accountability: Introducing mechanisms and structures for internal or external accountability over institutions will enhance accountability through the prospect of negative sanctions against rule-breakers. In time, effective accountability mechanisms can channel public or civil society demand for further reform or better delivery.	Attention to the existence, setup and functioning of internal and external accountability mechanisms; to their handling of exemplary or high-profile cases; to the legal framework and application of the law in these areas; and to the political independence of security and justice institutions at operational level. May also pay attention to higher-level security sector policy development as a contribution to improved governance.	An approach promoted in the OECD SSR Handbook but employed only by a minority of programmes. Emphasis on external mechanisms such as parliaments, civil society and ombudsperson's offices particularly lacking in these programmes. Often less attractive to partner governments.
Institutional reform/restructuring: Given the right institutional set-up (laws, procedures, management and recruitment systems and tools), security and justice institutions and those working within them will begin to function more rationally and effectively.	Strong focus on re-designing institutional structures and their operating procedures in line with perceived models of best practice.	Particularly common for creating operational separation between MOD/armed forces or interior ministry/police. May achieve little, however, if the institutional set-up is reformed without changing the attitudes and behaviour of the politicians and managers who run them and the staff who work for them.
Enhance national security: The main threats to security and justice are from external/international factors such as regional conflicts, terrorism, weapons proliferation, etc. Strengthening the capacity of national actors to combat these threats will increase security for everyone.	Significant resources spent on training and equipment for national security forces, particularly armed forces, policing and intelligence services. Often aims only to strengthen capacity (see above) with little thought given to whether this capacity will be used effectively and accountably.	The most prevalent form of donor assistance to security actors in recent years, though often not considered by donors as "SSR" or related to SSR principles and standards. Strong donor emphasis on counter-terrorism may distort national perceptions of security or redirect resources away from other pressing local needs.
Enhance community security: Many security and justice concerns are essentially local in nature. Addressing security at the community level will thus have the biggest impact for improving	Training of police in community policing approach, often with resources for new equipment and facilities. Community safety plans, based on local perception surveys, that encourage local state and non-	Community policing has been popular for several years, while a broader approach to community security is more recent. Relation to and impact on state/national security and conflict dynamics

security and justice and can reduce the risk of violent conflict.	governmental actors to work together to improve security.	sometimes unclear.
Comprehensive approach: Different elements of the security system influence each other and reform in one sector may fail if not supported by reform in an adjacent sector; therefore a comprehensive approach is needed which links reforms together across the security system.	Emphasises a system-wide approach to reform, with an overarching SSR strategy to which specific reforms of sectors and sub-sectors are subordinate. Often aims to build capacity of an overarching body such as a national security council to promote joined-up thinking.	OECD DAC policy emphasises the importance of integrated and whole-of-government approaches to SSR. However, a genuinely comprehensive approach would be highly complex and it is unclear how far or how effectively such approaches have been employed in practice.
Enhance service delivery: Security and justice are public goods like health and education. The public are service users who are entitled to security and justice. National security and justice institutions should be reformed with a focus on those issues and mechanisms most likely to meet demand in the near term.	Opinion surveys, workshops, etc. used to identify key public demands and expectations of security and justice; rights-based approaches that emphasise links to constitutional and legislative provisions; specific interventions for reform, support, capacity building then designed on this basis with public participation and support.	This approach is relatively new. Raises various dilemmas about how to treat security and justice as basic services in fragile and conflict-affected countries.

In addition to the above theories, others less frequently or not hitherto applied within SSR circles can be identified and may be adapted from other fields. These might include, amongst others, changing public attitudes, tackling societal violence, working with political elites or within peace agreements⁷. Note also that while this table has some similarities with the dimensions of change (Section 3.4), the emphasis is different. Identifying the dimensions of change is about clarifying the main areas in which the programme seeks to achieve change, and thus what it is most important for the programme to measure. By contrast, reviewing the *theory* of change is about whether the approach taken is appropriate given the context and thus whether the planned activities are likely to lead to results.

Box 10.9. Questions to consider in refining the programme design

- **What is/are the theory or theories of change that underpin the programme?** Has this theory of change been consciously chosen, or is the programme design based on assumptions of what will work and/or similar programmes in other contexts?
- **Is this theory of change appropriate?** Does this approach suit the context of this programme? Are there other ways of achieving the same goals that might be more appropriate or successful in this context? If the programme displays elements of several different theories of change, are they complementary or contradictory?

Other sources of information

- Sida (2007), *Looking Back, Moving Forward*, Sida, Stockholm. This evaluation manual contains a short but useful section on intervention logic (pages 66-68), available at: www.sida.se/Svenska/Om-oss/Publikationer/Visa-publikation/?iframeSrc=http://www2.sida.se/sida/jsp/sida.jsp%3Fd=118%26a=3148&language=en_US&searchWords=looking%20back

Allocating time and resources for monitoring, reviews and evaluations

Programme designers must ensure that plans allocate appropriate resources (including time) to monitoring and evaluation: there is no point in establishing indicators, targets and baselines if the programme does not have the money, staff and time to monitor these indicators and evaluate performance.

One of the most frequent problems is that the programme has no specific budget for monitoring and evaluation. Precise costs will depend on the complexity of the programme and which indicators have been chosen. Where programmes have identified the need to strengthen the M&E capacity and skills of local or international actors and institutions (Sections 3.1 and 3.2), this may require additional resources. In practice, programme designers may not be able to accurately predict the costs of all M&E activities, so it may be easier to allocate 5-10% of the overall budget specifically for M&E and to create a separate budget line for this. This would then cover all costs, including training in M&E for relevant staff members; collecting and analysing baseline data; establishing and maintaining data collection and storage systems; and consultant fees for independent evaluations. The budget must also include staff time to manage, analyse and review monitoring information and to prepare, manage and respond to evaluations.

As well as a budget, M&E also requires human resources. There should be institutional incentives to undertake M&E, such as a clear commitment to use monitoring information and evaluations to improve how the programme is delivered. It may also help to include responsibility for M&E in staff job descriptions.

It is also useful to establish when key monitoring and evaluation activities will take place. Some information collection requires considerable forward planning (e.g. a survey of public perceptions of security and justice). Approximate dates can be set for when management structures will review monitoring information and at what periods it will be possible to revise project strategies and implementation plans. It may also be possible to set dates for interim or end-of-project evaluations and impact assessments.

Box 10.10. Questions to consider when allocating time and resources

- **What resources has the programme allocated to M&E?** Is there a specific budget for M&E activities? Will it be possible to monitor all indicators with these resources? Do resources need to be allocated for training? Does the budget also allocate staff time to manage and respond to monitoring information and evaluation findings?
- **Do the programme plans set times for key M&E activities?** Have times been set for regular management meetings to review monitoring information? Have approximate dates been set for interim and/or end-of-project evaluations?

Box 10.11. Checklist: Key issues during programme design

- Does the programme meet all institutional requirements for M&E?
- Do the institutions and staff who will work with this programme (both partner country and international actors) have the skills and capacity to establish and manage a monitoring system? If not, is support available?
- Are the key changes anticipated by the programme:
 - Clearly defined?
 - Feasible to achieve?
 - Linked to relevant indicators and (where appropriate) targets?
 - Possible to measure given the context and available resources?
- As far as possible, have all key stakeholders reached agreement on what changes the programme will achieve and how?
- Will the programme monitor the right balance of output, outcome and impact indicators?
- Has the programme identified all necessary baseline data, or allocated resources to collect this information during the start-up phase?
- Is the programme's theory of change appropriate to the context?
- Does the programme allocate sufficient resources (money, human resources, time) to M&E?

4. Monitoring

Monitoring is primarily an internal management function which measures how a programme is performing. This allows managers and other interested parties to assess whether the programme is achieving the anticipated results, and to make corrections to programme design and implementation if necessary.

Monitoring is often confused with reporting. Many international donors require all major programmes to report at certain points of the programme (or at regular intervals, *e.g.* annual reviews). While reporting is certainly a monitoring activity, it is only part of monitoring. Monitoring also involves collecting, storing and analysing relevant information – programme managers often find it hard to report effectively because they only consider what information to collect and how to collect it when they are obliged to complete a report. Another challenge is that reporting is mostly driven by donors' institutional management requirements. Unfortunately, the reporting process is rarely designed and undertaken with the needs of the partner country in mind. A good monitoring system should be owned by both the partner country and the international actors providing support, and is just as concerned with collecting and analysing information as it is with reporting.

Research has shown that the majority of SSR programmes have weak monitoring systems. Why do SSR programmes find monitoring so hard? The answer is probably a combination of factors, some of which are about who is managing the programme, and some of which are about the nature of SSR programmes. As noted in Chapter 1, many security sector actors do not have much experience in M&E, and without additional training and support, this can be reflected in poor programme design (Chapter 3), a lack of awareness of the importance of monitoring, and/or a lack of relevant skills and capacity among staff and institutions both at partner country and international level.

However, SSR programmes also find it difficult because of the issues they address, and the contexts in which they work. Cultures of secrecy are a major obstacle to monitoring for most SSR programmes, since good monitoring depends on good information. It is usually much harder to gather or access information about security and the work of security sector actors than, for example, information about agricultural production. The other point is that SSR is often undertaken in fragile and conflict-affected countries. In such circumstances, not only is there likely to be limited infrastructure for data collection, but the situation can change rapidly and unpredictably. This makes monitoring even more important, but even harder to do well.

With this in mind, monitoring actually requires *greater* preparation and commitment than in other spheres. Programme managers should therefore read this chapter carefully and ensure that they have answered all the checklist questions in Box 10.15. Monitoring can essentially be broken down into two main activities: collecting information in a useful format; and interpreting this information as a management tool. This chapter is thus divided into two parts:

1. Establishing a monitoring system
2. Using monitoring information

Establishing a monitoring system

In order to ensure that the necessary information is collected, programme managers need to establish a monitoring system that will specify *what* information to collect, *how* to collect it, *when and how often* to collect it, and *who* should collect it. This system must balance the needs of the donor institution with the partner country's needs and capacities, and should be based on a principle of *joint ownership*. In fact, this is

unlikely to be a one-off activity, but rather an iterative process that gradually improves the quality of monitoring as the programme progresses. There is always space to improve the monitoring system at any point in the programme cycle.

Deciding what information to collect and how

A well-designed SSR programme will have clearly stated indicators (Chapter 3). This will help clarify the information that needs to be collected. Even in such cases, however, programme designers are unlikely to go into precise detail about the information to collect and how. Therefore, programme managers need to answer this question as early as possible, preferably in the programme start-up phase.

The questions of what information to collect and how are closely related, since there is no point in deciding to collect information which would be extremely difficult, costly, or dangerous to gather. Given the context in which most SSR programmes operate, programme managers normally have to be very strategic in deciding which information is most important to collect and finding the easiest way of doing so. The emphasis should be on deciding what information is “good enough” to allow managers and other interested parties to assess how the programme is performing. Box 10.12 suggests a number of questions that programme managers should consider when deciding what information to collect (Box 10.7 on ‘Good practice in using indicators and targets’ in Section 3.5 may also be useful). Some of the issues identified in the box are particularly significant for SSR programmes and thus deserve further discussion.

Firstly, availability of information is the biggest challenge for many SSR programmes. Programme managers should be aware that even when information is theoretically available, it may be difficult to access or to compile in a useful form. For example, in some countries, programme managers may find it very hard to gather official crime statistics from the relevant government agencies, particularly if they require the data to be disaggregated. Programme staff may end up spending significant time and effort chasing such information. Programme managers therefore have to be realistic about whether “available” information is in fact accessible. International actors may also benefit from discussing issues such as monitoring, transparency and accountability with partner country officials early on in the programme, so that there is time to build a shared understanding of why such information should be collected or made available. For example, partner country officials may feel threatened or offended if programme managers insist that government data needs to be compared with other sources of information.

Box 10.12. Questions to ask when deciding whether monitoring information is “good enough”

- **What information is essential to collect?** What is the bare minimum of information that *must* be collected to allow programme managers to assess performance? Which indicators are essential to measure (usually linked to the most important targets)? Does new information need to be collected to provide baseline data?
- **What information is already available?** Is this information routinely collected? Is it easy to get access to and to compile in a useful format? If it is not already available:
 - How can this information be collected most effectively?
 - How much will it cost to collect this information, and what staff and time will be required to do so?
 - Can the process of collecting information and/or the collected information be shared with other programmes and institutions in order to increase the benefit of collecting this information and avoid duplication?
- **How reliable is this (source of) information?** How reliable is each source? What does this information represent, and how accurately? How many different sources of information are being used, and are they enough to triangulate information and cancel out any biases?
- **Is collecting or using this information likely to be sensitive?** Does collecting this information pose any risks either to those doing the research or to any vulnerable communities? Is this information politically sensitive, or likely to affect conflict dynamics? If so, what are the implications of collecting and using this information, and how can any risks be mitigated?
- **Do these sources provide the right type of information (or right mix of types of information) for your needs?** Is this information quantitative (*e.g.* official statistics, public opinion surveys, number of people trained) or qualitative (*e.g.* key informant interviews, focus group discussions, follow-up interviews with training participants)? What is most useful for monitoring purposes (quantitative information is generally more representative for larger groups of people and is sometimes all that is required at output level; qualitative information is less generalisable but often gives deeper understanding of complex issues)? Has the right mix of qualitative and quantitative information been found?
- **What can this information tell you about different groups of people?** If this is quantitative information, can it be disaggregated into different categories (*e.g.* sex, age, location, ethnicity, religion, etc.)? Can qualitative sources, such as interviews or focus groups, provide more detailed information about the experiences of certain groups?
- **Is this source of information likely to identify any unexpected effects of the programme?** Does this information focus exclusively on the planned results of the programme, or is it also able to highlight any unexpected changes that may have occurred because of the programme?
- **What information is being collected at each result level (output/outcome/impact)?** Is this information appropriate at this result level? Overall, is the monitoring system capturing the right balance of information at each level?
- **What are the weaknesses of each piece of information?** No information is perfect, and most SSR programmes must make do with “good enough” information. What are the weaknesses of this information? Is it possible to formally note the weaknesses of this (source of) information to minimise the risk that data will be misinterpreted?

Where several donors are supporting reforms in the security sector, programme managers should also assess whether it is possible to collect information jointly, in order to avoid duplication and reduce the strain on partner country systems. This is in line with point 46 of the Paris Declaration on Aid Effectiveness, which commits donors to harmonise their monitoring and reporting requirements.

Secondly, SSR touches on some very sensitive issues. Therefore, both collecting and using information for monitoring purposes may come with certain risks. For example, it may not be possible to collect information in one part of the country because it is too insecure for researchers to travel there. Asking people about sensitive issues may put them in danger (for example if a powerful local individual suspects that respondents are telling researchers the “wrong” thing). Even if there is no such danger, people may be afraid to answer honestly, so the information they give may not be fully reliable. There are also various risks associated with how sensitive information is interpreted and presented (see Section 5.2 below).

Thirdly, it is well recognised that an SSR programme in one part of the security system is often affected by what is (or is not) happening in another part of the security system. For example, a police reform programme may struggle to boost public confidence in the criminal justice system because there have been no corresponding reforms of the justice sector, or because the reform has not addressed other security system institutions with a policing function (such as the armed forces in some countries). While it is unrealistic to expect the monitoring system to study all of these linkages, it should avoid focusing too narrowly only on the direct results that the programme aims to achieve. The monitoring system should make it possible to include information on any unexpected results of the programme.

Fourthly, programme managers should consider from the start how best to ensure that women and marginalised groups are able to participate fully and equally in the monitoring process. This goes beyond seeking representative interviewees; it requires the researchers involved in the monitoring process to ensure that everyone has the opportunity to participate. This might include, for example, scheduling focus groups and interviews at times and locations where both women and men can attend; providing childcare and transportation; holding separate focus groups for certain target groups so that they can express their views more freely; and using communication tools that are accessible for non-literate groups.

Lastly, it is worth re-emphasising that programme managers need to base their judgments about what is “good enough” on a realistic appraisal of what is available or possible to collect. While accurately compiled statistics or a detailed public perception survey may seem more objective and comprehensive, in many cases this is simply not possible. There are various monitoring tools that can provide at least some useful information, and it may be better to settle on such methods (while recognising their weaknesses) rather than demanding information that cannot realistically be collected. These tools include:

- Progress reports submitted by individuals or institutions
- Key stakeholder workshops
- Expert interviews
- User group meetings
- Suggestion boxes
- Go-and-see visits and spot checks (unannounced visits)
- Exchange of views during formal meetings (*e.g.* technical management committee meetings, strategic management/steering committee meetings, donor coordination meetings).

Roles and responsibilities: who does what, when and how often?

Once it is clear what information needs to be collected, responsibility for information collection must then be allocated to particular institutions and individuals. This includes any external contractors who are

involved in delivering all or part of the SSR programme. It may be helpful to agree a monitoring plan which identifies what information will be collected, when, how and by whom (Table 10.6).

Table 10.6. Sample monitoring plan

Indicator	Info sources	Collection method	Schedule	Responsibility	Required partners	Dissemination plan
Number of male and female police personnel relative to population	Official data from interior ministry	Request and review ministry data	Annual	Programme manager	Interior ministry	Quarterly stakeholders meeting
	Random spot-checks at police stations	Field-monitor visit	Biannual		National civil society organisation network on public safety	
Number of male and female people reporting that they are worried about being victim of crime	Survey of public perceptions of security International crime victimisation survey	Household survey undertaken by independent sociological research company	Annual	Programme manager	Independent sociological research company	Annual review panel Findings of survey published and advertised

Creating appropriate ownership of the monitoring system

How monitoring is carried out can be just as important as *what* is done. Monitoring must be informed by and respectful of key principles such as local ownership, gender equality and conflict sensitivity. Furthermore, without local ownership and participation, partner country actors are less likely to feel ownership of monitoring data and are thus less likely to act on them. There is thus a need to create an appropriate ownership structure for the monitoring system from the start.

The SSR programme should create a monitoring system that will be sustainable beyond the life of the programme, and should thus aim to gradually increase local capacity and ownership as the programme develops. However, for the duration of the programme, what is really needed is *joint ownership* of the monitoring system between the external actor(s) providing assistance and key stakeholders in the partner country. This might involve:

- Ensuring that all relevant local and donor stakeholders (including non-governmental actors) are included in the management and reviewing structures that will oversee the monitoring system and act on monitoring information.
- Ensuring that these structures allow for the full and equal participation of women.
- Ensuring that they also include other relevant actors, such as other official bodies (national statistics offices, audit offices, etc.), other relevant ministries beyond the security system (ministry of education, ministry for women, etc.), members of the legislature (including opposition parliamentarians), and also civil society participation where appropriate.

- Agreeing how information will be shared between donor and local actors (including the general public and/or non-governmental actors where it is appropriate to share such information), and putting procedures in place for this information sharing to take place.
- Identifying where M&E processes and structures most require improvement (at the national level, but possibly also at the donor level) and both donor and national government actors taking appropriate steps to make such improvements.

Box 10.13. Questions to consider when establishing a monitoring system

- **What information *must* be collected?** What information is essential to report on key indicators and targets set in programme documents? Do baseline data still need to be collected?
- **What information can be collected so that it is “good enough”?** (See Box 4.1)
- **How will this information be collected?** What methods and tools will be used to collect this information? How long will this take and how much will it cost?
- **When should this information be collected, and how often?** How long will it take to collect this information? How regularly should it be collected (monthly, quarterly, annually)?
- **Who is responsible for collecting this information?** Which people or institutions should collect this information? Who is responsible for ensuring that this information is received?
- **Is the monitoring system jointly owned?** Does the system promote local ownership, broad participation and gender equality? Are there plans in place to transfer ownership gradually to local partners as the programme develops?

Other sources of information

- The UNDP *Handbook on Planning, Monitoring and Evaluating for Development Results* has a full chapter on monitoring for results (pages 99-126): www.undp.org/evaluation/handbook

Using monitoring information

The primary purpose of collecting monitoring information is to allow programme managers to understand how the programme is developing. This can help to:

- Check that activities are being implemented as planned (by reviewing output indicators).
- Assess whether the programme is achieving the planned changes at outcome (and impact) level.
- Monitor the context in which the programme is operating.
- Assess conflict sensitivity, *i.e.* how the programme is affecting and being affected by conflict dynamics. This is particularly important for SSR programmes.

On this basis, programme managers can then decide whether any changes need to be made to programme design and implementation. They may also make changes to the monitoring system if better or different information is required.

Most monitoring information can also be shared more widely with other interested parties and with the general public. This can build support for the programme by demonstrating what it has achieved so far. It also helps to promote transparency and accountability in the security sector. However, it may not be possible to publish all monitoring information (see below).

The process for reviewing monitoring information depends on the needs of the programme and the type of information. Some information will be used by programme staff as it becomes available in order to inform day-to-day management. At a more formal level, most SSR programmes have regular management meetings, and this is an opportunity to review and respond to monitoring information more systematically.

Using monitoring information is largely self-explanatory and does not require detailed guidance. From an SSR perspective, there are two points that are worth emphasising. Firstly, it has already been noted that most SSR programmes will have to make do with information that has weaknesses but is good enough. However, there is a risk that when programme managers come to review monitoring information and make decisions, they do not take the weaknesses of this information into account. This might lead them to misinterpret the data or be over-confident about their conclusions. Monitoring information is only as good as the people and systems that interpret it. Indicators *indicate*, but they rarely show the whole picture, and the less good-quality information that is available, the more important it becomes for programme managers to use experience and common sense in interpreting the data they receive.

Secondly, it has already been noted that monitoring systems may need to collect very sensitive information. It is essential that not only are SSR programmes conflict-sensitive, but so are their monitoring systems. This means that programme managers must carefully decide which information to share with other stakeholders, including the general public. On the whole, programme managers should try to publish as much information as possible to promote transparency and accountability, but they must always consider whether this information might influence political or conflict dynamics. If this is a risk, this does not necessarily mean that the information should be suppressed; it does however mean that it needs to be presented in a responsible and conflict-sensitive manner, particularly when shared with the media.

Box 10.14. Questions to consider when using monitoring information

- **How is the programme performing?** What progress is being made with planned outputs? What progress is being made towards achieving planned results (outcomes and impacts)? What can be learned from recent programme activities?
- **Is the programme conflict-sensitive?** How has the programme been affected by conflict dynamics? What impact is the programme having on conflict dynamics?
- **Has the data been interpreted correctly?** What conclusions can realistically be drawn from the existing information?
- **How has the context changed?** What does this mean for the programme and for risk management? Are the project inputs, activities and outputs still relevant?
- **Do programme designs or implementation plans need to be revised?** Do programme workplans need updating? Does programme design require any fundamental changes? Are there any targets that are unrealistic or are encouraging perverse behaviour?
- **Does the monitoring system need to be revised?** Is there a better way of collecting the information that is required, or compiling it in a more useful format? Is any new information required? Is there any information that does *not* need to be collected anymore?

Box 10.15. Checklist: Key issues for collecting and using monitoring information

- ✓ Do all programme staff and partner institutions understand why monitoring is important, and have the skills and capacity to undertake monitoring?
- ✓ Is the information collected “good enough” for management purposes? What information must be collected?
- ✓ Is it clear *what* information will be collected, *how*, *when*, *how often*, and *by whom*?
- ✓ Is the monitoring system jointly owned between the partner country and the donor institution(s)?
- ✓ Are women and men participating equally in the monitoring system?
- ✓ Are programme managers interpreting monitoring information correctly, and using it as a basis for revising programme design and implementation plans?
- ✓ Is the system monitoring how the SSR programme is affecting and being affected by conflict dynamics, and is monitoring information presented in a conflict-sensitive manner?

5. Evaluation

Whereas monitoring is a continuous process, evaluation is a distinct activity that aims to assess in greater depth how a programme, policy, institution or web of institutions is performing or has performed. The main purposes of evaluations are to enhance accountability (*i.e.* to show how well the programme works, and thus whether resources have been used effectively and responsibly) and to learn lessons that can improve future programming (Chapter 2).

Unfortunately, both programme staff and partner country stakeholders often approach evaluations with limited enthusiasm. There are three main reasons for this. Firstly, the evaluation may feel imposed (this is not only a concern for local stakeholders; international staff at field level may also feel that the evaluation is being imposed by headquarters). Secondly, the evaluation may be poorly designed, leaving staff unclear why it has been commissioned or what it is supposed to achieve. Thirdly, programme staff may be reluctant to invest time and effort in the evaluation because they are sceptical about anything changing as a result. These problems can be avoided by ensuring that the commissioning process is owned by all stakeholders, that the purpose of the evaluation is clear and that a suitable methodology has been identified (Section 5.1). Senior programme managers must also demonstrate a willingness to respond to the findings of the evaluation (Section 5.4).

In most cases, evaluations should be external and independent. Since this toolkit is primarily aimed at SSR programme managers and designers, this chapter does not provide guidance on how to undertake an evaluation, so much as how to plan, manage and respond to one. However, this chapter will also be of interest for professional evaluators as it identifies how and why evaluations of SSR programmes may differ from evaluations in other spheres. The guide questions for SSR programmes (Section 5.2) are particularly useful when evaluators and programme managers are working together to agree or finalise the scope and methodology of the evaluation.

This chapter has four parts:

1. Commissioning an evaluation: Who designs the evaluation, what should it evaluate, and how?
2. Evaluation criteria and guide questions
3. Managing the evaluation
4. Follow-up: Publishing and responding to the evaluation

Commissioning an evaluation: Who designs the evaluation, what should it evaluate, for whom, and how?

As noted above, evaluations can feel imposed if the commissioning process is dominated by the donor, particularly if the demand for an evaluation is driven from headquarters. If local partners and programme staff feel little ownership over the evaluation, they are less likely to respond to its findings, and it is less likely to promote learning and accountability. The principle of local ownership should therefore be central to how the evaluation is designed, managed and used.

If the SSR programme's monitoring system is grounded in joint ownership (Section 4.1), it should be possible to use the same mechanisms for evaluation. Even if this has not been done, some form of joint management panel involving partners should still be established for the evaluation. Programme managers should also consider commissioning an evaluation jointly with other donors/institutions for co-funded programmes and for more ambitious types of evaluation such as sector evaluations and thematic

evaluations (see below). This promotes harmonisation of aid and reduces the burden on partner country M&E systems.

Establishing the purpose and scope of the evaluation

Having established a suitable management structure, the next step is to agree the purpose and scope of the evaluation. Evaluations are generally held to have two main purposes: to show whether the programme 'works' and thus demonstrate that the programme is accountable; and to allow lessons to be learned which can inform future programming (*e.g.* further security and justice provision in that country; further international assistance to an SSR programme in that country; similar SSR programmes/activities undertaken in other countries; or new policies and tools relating to SSR).

What is less widely acknowledged, however, is that there may be other purposes for commissioning an evaluation, such as:

- Informing future funding decisions.
- Judging the performance of external contractors, key stakeholders in partner countries, and even international actors.
- To send political messages to key stakeholders, where those commissioning the evaluation expect (or ensure) that it will reach certain conclusions.

These purposes may not be complementary. For example, programme staff and partners are less likely to talk honestly during an evaluation about the challenges they faced if they know its overriding purpose is to determine funding allocations. Such evaluations will therefore have limited use for lessons learning. In reality, these competing purposes will often exist in parallel and cannot always be reconciled. Nonetheless, whenever possible these factors should be consciously recognised when evaluations are being designed, and differences between competing purposes should be minimised.

It is also worth thinking carefully about the scope of the evaluation. Traditionally, most evaluations operate at the programme level because managers want to know whether the programme is working. Even at this level, however, evaluations can focus on several different levels of the results chain (Section 3.4), though they usually review results at outcome and impact level. Evaluations should only look in detail at outputs if there have been serious problems with programme implementation that require investigation.

Evaluations of a higher, more systemic nature are starting to become more common. These can look at the nature and quality of donor support to security and justice systems as a whole, evaluating the results of mechanisms such as sector-wide approaches (SWAPs) and budget support. They can also take a wider overview of the security system in the partner country, evaluating broader dynamics and security management and governance (and then perhaps assessing the impact of the SSR programme within this wider context). Such evaluations are more complicated, costly and time-consuming, and are unlikely to be commissioned for a specific SSR programme. Nevertheless, programme managers should be aware of such evaluation types as they can provide important insights into many vital practical and structural elements of SSR. The main donor institutions should thus consider whether it is possible to commission more systemic evaluations, preferably jointly. Table 10.7 provides examples of how SSR can be evaluated at these different levels.

Table 10.7. Levels of evaluation for SSR

Scope of evaluation	Description	SSR example
Project evaluation	Evaluation of individually planned activities designed to achieve specific objectives within a given budget and time period	Evaluation of project providing human rights training to police officers
Programme evaluation	Evaluation of a coherent set of activities in terms of policies, institutions or finances, usually covering a number of related projects or activities in one country	Evaluation of multi-year justice reform programme
Sector evaluation	Evaluation of a single sector/system or sub-sector	Evaluation of reform dynamics in the security system as a whole or in a sub-sector such as policing or defence
Evaluation of developmental aid instruments	Evaluation of a specific instrument or channel for development aid funding. Usually cross-country and cross-sectoral	Evaluation of effectiveness of technical assistance such as provision of equipment to police services; evaluation of sector-wide support to justice, law and order sector
Country programme evaluation (all types of development assistance to one country)	Evaluation of the combined cross-sectoral support provided by a single funding agency to a partner country. This could be done as a joint multi-stakeholder evaluation	Evaluation of all support provided to a country, including SSR, and link between SSR and other objectives, such as poverty reduction
Country development evaluation (all types of development assistance to a country <i>from a partner country perspective</i>)	Evaluation of the combined cross-sectoral support of all funding agencies to a partner country. Can include trade, donor and policy coherence, and often in relation to the country's poverty reduction strategy	Evaluation from country perspective of how donors have provided support, including to SSR, and how this relates to the country's own priorities
Evaluation of partnerships and global funding mechanisms and institutions	Include the evaluation of NGO partnership schemes, global funds, global public-private partnerships and global institutions such as the UN agencies.	Evaluation of agencies providing support for SSR, such as UNDP and the UN Department of Peacekeeping Operations (DPKO)
Thematic evaluations at global and national level	Evaluation of selected aspects of different types of development aid instruments, <i>e.g.</i> influence on trade negotiations, environment, gender, HIV/AIDS or evaluating a range of sector programmes in different countries	Evaluation of impact of SSR programmes as a form of assistance, evaluation of the effectiveness of SSR programmes in different contexts around the world

Source: Adapted from UK Department for International Development (2005), *Guidance on Evaluation and Review for DfID Staff*, DFID, London.

Table 10.6 moves from small, project-level evaluations through to global thematic evaluations. It should be noted that project and sub-sector evaluations would not strictly be SSR evaluations, since they look only at one aspect of the security system, rather than the system as a whole. For example, they may evaluate the police, border or prison sectors and/or programmes to reform these sub-sectors, but they do not address the cross-cutting nature of SSR.

Lastly, it should also be clear *when* the evaluation will take place. Evaluations may be held at several points in the project cycle, but they are most frequent either during implementation (interim evaluation) or after the programme is over (end-of-programme evaluation). Most end-of-programme evaluations happen

immediately after the programme is completed. However, SSR usually involves long-term changes that are sustained long after the programme has finished; there is thus a need for assessments that are carried out two to three years later to evaluate whether any lasting impact has been achieved.

Choosing an appropriate methodology

It is also necessary to decide what methodology the evaluation should employ. It is common for those commissioning the evaluation to define the methodology in some detail in the terms of reference (TOR) for the evaluation. However, while it may be useful to have a general idea of how the evaluation will be carried out, it is better not to be too prescriptive at this stage. Evaluation specialists are more likely than programme managers to know what is the most appropriate way of running the evaluation, and it would be better to reach mutual agreement once an evaluation team has been selected (see Section 5.3). Furthermore, it often becomes necessary to adapt the methodology during the evaluation process as evaluators understand more about the programme and the context in which it is operating.

Nonetheless, programme managers should be aware of some of the main methods that are available, particularly when it comes to analysing how the programme has contributed to the achievement or otherwise of higher-level impacts (Box 10.16).

Box 10.16. Methods for measuring impact

As noted in Section 3.4, it is not normally feasible to attribute changes that occur at the impact level directly to specific programmes and interventions. However, this does not mean that it is impossible to study the impact of a programme and the degree to which it has contributed to higher-order changes. But this does require more sophisticated techniques than more basic evaluations usually employ. A vast range of methods have been developed in recent years to address such questions. Some of the most popular include:

- **Impact assessment/evaluation.** The terms “impact assessment” and “impact evaluation” are used by different actors in different ways and can refer to several different forms of evaluation. Here we use it to refer to the use of “counterfactual” analysis of the impact of an intervention on final outcomes. A counterfactual analysis compares what actually happened with what might have happened had the intervention not taken place. This is done by comparing a range of data for areas/groups where the intervention took place (sometimes selected randomly) and for similar areas/groups not affected by the intervention.
- **Contribution analysis.** Contribution analysis seeks to provide plausible evidence of the difference that a programme is making to observed outcomes. It was first proposed by John Mayne in Canada in 1999, and has since been adapted for use in development contexts. The idea is to assess the programme logic, analyse the results that have been achieved (usually using existing M&E structures), to consider alternative explanations for these results, to build a “story” about the contribution the programme has made, and then test this story with stakeholders.
- **Outcome mapping.** Designed by the International Development Research Centre, outcome mapping is an evaluation tool that focuses outcomes on changes in behaviour of individuals, groups and institutions and the relationship between these individuals and groups. It thus approaches impact in a very different way from traditional methodologies that focus on more tangible “products” of a programme. This methodology may be particularly useful for SSR programmes where changes in attitudes and behaviours (e.g. among senior officials working in security and justice institutions) is just as important as any specific practical changes that occur.

More information on these forms of evaluation can be found in the Additional Resources section at the end.

Programme managers also need to decide whether the evaluation will be external, internal or participatory (Table 10.8). External and internal evaluations are both driven by expert evaluators; the main difference is that while external evaluations should be done by evaluators who are totally independent of the

programme and have no stake in its outcome, internal evaluations are done by staff from within the institution (but not by programme staff themselves). Many donor institutions require most evaluations to be external and independent so that they are credible for accountability purposes; other internal mechanisms may therefore be needed to promote lessons learning.

By contrast to these forms of evaluation, in participatory evaluations the evaluators play the role of facilitators rather than experts. The aim is to mobilise a community to evaluate a project according to its own terms and criteria. Participatory evaluations help to create local ownership of programmes and to ensure that they are genuinely in line with local needs and priorities; in practice, however, they have rarely been used for SSR programmes, which have tended to employ a top-down logic of intervention and evaluation.

Table 10.8. The differences between external, internal and participatory evaluation

	External	Internal	Participatory
Who?	Expert evaluators with no stake in the outcome of the evaluation Truly independent evaluation allows evaluators to set scope of/ questions for evaluation	Expert evaluators within the institution(s) commissioning the evaluation, but operationally separate	Community members, in collaboration with project staff and external facilitators
What?	Performance evaluated against formally-defined standards, often with reference to formal goals and objectives		Community members and other participants set their own standards of success
How?	Evaluators control data collection and analysis. Scientific criteria of objectivity. Outsider perspective		Self-evaluation. Collaborative process of data collection and analysis
Advantages/disadvantages	Most likely to be objective Independent evaluators required for accountability purposes External actors may bring fresh eye and new perspectives to programme External evaluators may not sufficiently understand context, structural obstacles, opportunities for change	Internal evaluators better understand institutional specifics May be most useful for lessons learning as less “threatening” to programme staff Less credible if shared with other stakeholders	Participation puts beneficiary voices first Participation encourages local ownership and empowerment Greater scope for ensuring participation of women, minorities and marginalised groups Evaluation may be more relevant to local needs/perspectives Evaluation may not answer questions that are important to other stakeholders

Box 10.17. Questions to consider when conducting an evaluation

- **What is the primary purpose of the evaluation?** Is it for accountability purposes? Is it to assess how well a specific SSR programme has performed? Is it to learn lessons? Are there other management objectives driving the evaluation, such as funding pressures?
- **Who controls the commissioning process?** Do the partner country and the donor jointly own the commissioning process? Does this process take into account and balance the views of partner government officials, donor headquarters, local programme staff and other key stakeholders and beneficiaries? Are there any opportunities to commission a joint evaluation with other donors or institutions?
- **When will the evaluation take place?** Is this a mid-term or end-of-programme evaluation? Is there any opportunity to evaluate the impact of the programme long after it is over?
- **Will the evaluation be external, internal or participatory?** Will the evaluation be run by independent external evaluators? Or by in-house evaluators? Or will a participatory approach be used, where local communities shape the evaluation with external facilitation?

Action point

- **Commission more evaluations at the systemic level.** In recent years, some of the major international institutions supporting SSR have begun to undertake more evaluations that go beyond programme evaluation, such as sector-wide and thematic evaluations. Major supporters of SSR should therefore undertake more such evaluations, preferably jointly.

Other sources of information

- The following OECD DAC documents provide relevant standards and guidance on evaluation, including in the related field of conflict prevention and peacebuilding:
 - OECD (2002), *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Paris, available at www.oecd.org/dataoecd/29/21/2754804.pdf
 - OECD (2010), *Evaluating Development Cooperation: Summary of Key Norms and Standards*, OECD, Paris, available at www.oecd.org/dataoecd/12/56/41612905.pdf
 - OECD (2010), "Quality Standards for Development Evaluation", *DAC Guidelines and Reference Series*, OECD, Paris, available at www.oecd.org/dataoecd/55/0/44798177.pdf
 - OECD (2008), *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities*, Working draft, OECD, Paris, available at www.oecd.org/secure/pdfDocument/0,2834,en_21571361_34047972_39774574_1_1_1_1_00.pdf
- Sida's evaluation manual, *Looking Back, Moving Forward*, is a very useful general guide and also contains a good discussion of participatory evaluation on pages 17-20. Sida (2007), *Looking Back, Moving Forward*, Sida, Stockholm, available at www.sida.se/Svenska/Om-oss/Publikationer/Visa-publikation/?iframeSrc=http://www2.sida.se/sida/jsp/sida.jsp%3Fd=118%26a=3148&language=en_US&searchWords=looking%20back

Evaluation criteria and guide questions

The OECD DAC has agreed five standard criteria for the evaluation of development: relevance, efficiency, effectiveness, impact and sustainability. These provide a basis for the evaluation of SSR programmes, but they can be expanded to include other important issues that are often overlooked or only poorly addressed. This toolkit thus suggests four extra criteria and expands the focus of sustainability to emphasise questions of ownership:

- **Coherence.** OECD DAC guidance promotes "whole-of-government" approaches to development support to SSR. This criterion focuses on whether the support and initiatives of different international actors and departments are operating together in a coherent and strategic fashion.

- **Co-ordination and linkages.** While coherence focuses on how donor support is provided, this criterion looks at co-ordination between different parts of the security system and how this is promoted by SSR programmes. It also looks at links between SSR and other government strategies, such as poverty reduction strategies.
- **Consistency with values.** The long-term impact of SSR programmes often depends as much on whether it has succeeded in promoting key principles and values in security governance as it does on providing specific skills, equipment or structures. It is also essential that the SSR programme itself upholds these values. Consistency with values is thus an important focus for evaluation.
- **Sustainability and ownership.** Although in theory it is obvious that programmes must be locally owned to achieve sustainable results, many SSR programmes have not paid sufficient attention to local ownership of either the programme itself or the system it creates.

Table 10.9 translates these criteria into guide questions for evaluations. This list is not intended to be comprehensive, but to provide inspiration to evaluators and those commissioning evaluations. It should help them to decide how deeply to focus on each criterion and to choose which questions are most important for each evaluation.

Table 10.9. Evaluation criteria and guide questions

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, as well as for women vs. men, in light of their different security needs and the way in which these are met? • 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?
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

	<p>stakeholders, including women, men, boys and girls?</p> <ul style="list-style-type: none"> • 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? • Is the intervention consistent with partners' priorities and effective demand? Is it supported by local institutions and well integrated into 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, women and men, representatives of marginalised groups) 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 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 and gender equality plans? • 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, men and women?

Managing the evaluation

Since evaluations are generally run by independent evaluators, programme managers should not interfere in the actual process of undertaking an evaluation. Nonetheless, there are various practical steps for which programme managers are responsible, some of which require extra thought for SSR programmes. These include:

- Drafting terms of reference and hiring an evaluator or evaluation team
- Clarifying stakeholder involvement
- Assessment of risk and conflict sensitivity

- Summarising the theory of change for evaluators

The terms of reference (TOR) state what is required from the evaluation. They provide a template for hiring independent evaluators and then act as the main guidance document to ensure that the evaluation is implemented correctly. They include information such as the objectives of the evaluation, available resources, timing, cost, management arrangements, and the products that the evaluation is expected to deliver. They usually also set out the skills, knowledge and other qualities that an evaluator or evaluation team should have. From an SSR perspective, there are three key criteria for evaluators/evaluation teams that can be difficult to meet simultaneously:

- **Evaluators who combine technical knowledge of SSR with M&E experience, including experience of gender and security issues** (or an evaluation team which together combines the appropriate knowledge).
- **Evaluators with knowledge of the political and social context in the country/area where the SSR programme is taking place.** If there enough resources to hire a team rather than an individual, the team should include a national of the partner country wherever possible; at the least, the evaluation team should include evaluators who speak the local language(s),
- **Evaluators who have or can easily acquire security clearance.** For some sensitive evaluations (*e.g.* intelligence reforms, SSR in stabilisation environments), evaluators may need security clearance in advance, or vetting time must be factored into planning.

Programme managers should try to ensure that the evaluation encourages broad stakeholder involvement and participation. Evaluations of SSR programmes are often based on a limited number of interviews with international programme staff and partner country officials, with little or no attempt made to seek the views of other interested parties, including the communities that are intended to be the ultimate beneficiaries of the programme. Another common flaw is that often little is done to ensure the suitable participation of women and men (and where appropriate, boys and girls as well). Managers should not be prescriptive about how evaluators should meet, but they can help evaluators to seek the views of as many (groups of) people as possible by suggesting possible targets, sharing contact information, and helping to organise meetings and other research events if necessary.

Another important task for SSR programme managers is to assess any potential risks related to the evaluation. What controversies or issues may surround the evaluation (values, politics, personalities), and what impact might they have? What does this mean in terms of conflict sensitivity, and is there a more conflict-sensitive way to approach the evaluation?

In particular, SSR programme managers should be aware of any risks to physical security related to the evaluation. An important question to ask is: Who or what might be at risk? Are there any physical threats to evaluators (for example by travelling to certain areas or researching certain issues)? Could other people be placed at risk by any evaluation activities: for example, if evaluators wish to interview opposition groups or marginalised and vulnerable communities about their perceptions of security and the SSR programme, but powerful stakeholders are opposed to this, what are the risks to these groups? How will any sensitive information be handled? How will expectations be managed so that interviewees do not feel that their honesty has been exploited as they see nothing in return?

When any such risks have been identified, managers must then develop a strategy for managing or avoiding them. The conflict sensitivity of the evaluation needs to be monitored *constantly* (not only at the start), since conflict dynamics may change rapidly in fragile environments.

Programme managers can help evaluators immensely by providing a summary of the theory of change underpinning the programme (Section 3.6). As noted above, SSR programmes are often vague about their overall objectives, either because of poor design, because the programme set deliberately vague objectives to allow it space to develop over time, or because the real objectives have been deliberately obscured in programme documents due to their sensitivity and the risk that key stakeholders might resist if the objectives are clearly stated from the start. If evaluators are not aware of the programme's theory of change, the evaluation may miss the point; if this has never been explicitly stated, evaluators may have to spend considerable time with managers and project staff seeking to draw out the theory of change. Therefore, if programme managers can provide a summary of the programme logic to evaluators (in some cases, this may need to be done privately and not included in evaluation documents), this can greatly improve the quality of the evaluation.

Box 10.18. Questions to consider for managing the evaluation

- **Do the evaluator(s) have the right skills and characteristics?** Do they have knowledge of both SSR and M&E? What is the gender balance and are there any evaluators with knowledge of gender and security issues? Do they understand the local political and social context? Do they have or do they need security clearance for this evaluation?
- **Are the evaluator(s) meeting with as many (groups of) people as possible?** Can programme staff suggest and/or help to organise meetings with a wide range of stakeholders and beneficiaries? Are separate meetings necessary with women, men, boys and girls?
- **Are there any risks to physical security or conflict sensitivity associated with the evaluation?** Do any evaluation activities threaten the physical security of either evaluators or those individuals and communities that are evaluated? Are there any questions or activities that are particularly sensitive and need to be handled very carefully?
- **Do the evaluator(s) understand the programme's theory of change?** Is the theory of change obvious from programme documents? Are there any objectives that have deliberately not been mentioned in programme documents? Is it necessary for programme managers to provide a summary of the theory of change to evaluators?

Following up: publishing and responding to the evaluation

The follow-up period begins long before the evaluation is officially published, from when the evaluators begin to draft the final report.

One reason why many evaluations do not have much impact is that decision makers find their format inaccessible. While this problem is not specific to SSR, it is a very important one, and thus deserves brief guidance. Programme managers should discuss presentation with the evaluators early in the report-drafting process. Several key principles should apply:

- **Use accessible language:** The evaluation report should avoid jargon and present information in as clear and simple a format as possible. This is particularly important so that it is accessible to partner country stakeholders for whom English may not be their first language.
- **Be realistic.** The evaluation needs to be based on a realistic analysis of the wider context in which the programme is taking place, and acknowledge any structural and contextual limitations that affect what is possible.
- **Provide actionable recommendations.** Recommendations must not only be realistic, they must also be actionable, *i.e.* it must be possible for a specific actor (the partner government,

international actor(s), local civil society organisations, etc) to act on the recommendation in a coherent fashion.

- **Prioritise recommendations.** Recommendations should be prioritised, indicating which are most important and which are of secondary importance. Long lists of recommendations that are not prioritised are off-putting and likely to be ignored.

Once the evaluators have prepared an initial draft of the evaluation report, it will normally be appropriate to provide key stakeholders with a chance to review and respond before it is circulated more widely. This may take the form of a validation workshop, where the main findings of the evaluation are presented to stakeholders, or the report may be circulated for comment. This acts as a quality control, allowing key stakeholders to propose corrections regarding any information they believe to be factually wrong and to suggest clarifications/ improvements where they disagree with the findings or recommendations. However, no one should have any form of veto over the evaluation: key stakeholders can review the evaluation and propose changes, but the evaluators retain independence and the right to present the findings as they believe to be accurate.

Conflict sensitivity is a particular concern when publishing evaluations of SSR programmes. In some cases, it may be appropriate to prepare two versions of the evaluation report: a full report which is shared with a restricted number of people; and an edited version that is distributed more widely. This is because evaluations of SSR can touch on issues that are highly sensitive, especially if they are critical of certain actors within the partner country. In such circumstances, it may be appropriate to restrict access to some findings (rather than removing these findings from the evaluation entirely).

Lastly, once the evaluation report has been published, a response must be prepared. This avoids the report being left unread on a shelf. Key stakeholders at both international and partner country level should review the evaluation and prepare a response indicating what action they intend to take. Some institutions already require senior decision makers to publish an official response to certain types of evaluations; while this would appear to be good practice, there is still a risk that a formal response will not lead to appropriate changes, even if a commitment is made on paper. There needs to be genuine commitment, across all levels of the institution, to learn from the evaluation. Building this depends on the institutional culture and the incentives to learn (see action point in Box 10.19). If there is not a positive institutional framework for M&E well before the evaluation is published, the evaluation itself is unlikely to lead to significant change. Similarly, if local ownership of the evaluation has not been established at an earlier stage, local stakeholders may feel little obligation to act upon its findings.

Box 10.19. Questions to consider for the evaluation follow up

- **Is the evaluation presented in an accessible format?** Is the language easy to understand? Are recommendations actionable (*i.e.* realistic to act upon)? Is there any prioritisation of recommendations indicating which are the most important?
- **Have key stakeholders had the chance to review and respond to the evaluation before publication?** Do they understand that they can suggest – but not demand – changes and clarifications?
- **What is the likely impact on conflict dynamics of publishing the evaluation?** Is it necessary to restrict access to certain findings? Can an edited version be made available?
- **What response is required to the evaluation and by whom?** Is there any institutional obligation to respond to the evaluation? Should senior decision makers and/or programme managers publish a formal response to the evaluation, detailing what actions they intend to take? Can international actors issue a joint response with partner country officials?

Action point

- International actors should review their requirements for responding to evaluations, and consider making it obligatory to publish a response to certain types of evaluation. This response would list what actions they intend to take following the evaluation.

Box 10.20. Checklist: Key questions when designing, managing and publishing an evaluation

- ✓ Who controls the process of commissioning and sharing the evaluation? Can local stakeholders (including field-level international programme staff) input into its design and management?
- ✓ Is the primary purpose and the scope (programme/sector/systemic/thematic) of the evaluation clear, and has it been agreed by all key stakeholders?
- ✓ Is the methodology appropriate, and has it been agreed with the evaluator(s)?
- ✓ Do the evaluator(s) have all of the necessary characteristics, such as knowledge of M&E, knowledge of SSR programming, gender expertise and knowledge of the local context? Do they have or need security clearance?
- ✓ Has there been an assessment of the risks to physical security both for evaluators and the individuals and communities that are contacted? If risks have been identified, is there a strategy for risk management/avoidance?
- ✓ Is the evaluation conflict-sensitive, *i.e.* is there any risk that the evaluation could negatively affect conflict dynamics?
- ✓ Who should respond to the evaluation findings, and how?

6. Common challenges and possible solutions

Chapter 3 makes it clear that many of the common challenges in monitoring and evaluation could be avoided by addressing them during the planning phase. Nonetheless, no programme is ideal and problems are likely to arise regardless of how well the M&E system has been designed. This chapter suggests a simple method for “troubleshooting” any problems, and then looks briefly at some of the issues that are most likely to arise and suggests some possible solutions in table form.

Troubleshooting M&E problems

While many problems encountered during monitoring and evaluation are caused by a failure to invest proper resources (including time) in developing and maintaining a suitable M&E system, it is not always the case that these problems can be solved simply by doing more or better monitoring and evaluating. In fact, many such problems are symptoms of deeper issues, such as the overall design of the programme, the relationship between programme partners, misunderstandings or disagreements about the purpose of certain M&E activities, etc.

For this reason, when problems arise that are perceived as “M&E problems”, it is worth following the three steps to diagnose the reasons for these problems and identify suitable solutions:

1. Classify the symptoms
2. Analyse underlying problems and deeper causes
3. Identify short-term and longer-term responses.

This is best illustrated through an example. Imagine that a problem has arisen a year into a programme because of a disagreement between the international donor and the partner government about monitoring public perceptions of security. The donor wishes to commission an independent survey from a non-governmental organisation, but the partner government voices opposition. Despite the donor’s insistence that this was a fundamental element of measuring success, the partner government continues to block the survey. What should be done in such circumstances?

Firstly, start by looking at the symptoms. At least two can be identified. The first is the partner government’s hostility to perception surveys. Yet there is also a second – a lack of shared ownership of all aspects the M&E system.

The second step is to analyse the underlying causes of these symptoms. The partner government’s hostility to perception surveys may be explained by several things – perhaps a general distrust of non-governmental actors, or a concern that the survey will not produce “positive” results and will therefore make the government look bad in the eyes of its citizens.

Similarly, there may be several causes for the lack of shared ownership of the M&E system: Is it simply a largely technical misunderstanding about certain M&E tools, or is there a fundamental gap in approaches and expectations where M&E issues are no more than a symptom of much deeper problems? For example, the concept of results-based management may not be the guiding principle for the partner country’s security sector, and therefore the very idea of producing information that identifies a “weak” starting point (as baselines are intended to do) may be very challenging. It may also be that partner government officials did not fully understand what was involved and the implications. Perhaps more could have been done to generate genuine (rather than formal) ownership of such M&E tools. Or there may have been firm agreement during the design phase, but the government has since backtracked – if so, why?

Only once such questions have been answered is it possible to identify short- and longer-term solutions (Table 10.10). In this case, there may be several short-term solutions, depending on what the international programme manager thinks is most appropriate. One option is to drop the demand for perception surveys and find alternative methods of obtaining similar monitoring information. Another is to explain the reasons for undertaking perception surveys in more detail, and thus to persuade the partner government why it is worth doing so. Over the longer term, there may be any number of steps that can be taken to build stronger joint ownership of the M&E system, from reviewing the whole set-up in detail through to further capacity building.

Table 10.10. Exploring common problems, underlying reasons and possible solutions

Problem	Underlying issues/ questions to ask	Possible solution
Lack of baseline data/study	Why was this data not collected? Because of poor planning? Or because the information is hard to access or collect?	It may be possible to recreate a baseline by collecting data from other sources relating to the time immediately before the programme started Refer back to the analytical baseline in original project documents, <i>i.e.</i> the analysis on which the programme design was originally based Consider using participatory monitoring techniques to identify the most important information
Impossible to identify or collect suitable indicators	What are indicators needed for? Why is it not possible to collect them?	Indicators may not be the most appropriate way of measuring change in this context; other methods can be found In fragile and conflict-affected environments, it may not be possible to use standard indicators, but proxy indicators may provide similar information
Partner government information systems do not produce reliable information	Is this due to lack of capacity? Or because statistics are deliberately distorted for political purposes?	Over the long term, the solution is to invest in building the capacity of national M&E systems Over the short term, it may be possible to find similar information from other sources, such as surveys by non-governmental organisations
Information required for reporting not available	Is this information available anywhere? If it is not available, can it be easily collected?	Sometimes, data have simply not been collected due to the lack of a proper monitoring system. In such cases, asking key stakeholders may provide the required information – but ask sensitively and acknowledge the failure to request this information earlier This also indicates a need to invest in better monitoring systems in future
Programme staff observe unexpected negative results being generated by programme	Why are these negative results occurring? Is it a result of perverse incentives from indicators/targets?	If it appears that an indicator or target is causing programme/partner country staff to behave in a perverse way, consider how the target could be revised to achieve the same objective without encouraging inappropriate responses
Key stakeholders resist evaluation process or co-operate to the minimum degree necessary	Do stakeholders understand what the evaluation is for? Have they been involved in the M&E process up to now?	Stakeholders may simply not understand why the evaluation is taking place and what the information will be used for. Explaining why the evaluation is necessary and how it will help improve programming

		<p>may make them more co-operative.</p> <p>This may also suggest a lack of participation in the M&E system. If it is not too late to do so, revise the management of the evaluation and/or its methodology to give stakeholders greater involvement in its successful conclusion</p>
Evaluation does not produce useful results	Is this because the evaluation was poorly designed? Or because the evaluators do not have the right skills?	It may be helpful to organise a meeting with the evaluation team to discuss the problems. Discussing the programme's intervention logic and the scope and methodology of the evaluation is most likely to identify where the problems lie. If possible, revise the evaluation methodology to ensure a more appropriate result
Evaluation is ignored/rejected by key stakeholders in partner country	Is this because they do not feel any ownership of the evaluation? Is it because the findings are too sensitive, or they find certain conclusions offensive?	<p>If still possible, give key stakeholders the chance to prepare a response to the evaluation report, identifying any areas that they believe are factually incorrect and suggesting improvements they would like to see to the evaluation (while emphasising that they do not have the right to veto the evaluation)</p> <p>Assess the conflict sensitivity of the evaluation and take account of any findings that are likely to be controversial, if necessary producing an edited version of the evaluation that is disseminated more widely</p> <p>Consider presenting various versions or summaries of the evaluation findings that are more accessible to key audiences, and organising meetings to share and discuss the findings</p>
Limited participation by women in the M&E system	What efforts have been made to ensure women participate and are represented?	<p>Try to involve the ministry responsible for women's affairs and/or gender.</p> <p>Consult with both national and local women's organisations on how to reach out to and involve women. It may be necessary to organise focus groups or meetings with women; if so, these should be planned to ensure they are at appropriate times and in accessible places.</p>

Additional resources

The following are recommended resources for further information on monitoring and evaluation and security sector reform:

Bamberger, M. (2005), *Designing Quality Impact Evaluations Under Budget, Time and Data Constraints*, World Bank, Washington, DC, http://siteresources.worldbank.org/INTISPMA/Resources/Training-Events-and-Materials/Designing_quality_IE_under_constraints.pdf.

Caputo, E. *et al.* (2008), "Methodology for Evaluations of Budget Support Operations at Country Level", *Issue Paper*, EuropeAid Evaluation Unit, Brussels.

Chapman, J. and A. Mancini (2008), *Impact Assessment: Drivers, dilemmas and deliberations*, Sightsavers International, West Sussex.

Church C. and J. Shouldice (2002), *The Evaluation of Conflict Resolution Interventions: Framing the State of Play*, INCORE/United Nations University, Londonderry.

Dakolias M. (2005), "Methods for Monitoring and Evaluating the Rule of Law", in Centre for International Legal Cooperation (2005), *Applying the "Sectoral Approach" to the Legal and Judicial Domain*, CILC, The Hague.

Davies, R. and J. Dart (2005), *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*, MandE/Clear Horizon, Victoria, Australia.

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

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

International Development Research Centre website. Contains various materials about outcome mapping at: http://www.idrc.ca/en/ev-26586-201-1-DO_TOPIC.html

Johannsen, A.M. (2009), "Training Resources on Security Sector Reform Assessment, Monitoring and Evaluation and Gender", in M. Bastick and K. Valasek (eds.), *Gender and Security Sector Reform Training Resource Package*, DCAF, Geneva, www.gssrtraining.ch/index.php?option=com_content&view=article&id=145&Itemid=33&lang=en

Jones, R. and V. Young (2004), *CIDA Evaluation Guide*, Evaluation Division, Performance & Knowledge Management Branch, Canadian International Development Agency, Quebec.

Kusek, J.Z. and R.C. Rist (2004), *Ten Steps to a Results-Based Monitoring and Evaluation System*, The International Bank for Reconstruction and Development/The World Bank, Washington, DC.

Mayne, J. (2006), *Exploring Attribution through Contribution Analysis*, available at www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief16_Contribution_Analysis.pdf

Molund, S. and G. Schill (2007), *Looking Back, Moving Forward: SIDA Evaluation Manual*, Sida, Stockholm.

Monitoring and Evaluation News website, <http://mande.co.uk/>

Moser A. (2007), *Gender and Indicators: Overview Report*, Institute for Development Studies, Brighton.

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

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

OECD (2006), *Guidance for Managing Joint Evaluations*, OECD, Paris.

OECD (2008), *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities*, working document for application period, OECD, Paris.

OECD (2008), *Network on Development Cooperation, Evaluating Development Co-operation: Summary of Key Norms and Standards*, OECD, Paris.

Popovic N. (2008), "Security Sector Reform Assessment, Monitoring and Evaluation and Gender", in M. Bastick and K. Valasek (eds.), *Gender and Security Sector Reform Toolkit*, DCAF and OSCE/ODIHR and UN_INSTRAW, Geneva, www.dcaf.ch/Publications/Publication-Detail?lng=en&id=47416

Roche, C. (1999), *Impact Assessment for Development Agencies: Learning to Value Change*, Oxfam GB, Oxford.

Rynn, S. and D. Hiscock (2009), *Evaluating for Security and Justice: Challenges and opportunities for improved monitoring and evaluation of security system reform programmes*, Saferworld, London.

Sartorius R. and C. Carver (2006), *Monitoring, Evaluation and Learning for Fragile States and Peacebuilding Programs: Practical Tools for Improving Program Performance and Results*, Social Impact, Washington, DC.

Smutylo, T. (2001), *Crouching Impact, Hidden Attribution: Overcoming Threats to Learning in Development Programs*, International Development Research Centre, Ottawa.

UNDP (United Nations Development Programme) (2006), *Measuring Democratic Governance: A Framework for Selecting Pro-Poor and Gender Sensitive Indicators*, UNDP, New York.

United Nations Development Programme Evaluation Office (2009), *Handbook on Planning, Monitoring and Evaluating for Development Results*, UNDP, New York.

Endnotes

- 1) See Section 3 of this handbook, as well as Clingendael: Netherlands Institute of International Relations (2003), *Enhancing Democratic Governance of the Security Sector: An Institutional Assessment Framework*, report prepared for the Netherlands Ministry of Foreign Affairs, Clingendael, The Hague.
- 2) See OECD (2005), *The Paris Declaration on Aid Effectiveness*, OECD, Paris, available at www.oecd.org/dataoecd/30/63/43911948.pdf.
- 3) See OECD (2007), *OECD DAC Principles for Good International Engagement in Fragile States and Situations*, OECD, Paris, available at www.oecd.org/dataoecd/61/45/38368714.pdf.
- 4) See DFID (2009), "Guidance on using the revised Logical Framework", *How to note: A DFID Practice Paper*, February 2009, DFID, London, available at: www.dfid.gov.uk/Documents/publications1/how-to-guid-rev-log-fmwk.pdf
- 5) These categories are only intended to make this list more accessible; it should be noted there is some overlap between the categories.
- 6) These theories are mostly not antagonistic or exclusive; more complex programmes may employ several theories of change at the same time.
- 7) See for example, OECD (2008), *DAC Guidance on Evaluating Conflict Prevention and Peacebuilding*, Working draft for application period, OECD, Paris, p 82. Available at www.oecd.org/secure/pdfDocument/0,2834,en_21571361_34047972_39774574_1_1_1_1,00.pdf