Helpdesk Research Report: Police reform evaluations

Date: 1/03/2012

Query: Provide research on successful evaluations of police reform programmes. If possible focus on evaluations that aim to measure impacts on increased state capacity, citizens’ level of trust and the effect of increased stability on poverty.

Enquirer: DFID Pakistan

Author: Freida M'Cormack, freida@gsdrc.org

Contents
1. Overview
2. Evaluating evaluations
3. Lessons from programme evaluations and evaluation design
4. Best practice guidelines
5. Sources

1. Introduction

Police reform often comes under the remit of broader security sector reform (SSR). The two are increasingly promoted in post-conflict, transitional and fragile states as a means of providing a stable environment within which wider social, economic and political development can take place. Numerous studies of police and security sector reform exist, as do comprehensive guidelines for best practice, which include sections on monitoring and evaluation (M&E). An example of this is the DAC OECD Handbook on Security Sector Reform.

Despite this, however, researchers and practitioners argue that there is very little adequate M&E of security sector reform processes (Rynn with Hiscock 2009). Hermsmeyer (2010: 6) argues that ‘the lack of systematic and meaningful evaluations of program impact and effectiveness may be the most glaring deficiency among SSR stakeholders. To the extent evaluation is performed, it tends to focus on outputs (e.g. amount of dollars spent, number of security forces trained) rather than outcomes’. He goes on to assert that ‘[m]ore work is needed on both quantitative measures of impact, including more accurate measures of readiness resulting from new training and equipment, and qualitative measures of effectiveness, such as improved public perceptions of safety and security and behavioral changes reflecting increased respect for human rights among security forces’ (Hermsmeyer 2010: 6).
Many SSR programmes make little provision for M&E during project design; some do not address the issue in project documents at all. Additionally, M&E is often under-resourced (Rynn with Hiscock 2009). Speaking specifically of community policing, Groenewald and Peake argue ‘certain challenges of evaluation remain – for instance, on the most appropriate and realistic benchmarks and indicators for measuring progress’ (2004: 9). Further, a survey of the literature on international civilian policing highlights the absence of systematic and comprehensive measurement tools and the arbitrary nature in which successes and failures are defined (Bajraktari et al 2006).

Consequently, few instances of ‘successful evaluations’ that provide examples of how to proceed with the M&E of police reform exist. Lessons can be learnt, however, from examining the challenges that faced previous evaluations. For instance, as a result of a Saferworld Project there are ‘evaluations of evaluations’ of police and security sector reform programmes. Particular programme evaluations also identify lessons for the design and execution of M&E programmes for police and security sector reform. Finally, general best practice guidelines are also available.

The report particularly focuses on efforts to measure social outcomes, principally, the impact of police reform on state responsiveness, people’s levels of trust in the state, and levels of poverty. As such, it does not include information about the numerous evaluations that evaluate police reform from a technical perspective only (e.g. performance, building capacity, financial management etc).

The research highlights key aspects for designing police reform evaluation. These include:

Ensure that a **baseline** survey is undertaken at the beginning of the programme’s implementation against which evaluators can subsequently assess findings.

**Clarity and purpose:** Ensure that higher level indicators are broken down into specific, measureable elements, such as percentage of the population who regard police presence to have a positive impact on the safety/security situation.

**Local ownership and participation:** The inclusion of beneficiaries and stakeholders external to the programme in the design and conduct of evaluations is critical to their success.

**Provide adequate time and resources:** Plenty of time needs to be built in for gathering evidence and reviewing the programme, particularly when the programme has a wide scope.

**Public opinion polling** is a valuable M&E tool, especially for measuring the development of a population’s sense of security. Such surveys can offer quantitative ‘proof’ as to whether observed changes in one area are attributable to programme activities, through a comparison with ‘control’ areas.

The **gender dimensions of policing** are an important, often overlooked, aspect. Evaluations should consider the programme’s impacts on gender roles, expectations and outcomes, including matters such as domestic violence and sexual abuse.

2. **Evaluating evaluations**

In 2009, Saferworld completed a research project, ‘Evaluating for Security: Developing specific guidelines on monitoring and evaluating Security Sector Reform interventions’. Among the questions it sought to answer was ‘What should we be measuring when monitoring and
evaluating SSR and how? The project conducted five country case studies (Albania, Bosnia and Herzegovina, Papua New Guinea, Sierra Leone and Uganda – lessons from the first four are presented here) to investigate how individual SSR programmes have been monitored and evaluated. This section draws largely on the project’s findings and papers.

This report compiles the results of a research project on the M&E of SSR programmes (which includes police reform) and provides material from which tailored guidance could be prepared to meet the needs of interested parties. It highlights a number of specific challenges in the M&E of SSR. These include the complexity of the sector and its culture of secrecy. Donor policy, bureaucracy and politics are also problematic. Participation is often an issue, with lack of sufficient stakeholder input and the fact of the fragile context in which reform is being implemented presenting its own challenges.

The report proposes eight evaluation criteria (with corresponding prompt questions) which can be used to inform the design of evaluations, taking into account the specifics of SSR programmes: relevance/appropriateness; effectiveness; efficiency; impact; sustainability and ownership; coherence; co-ordination/linkages; and consistency with values. The report also provides information on publications which provide information and advice on M&E.

http://www.saferworld.org.uk/downloads/pubdocs/SierraLeoneCaseStudy.pdf
This report assesses the monitoring and evaluation (M&E) arrangements put in place by the UK’s 1999-2008 Sierra Leone Security Sector Reform Programme (SILSEP). It identifies some challenges associated with the Output to Purpose Reviews conducted in 2005 and 2007. These include: limited availability of data, lack of local ownership, time limitations, and a lack of overall strategy or management. A key lesson is that socio-economic issues need to be factored into evaluation: the issue of how the peace dividend could be used to link up with economic development was not appropriately addressed by the review. The report notes, however, that if such issues are to be incorporated into evaluations, this must be clearly stated in Terms of Reference. It also notes that the relationship between SSR, other programmes and wider socio-economic issues must be dealt with intelligently, including perhaps via the involvement of mainstream development specialists or economists in the evaluation team. Other lessons from the review include:

- Attend to human resource and project management issues.
- Prioritise and contextualise recommendations and tie these into management planning.
- Ensure adequate inputs to evaluations: the review teams should include members with broader expertise from outside the security sector with expertise in civil society and
oversight issues, economic development and financial management, gender and ethnicity.


This report assesses the M&E arrangements for the AusAID-supported Law and Justice Sector (which includes the police) in Papua New Guinea (PNG), between 2004-2008. The Government of PNG’s sector M&E framework (PMF) comprises a limited number of Key Performance Measures (broken down into 58 sub-measures) against which each of five sector goals is measured. The five goals are: **improving policing, safety & crime prevention** (sub-measures for this include ‘public perception of police performance and discipline improves’); **improved accountability & reduced corruption** (sub-measures include the community perceives that corruption is decreasing, and that the country improves its position on the transparency international corruption index); **increased access to justice and just results**; **improved reconciliation, reintegration & deterrence**; and **improved ability to provide law & justice services**. AusAid also has its own M&E mechanism, which includes four outcomes, but the review critiques these for not focusing enough on measuring the quality and impact of activities through qualitative data.

The review highlights the following lessons for the M&E of security and justice sector reform:

- **Clearly describe roles and responsibilities**: who is responsible for carrying out and resourcing M&E and at what level (input, output, outcome or impact) should this occur?
- **Describe baseline data needs during the design phase**. Since data needs vary it may be necessary to design separate baseline studies at different levels or with different foci.
- **Use a staged approach for reliance on host information systems and data collection**. This may include an element of capacity building, including of decentralised and sub-national institutions. Specific indicators should be developed to measure capacity building (including its intangible aspects).
- **Promote national ownership** by clarifying the content and purpose of the M&E in ways that are understood by national stakeholders.
- **Link sector and programme priorities on cross-cutting issues such as gender**.
- **Create financial incentives for M&E**, for instance by linking specific indicators for M&E with performance payment assessment.
- **Clearly link programme planning and the reporting system**, which should be driven by the sector’s arrangements.
- **More focus should be placed on measuring outcomes** (as opposed to activity-focused outputs) but at a pace that is realistic given the context and the capacities of the country’s institutions.
- **Donor flexibility**: it is important that donor requirements do not overshadow those of the government and lead to a situation where the sector and agencies are overwhelmed by demands to plan and report.
- **Consider adopting a ‘theories of change’ approach** – which focuses primarily on assessing changes that have happened at the impact level and then working backwards to identify what has contributed to this.
It is important to carry out ‘true’ evaluations in addition to reviews, which are easier to carry out but tend to have a more limited scope. Tailor standard evaluation criteria (such as the OECD DAC guidelines, below) and include aspects such as ownership, participation of beneficiaries in M&E, formal/informal justice sector links and institutional capacity building.

http://www.saferworld.org.uk/downloads/pubdocs/BosniaCaseStudy.pdf
This report analyses the M&E arrangements of a community policing project in Bosnia and Herzegovina (BiH). The report focuses on how the community policing project in BiH was evaluated but it does not attempt to link the M&E of this particular project with wider safety and security concerns within BiH. The review noted that:

- No baseline survey had been undertaken at the beginning of the programme's implementation against which the evaluator could assess his findings.
- Higher level indicators are vague, and it is unclear how progress towards them is to be achieved. It would be useful to break these indicators down into more specific elements which outline how ‘police presence’, for example, would be measured in qualitative and quantitative terms, e.g. numbers of police on the streets and percentage of population who regard the police presence to be actually having a positive impact on the safety/security situation.
- Some interviewees criticised the programme for being based on assumptions about community needs and not testing these through consulting beneficiaries in programme design and M&E. They felt this posed a number of risks to efficiency, impact, sustainability and local buy-in.

The review found that it is important to validate monitoring information and rely on external views as well as internal sources. In this context public opinion polling is of value. In addition to evaluating observable changes, it also ensures the participation of external stakeholders and beneficiaries and can foster a sense of local ownership over the monitoring of over the programme. Including external stakeholders in the design process, using participatory approaches is also important.

Other lessons included the need to:

- provide adequate time and resources to M&E
- ensure clarity regarding purpose, method and focus of evaluations
- specify and define key terms
- take account of context
- widen data sources and participation during evaluations and prior to publication.

http://www.saferworld.org.uk/downloads/pubdocs/BosniaCaseStudy.pdf
This report assesses the M&E arrangements for the United Nations Development Programme (UNDP) project, ‘Support to the Security Sector Reform (SSSR) programme in Albania’. Its three
key objectives were to foster greater awareness among citizens of the post-communist police-public relationship; enhance the police’s professional capacity; and increase co-operation between the public and the police through the establishment of Community Problem Solving Groups.

Lessons from the review included the importance of:

- tying indicators to statistics used by the state police and regularly using publicly available crime and police activity statistics in reporting, monitoring, and evaluation exercises
- creating a systematic format for monitoring the programme
- conducting evaluations that are insulated from funding concerns
- conducting evaluations before expansion
- co-ordination within the organisation
- understanding the limits of research methods used in evaluations
- increasing the sources of information
- a stronger evaluation focus on the competency of the programme’s management, as individuals and as a structural system.

3. Lessons from programme evaluations and evaluation design

http://www.ssrnetwork.net/uploaded_files/2453.pdf
This paper focuses on the evaluation of two UN programmes – the minority police recruitment policy and the certification process for Bosnian police officers. Under these programmes, police reform became an integral part of the international agenda of transforming the Bosnian state and society. Although the paper suggests that the UN Mission in Bosnia and Herzegovina (UNMIBH), and its International Police Task Force (IPTF), have undoubtedly made a positive contribution to the political and economic reconstruction, and social rehabilitation, the findings question the assumption that police reform can be used to engineer broader social change and further suggests that the ‘democratic policing’ mandate could have the unintended consequence of undermining the morale and efficiency of the police.

While this paper provides a general review of policing in Afghanistan, with regard to evaluation it also highlights some problems with public opinion surveys, which are popular for gauging public perceptions about police performance. The authors found that in Afghanistan, the evidence concerning public perceptions of threats to security is mixed. Most opinion polls have been criticised as lacking robust methodologies. Some opinion polls and field interviews provide evidence that crime is seen as more important to citizens than the insurgency which preoccupies the international community. However, in other polls, Afghans in every region put the Taliban, Al-Qaida and other anti-government forces as a greater source of insecurity in their local areas than
criminals and smugglers. In the spring of 2010, perceptions of the level of security ranged from almost three quarters of Regional Command (RC) Capital respondents believing security was 'good' and only 2.5 per cent considering it 'bad', through to RC South where only 28.2 per cent saw security as good while 33.9 per cent saw it as bad. There is therefore a clear need for accurate, universally accepted data on the basis of which national and international actors can monitor progress in policing. The authors also assert that lack a shared vision of what policing should mean has undermined reform efforts.

This report looks at the role of the police and how it can be strengthened through systematic measurement and performance evaluation. It proposes an evidence-based police measurement system, PRIME (Police Reform Indicators and Measurement Evaluation), a diagnostic tool that will give the UN and donor community a systematic way to assess police reform outcomes in post-conflict environments. It was based on based on academic and exploratory field research that evaluated police reform efforts in Kosovo, Sierra Leone, and Timor-Leste and is a qualitative assessment of 16 core indicators divided into four main pillars of outcomes: i. Performance Effectiveness; ii. Management and Oversight; iii. Community Relations; and iv. Sustainability. The four indicators under the Community Relations outcome are: human rights; cooperation; corruptibility and public acceptance.

This paper draws examples from Sierra Leone, Serbia and Northern Ireland to advocate thorough contextual analysis and a needs and resources assessment to inform nuanced design, planning, implementation and M&E of the intended engagement. It asserts that the analysis should go beyond narrow considerations of police and crime to probe the broader social, economic, and political situation within which the police and communities interrelate. Gender dimensions of policing should also be considered.

The paper identifies nine features of assessment, which should support M&E planning:

- **Risk and conflict analysis**, including an assessment of social, economic and political issues.
- **Crime trends analysis**, considering crime problems of the country and their social and economic costs. Combined with and verified against other assessments and analyses, measures perceptions of crime and insecurity, and can serve as a baseline indicator by which changes in crime/criminal trends can be measured over the course of the reform period.
- **Assessment of the police organisation**, including mapping structures, resources and mandates, evaluating the level and management of police resources and considering the level of corruption and reasons for it.
- **Baseline survey of perceptions of policing**: both the police and the public should be canvassed to determine what their perceptions are of key safety and security issues, and what the police can and should be doing to address these. It should also measure perceived problems with policing, public and police perception of police priorities, public expectations of the police, and the nature of the relationship between the police and the public.

- **Review of the legal and constitutional framework**, including relevant laws governing police behaviour and legislative changes needed to implement reform.

- **Evaluation of the relationship between police and the broader criminal justice and security sectors**: the police are the most visible institution of the security sector as well as a central element of the criminal justice system. It is therefore important to recognise this role and the linkages between the police and other institutions in this sector.

- **Review of existing social and governance structures** that pertain to safety and security. It is critical to have an understanding of informal structures that deal with public safety and security, and their levels of legitimacy and (both positive and negative potential) contributions.

- **Stakeholder and leadership analysis**: a successful and sustainable reform process requires the involvement and constructive support of local stakeholders and leaders.

- **Financial assessment and viability**. An accurate assessment of the financial needs and absorptive capacity of the local government and police organisation is crucial to inform realistic decisions on resource allocation and budgeting.

In terms of evaluation specifically, the paper argues for the full involvement and utilisation of local perspectives throughout, to ensure a broader range of perspectives, give legitimacy to the results of evaluation, and provide a learning opportunity for local stakeholders taking the work forward. Medium- and long-term assessments should measure the impact of reform on safety, security, and development, as well as its impact on the pre-existing risks and root causes of conflict/fragility.

*Office of the Oversight Commissioner, 2007, 'Report 19', Office of the Oversight Commissioner for Northern Ireland, Belfast*


Appendix C of this report, ‘the policing oversight evaluation methodology’ presents a summary of the evaluation strategy for police reform in Northern Ireland, which revolved around the concept of an Independent Commission on Policing for Northern Ireland and Oversight Commissioner, considered a ‘success story’ not just as a stocktaking function, but also as a critical impetus to the process of transformation and reform.

The decision to appoint an independent Oversight Commissioner from outside the United Kingdom or Ireland (when the programme ended in 2007 this post was held by a Canadian) was made to provide independent supervision of the implementation measures. It put the spotlight on systematic public reporting that could assure the community that all aspects of the reform were being implemented or seen to be implemented. The first Oversight Commissioner established an international evaluation team of experienced law enforcement and academic executives who developed 772 Performance Indicators against which progress would be evaluated. The indicators were validated by comparison to model policing protocols, existing monitoring
programmes and an intensive peer-group review. The oversight team also conducted lengthy meetings with the heads of the agencies subject to review, to confirm that all performance indicators were fully understood.

Several elements of the Independent Commission’s recommendations coalesced to make this concept a success:

- The fact that the Commission recommended an ‘outsider’ as Oversight Commissioner was important to avoid the perception of bias
- The fact that all of the stakeholders were required to supply the Oversight Commissioner with information, and that the Oversight Commissioner was required to consult with all stakeholders, ensured that there was extensive exchange of information.
- The power of the Oversight Commissioner was merely to ‘name and shame’. It proved to be a powerful and effective tool.
- The assembly of an evaluation team who were experts in their domain was critical to acceptance and judgment on policing progress.
- This oversight process provided a change assurance process, and was important for the building of trust in a democratic policing system.

4. Best practice guidelines

http://www.dcaf.ch/content/download/35483/526183/file/tool_11.pdf

Security needs, perceptions, roles and participation in decision-making differ according to socio-cultural gender roles. The integration of gender issues into SSR assessment and M&E processes, in addition to being mandated by international and regional laws and instruments, can strengthen the delivery of security and justice services, support participatory SSR processes and build non-discriminatory, human rights promoting, and representative security sector institutions. General areas for gendered M&E include:

- changes in the incidence of gender-based violence
- access to public security services for women, men, boys and girls
- access to decision-making positions
- public awareness of gendered insecurities
- recruitment, retention and position of female staff
- the impact of gender training.

In terms of processes, the toolkit advocates results-based management using clear, measurable and context-based gender-sensitive indicators.

This guidance suggests that when reviewing and evaluating programmes and projects, it is useful to consider the following DAC criteria: relevance, effectiveness, efficiency, impact and sustainability. Important areas to consider include: building local ownership; managing information to inform evaluations; and ensuring findings inform revisions.

Lessons from reviewing UK support for SSR in Sierra Leone include: use multidisciplinary teams for reviews; assess coherence with wider initiatives; and move beyond institutions and assess impact.


This section of the handbook provides practical guidance on M&E of SSR programmes. It argues there are three key factors involved in deciding what to measure: defining feasible and measurable impacts; identifying ‘dimensions of change’ and which are the most important; and agreeing among stakeholders on the changes the programme is seeking.

With regard to measuring impacts, realism is needed about what the programme can achieve, how the programme contributes to wider goals, and the context in which these goals can be achieved. Methods such as impact assessments, contribution analysis and participatory evaluation may help evaluate whether SSR programmes are making a contribution to such goals as peace, security and democratic governance. The section also provides further resources for indicator development in the security system.

7. Sources

References


**Key websites**
OECD-DAC Evaluation Resource Centre (DEReC)
http://www.oecd.org/findDocument/0,3770,en_35038640_35039563_1_35074428_1_1_1.00.html
Global Facilitation Network for Security Sector reform (GFN-SSR)
http://www.ssrnetwork.net
International Security Sector Advisory Team/Democratic Control of Armed Forces
http://issat.dcaf.ch

**Experts consulted**
Megan G. Kennedy-Chouane, OECD Development Co-operation Directorate
Peter Albrecht, Danish Institute for International Studies
Rohan Burdett, Coffey International
Paul Jackson, Birmingham University
Piet Biesheuvel, Coffey International
Kristie Drucza, Independent Consultant
Kenneth Annett, Police Service of Northern Ireland (retired)

---

**About helpdesk research reports:** Helpdesk reports are usually based on two days of desk-based research. They are designed to provide a brief overview of the key issues; and a summary of some of the best literature available. Experts are contacted during the course of the research, and those able to provide input within the short time-frame are acknowledged.