



RESEARCH REPORT

Survey of the United States Government's arrangements for monitoring and evaluating support to security sector reform

By Jake Sherman

January 2009

1. Introduction

This report provides an overview of the United States Government's arrangements for monitoring and evaluating (M&E) the support it provides to security sector reform (SSR). It examines the M&E systems that already exist for similar types of work as well as looking at any specific treatment given to SSR, before also identifying outstanding needs, challenges and any trends and opportunities that exist for improving M&E in this area.

This is one of five donor surveys carried out as part of the Saferworld project 'Evaluating for Security: Developing specific guidelines on monitoring and evaluating Security Sector Reform interventions'.¹ While the donor surveys unpack the institutional arrangements for M&E within the major SSR donors, they do not provide a detailed analysis of how this is carried out at the country level in particular cases. Instead, five separate country case studies have been carried out to investigate how individual SSR programmes have been monitored and evaluated. Together with a wider desk review, the case studies and donor surveys provide the evidence base from which specific guidance on monitoring and evaluating SSR can be developed.

Interviews for this case study were conducted during August 2008 with individuals at the following US Government (USG) departments and agencies: the US Department of State (DoS),² Department of Defense (DoD)³ and the Agency for International Development (USAID)⁴ – the three primary US Government entities responsible for SSR; and the Government Accountability Office (GAO)⁵ – the investigative arm of Congress. Interviews were also conducted with staff from the US Institute of Peace (USIP), an independent, Congressionally-funded institution that has been involved in the development of indicators for post-conflict stabilisation, with an independent consultant on rule of law, criminal justice, and post-conflict public security, and with two anonymous employees of DynCorp, International. (See annex). The M&E terminology used in this study is consistent with that used by the US Government and wherever appropriate, footnotes have been used to elaborate some specific terms in more detail.

The US Government – unlike the United Kingdom, Netherlands, and United Nations – does not yet have an official definition or statement of policy for SSR, though a draft inter-agency document is

¹ Other donor surveys include the institutional arrangements for M&E of SSR programmes by the, UK and Dutch Governments and the UN and EU.

² The Regional and Security Affairs/Bureau of African Affairs and the Office of Plans, Policy and Analysis, Bureau of Political-Military Affairs.

³ The Office of Secretary of Defence/International Security Affairs/Africa and the OSD/Global Security Affairs/ Partnership Strategy.

⁴ The Office of Democracy and Governance and the Haiti Stabilisation Initiative.

⁵ The International Affairs and Trade Team (IATT).

currently under consideration. In the absence of such a policy, most USG agencies and departments use the OECD definition – if they refer to SSR at all – which the USG endorsed in 2004. Up until very recently, SSR projects and activities undertaken by USG agencies and departments have not been called 'SSR'. The term is gaining use, but there is still a lot of conceptual clarity needed, for example, whether rule of law and justice work are SSR, or vice versa. According to USAID, SSR encompasses four pillars: police, small arms and light weapons, reintegration and civil/military relations. It notes that SSR includes efforts to 'improve judicial and penal systems, reorient police and similar bodies, and upgrade the ability of elected and appointed civil authorities, such as legislatures and the Executive and the Defence Ministries, to provide oversight and civilian control.'⁶

DoS, DoD, and USAID represent the '3Ds' of diplomacy, defence, and development, respectively. When the US Congress first established Foreign Assistance, it split funding between economic (USAID) and military (DoD); by law, no overlap is permitted between the two. USAID, which is legally prohibited from engaging in defence aspects of SSR work, has responsibility for activities relating to rule of law, including supporting legal reform, improving the administration of justice, and increasing citizens' access to justice. DoD assistance focuses primarily on training and equipping foreign militaries. The precise nature of these restrictions, however, is not spelled out anywhere in a policy or guidance document. As the US has no national police service and its Department of Justice is predominantly domestic in focus, support to policing initially fell through the cracks. The 1994 US intervention in Haiti began to change this, with DoS taking the lead on overseas police assistance. These divisions are increasingly blurred, however, in the context of US stabilisation missions. Due to 3D policies, eighteen percent of US Official Development Assistance is now channelled through DoD.⁷

Within each of the three main agencies or departments, programme authority is decentralised down to the embassies/missions at the country-level. Responsibility for programme M&E therefore, lies at this same level. Moreover, at DoS and USAID, the employment of relatively few programme staff (whether for SSR or other sectors) has led to a practice of programme implementation through outside contractors (who are required to monitor and contract independent evaluation of their work). In the case of DoD, implementation of defence assistance is undertaken through the regional 'Combatant Commands'. Co-ordination of SSR activities by different departments and agencies occurs primarily through the embassy country team (for example, the US Ambassador, Defence Attaché, and USAID chief of mission).

The decentralised nature of programme implementation made conducting interviews with those directly engaged at the country level in SSR programme execution and routine monitoring difficult. This case study drew more on policy level actors and is thus focused predominantly (though not exclusively) on evaluation.

Department of State

Within the DoS,⁸ SSR priorities and programmes are identified by US embassies/missions in the field, in co-operation with regional bureaus at headquarters (for example, African Affairs, Near East, and South and Central Asia) and with the technical bureaus, particularly the Bureau for Political-Military Affairs (PM). PM is responsible for providing policy direction for international security, security assistance to foreign governments, military operations, and defence trade.

⁶ See http://www.usaid.gov/our_work/cross-cutting_programmes/conflict/focus_areas/security.html, accessed 11 December, 2008.

⁷ For example, 'Commanders at the provincial and local levels use their Commanders' Emergency Reconstruction Programme (CERP) funds for projects to achieve rule of law objectives. CJTF-82 provides training to prosecutors and judges as well as logistical support for training and for the distribution of USAID publications. Military units at PRTs deliver USAID materials and some of their own creation to legal professionals and the general population.' OIG Report No. ISP-I-08-09, Inspection of Rule-of-Law Programmes in Afghanistan, January 2008, p 44.

⁸ For a concise overview of SSR definitions, authorities, funding accounts, and programmes within the US Department of State, see http://www.ndu.edu/ctnsp/Stab_Ops/Rosati%2016%20Apr.pdf, accessed on 09 September, 2008.

Through its Office of Policy, Plans and Analysis (PM/PPA), the Bureau is responsible for directing finances for US military assistance – some \$5 billion annually – to foreign governments through policy development, budget formulation, and programme oversight. PM/PPA is the responsible entity within DoS for security sector policy and is working to facilitate an integrated approach to SSR across regional and functional bureaus of the Department in order to improve joint planning and execution. This includes participating in assessment missions for SSR and providing advice on SSR programmes at the country-level. PM also co-operates with the DoD so that the latter's planning and military activities are consistent with US foreign policy objectives.⁹

The technical bureaus are responsible for global programmes and those that are not confined to a particular geographic region. Where a technical bureau is in charge of programme execution – for example, the Bureau for International Narcotics and Law Enforcement Affairs (INL) – that bureau is responsible for monitoring and evaluation through its representative in the embassy. INL is increasingly the lead agency within the USG for rule of law activities. Few INL staff at headquarters level are law enforcement professionals. Advisors with the necessary technical expertise are hired and deployed for specific projects, such as training police and counter-narcotics forces in Afghanistan, and operate with a high degree of independence from headquarters. Programme implementation is conducted through outside contractors, for example, DynCorp International and PAE.

Among the regional bureaus, the Bureau of African Affairs (AF) – particularly its Regional and Security Affairs (AF/RSA) office – has acquired considerable experience in designing, implementing, and assessing SSR programmes in Burundi, DRC, Liberia, and south Sudan. Since M&E of SSR activities is the responsibility of the individual embassy/mission, as noted above, AF – like other regional bureaus – does not have specialised evaluation staff. However, RSA does help embassies to develop performance indicators and track progress at programme level. RSA programme staff have also participated in inter-agency SSR assessment teams to countries receiving assistance. RSA co-operates with the SSR specialists within the Bureau for Political-Military Affairs and in USAID's Office for Democracy and Governance (ODG). PM and ODG each have one senior SSR specialist. Reflecting the absence of an official policy framework for SSR and the relative newness of the concept, there are few explicitly SSR-focused staff within the USG as a whole – in fact, USAID's 'Senior Security Reform Advisor' in ODG may be the only position with SSR in the title).

Department of Defense

The Office of the Secretary of Defense for Partnership Strategy (OSD/PS) provides overall global SSR policy guidance on restructuring and reform of foreign armed forces and defence sectors. These activities fall into four areas: i) operational capacity building, for example, the 'Global Train and Equip' programme¹⁰ and 'Overseas Humanitarian, Disaster and Civic Assistance'; ii) human capacity building, including the 'Regional Centres for Security Studies' and 'Combating Terrorism Fellowship Programme'; iii) institutional capacity building (for example, National Guard State Partnership Programme) and iv) funding provided to the DoS for whole-of-government efforts – 'Security and Stabilisation Assistance' – under Section 1207 of the National Defense Authorisation Act (NDAA) of fiscal year (FY) 2006 and FY 2007 and Section 1210 of NDAA FY 2008. With OSD, the geographic offices of the Assistant Secretaries of Defence are responsible for determining how

⁹ For example, under the 'Foreign Military Financing' account – one of the USG foreign military assistance programmes historically administered by State – a US embassy will make a recommendation on how much should be allocated, which then flows up to State and DoD (the latter being the primary implementer for military assistance) for actual allocation and prioritisation. PM is also responsible for the International Military Education and Training (IMET) programme, the Peacekeeping Operations account, which includes the African Contingency Operations Training and Assistance (ACOTA) programme, and it co-ordinates DoS involvement in 'Section 1206' programmes, under which DoD funds are used to improve foreign military capacities in counter-terrorism or support to US-led stability operations.

¹⁰ Section 1206 of the National Defense Authorisation Act (NDAA) of FY 2006 and FY 2007.

to implement high-level policy in communication with the Combatant Commands (for example, CENTCOM, AFRICOM) and the USG country team in the embassy.

The Combatant Commands implement policy for SSR undertaken by DoD in theatre in much the same way that DoS and USAID activities are led by the embassy/mission. These activities fall into two broad categories – ‘steady-state’ assistance in non-post-conflict or transition countries (for example, Liberia), and ‘stabilisation’ assistance where US forces are directly involved in helping countries to emerge from conflict (Afghanistan). For security, stabilisation, and reconstruction activities under the ‘1207’ and later ‘1210’ programmes, proposals for service or funding are submitted to the policy level (the State Department’s Co-ordinator for Reconstruction and Stabilisation (S/CRS)) by embassies, DoS regional bureaus, USAID, or Combatant Commands, but ‘all proposals must be closely co-ordinated with the affected Embassy and submitted by the Ambassador’.¹¹ S/CRS and DoD’s OSD then jointly recommend which proposals should be funded. All proposals have to be cleared with the relevant Combatant Command.

US Agency for International Development

USAID, like DoS, is structured along regional and technical offices with a high level of decentralisation of programme authority down to the embassy or mission at the country level. The USAID head of mission and the US ambassador work together to draw up the country team’s strategic plan for engagement in any given country (then approved by principles at the policy level). The strategic plan covers *all* embassy activities (this includes: USAID assistance, DoS and DoD security assistance, trade affairs, consular affairs, etc.). Programme level SSR decisions are therefore made at the mission level, and often implemented through contractors.

Technical offices at headquarters, including the Office of Democracy and Governance – responsible for rule of law programming and where the Senior SSR Advisor is located – have no direct programme authority, serving instead in an advisory capacity. ODG undertakes country assessment, helps with project design, assists with identification of implementers, and provides technical assistance. The Senior SSR Advisor also has a role at the policy/strategic level, setting the SSR agenda and shaping debate both within USAID and the interagency process involving DoS and DoD.

USAID is at a comparative disadvantage, relative to its counterparts in other donor countries – for example, DFID in the United Kingdom, which has the balance of power and project autonomy. In post-conflict settings, the US relies heavily on military, rather than civilian instruments, for implementation of security assistance, which is focused on a more narrowly defined security sector than in other major bilateral donors supporting SSR.

Government Accountability Office

The GAO is mandated by Congress to review any and all federal programmes, including – evaluating whether they are achieving their objectives. The GAO is not an implementing agency, so it only undertakes evaluation, not monitoring. It has no authority to set SSR priorities or shape programmes, as its Congressionally-mandated role is to audit programmes of other USG agencies and departments. One assistant director in the International Affairs and Trade Team (IATT), which has audit authority for security assistance programmes, stated that the “GAO is likely to look at all large [SSR] programmes – and they may not be as large as Afghanistan/Iraq. In the past, we’ve looked at US assistance in Bosnia, Pakistan, Sudan, and the West Bank”.¹² GAO is responsible for the design and execution of the evaluations, which are conducted at the programme rather than project level.

¹¹ Robert Perito, ‘Integrated Security Assistance: The 1207 Programme’, *United States Institute of Peace Special Report No. 207*, July 2008, p 15, (Appendix 2: Guiding Principles).

¹² Interview, GAO official, 22 August, 2008.

2. Existing monitoring and evaluation arrangements

Monitoring¹³ and evaluation¹⁴ structures, practices, and attitudes

Since the 2006 foreign assistance reform, which established the Director of US Foreign Assistance (concurrently the USAID Administrator) to better align aid and policy, DoS and USAID have jointly developed standard foreign assistance ('F') indicators for measuring the success and impact of USG foreign assistance funds. Since FY 2007, each USAID mission and US embassy is required to report against the F indicators. Prior to the reform, the two agencies developed and tracked separate indicators, which made it difficult to get a complete picture of results. Today, the common indicators, published in a master list,¹⁵ are divided out according to the US foreign assistance framework ('peace and security' and 'governance and democracy' being the two categories under which the majority of SSR programme activities fall). These indicators are broken down into strategic, programme-area, and 'element' levels. Programme area indicators measure results beyond what is achievable solely through the USG programme (that is they include other donors and host government activities), while element level indicators are of outputs directly attributable to USG activities.

According to the DoS,

'Data are collected primarily by implementing partners, and targets are set by USG agencies and their partners against these indicators on an annual basis. Information on standard indicators at the objective and programme area levels will be collected in Washington. Operating Units will collect data and report on indicators at the programme element level through Operational Plans and Performance Reports. These standard indicators are complemented by 'custom indicators' that are selected by each Operating Unit to measure and monitor performance in achieving results that are critical to the attainment of foreign assistance objectives by the particular bilateral, regional or Washington-based programme. Such custom indicators should be included in the particular Operating Unit's performance management plan, and be reported as appropriate in the annual Operational Plan and Performance Report.'¹⁶

Core components of SSR are divided among the standard indicators for 'peace and security' and for 'governing justly and democratically'. The indicators for governing justly and democratically, for example, are divided into four programme areas: rule of law and human rights; good governance; political competition and consensus building; and civil society, which are then further subdivided. The standard indicators are overwhelmingly quantitative in focus – for example, number of justice sector personnel that received USG training, number of legal institutions and associations supported by USG. This presents a narrow reading of what is being accomplished;

¹³ USAID is the only USG department to provide a definition for monitoring; 'systems to track and alert management as to whether actual results are being achieved as planned. Monitoring is built around a hierarchy of objectives logically linking USAID activities and resources to intermediate results and strategic objectives through cause-and-effect relationships. For each objective, one or more indicators are selected to measure performance against explicit targets (planned results to be achieved by specific dates). Performance monitoring is an ongoing, routine effort requiring data gathering, analysis, and reporting on results at periodic intervals'. See, <http://dec.usaid.gov/partners/evalweb/resources/definitions.cfm#diff>

¹⁴ USAID is the only USG department to provide a definition for evaluation; 'systematic analytical efforts that ask why certain results are being achieved. They are planned and conducted in response to specific management questions about performance of USAID-funded development assistance programmes or activities. Unlike performance monitoring, which is ongoing, evaluations are occasional-conducted when needed. Evaluations not only focus on why results are or are not being achieved, they may also address issues such as relevance, effectiveness, efficiency, impact, or sustainability. Often evaluations provide management with lessons and recommendations for adjustments in programme strategies or activities'. See, <http://dec.usaid.gov/partners/evalweb/resources/definitions.cfm#diff>.

¹⁵ See, www.state.gov/f/indicators.

¹⁶ *Ibid.*

furthermore, it is unclear that the F indicators alone effectively demonstrate 'the collective impact of foreign and host-government efforts to advance country development', since without additional qualitative indicators the impact of, for example more judges, on the overall performance of the judicial system cannot be determined. One respondent acknowledged that the F indicators are still relatively new and 'a work-in-progress'.

USAID is recognised as the leader among USG agencies and departments in performance measurement. The purpose of its M&E is to assess whether activities are achieving their intended purpose, to learn from and improve the effectiveness of its programmes and projects, and to share these findings with others in the development community. Accordingly, USAID requires every project (including SSR) to have an M&E plan, and each USAID country mission is required to establish targets for and report against multiple sets of indicators. This includes each mission's own performance monitoring plan (PMP), which identifies how performance data will be collected. The Agency's M&E requirements focus more on project effectiveness rather than on the overall contribution to stability goals, since this "requires taking into account impact of more projects than just those of one single Agency".¹⁷ According to one official, the organisational structure and manner in which programme funding is allocated creates obstacles to USAID and other USG departments from looking beyond their specific sector to the broader strategic level. The Measuring Progress in Conflict Environments project – described below – attempts to measure the latter. These indicators may vary widely among missions depending on the strategy and programmes of each.

ODG's 'Handbook of Democracy and Governance Programme Indicators'¹⁸ remains the definitive agency guide for identifying indicators and establishing baseline data for rule of law and governance projects. It is frequently used by USAID contractors to design their monitoring and evaluation plans, though the indicator list "does not always fit with the reality on the ground."¹⁹ For example, the justice indicators disproportionately concern the formal legal system; there is one indicator concerning the number of cases using 'alternative systems' despite the fact that the majority of people may use such systems. These monitoring and evaluation plans are designed solely by the contractor based on existing USAID guidelines and requirements; there is little direct interaction between USAID programme or evaluation staff programme and contractors. Staff from the relevant technical and regional bureaus comprise the review panel that awards the contract. It was noted that it is hard to generalise about the relationship between programme staff at headquarters or the embassy and the contractor during the project implementation phase because it varies from country to country, as well as from agency to agency. In some places, the contractor is very loosely overseen at the country level with no oversight from the HQ level, but the other extreme exists as well.²⁰

USAID openly publishes its guidelines on programme assessment and learning.²¹ It has standard requirements on reporting and evaluation (quarterly, annually, as well as third-party evaluations at the project mid-term and on completion). Reports must be provided to and are posted on the publicly accessible USAID Development Experience Clearinghouse database (www.dec.org).²² However, in the view of one respondent, the incentive for contractors to make their evaluations publicly available – despite the requirement – is often lacking, as this "would provide information

¹⁷ Interview, USAID official, 18 August 2008.

¹⁸ USIAD, 'Handbook of Democracy and Governance Programme Indicators,' August 1998 via http://www.usaid.gov/our_work/democracy_and_governance/publications/pdfs/pnacc390.pdf.

¹⁹ Interview, confidential source, 22 September, 2008.

²⁰ *Ibid.*

²¹ <http://www.usaid.gov/policy/ads/200/203.pdf>.

²² The database is easily searchable for 'SSR'; a quick scan for evaluations using the key words 'police,' 'police services,' 'parliamentary systems,' 'judiciary,' and 'judicial reform' and 'security' under 'democracy and governance' yielded 155 hits, including evaluations of 'justice sector development' in Bosnia, 'justice system reform' in Kosovo and Honduras, 'judicial reform activities' in Kazakhstan, community-police relations in Mindanao, and community policing in northern Uganda. However, there appears to be a lack of consistency in terms of what types of evaluations are available across projects, as well as in evaluations over the life of the project.

to the competition when contractors want to ensure that they are awarded the contract if it is renewed.”²³The enforcement by USAID of the requirement is also often weak.

DoS is trying to encourage a more evidence-based approach to its programming. In contrast to the ‘social sciences’ approach used at USAID, one respondent described the traditional DoS approach as less rigorous, particularly within INL. In contrast with USAID, baseline measurement has been less common and there is not yet any public repository for DoS evaluations. In fact, it was unclear with what regularity DoS required evaluations of its projects.

Awareness that measuring progress is important is growing both in terms of the need to do it and what measures are appropriate. AF/RSA commented that it has responded to critiques that it needed to require contractors to include programme indicators and is now undertaking a rebidding of its major projects. It was estimated that the requirement is currently included in 60-70 percent of contracts, but as a result of the rebidding process, will cover 100 percent.²⁴ In the past, DoS tended to see output measures as definitive, driven largely by the need to show that public money spent was well spent. This was influenced by the fact that quantitative measurement tends to be easier than qualitative measurement, which tends to focus on outcome. Nonetheless, the focus is now shifting increasingly towards measuring the latter. This is a positive development and interviewees were able to point to some examples of where this had proved beneficial.

DoS’s Africa Bureau noted positive results in programme outcomes as a result of feedback on assessments. The case of Liberia was highlighted, in which joint evaluation results enabled mid-course corrections in SSR assistance. The bureau realised that, in addition to US assistance to the Liberian military, police programming needed to be stepped up if the overall outcome of achieving prosperity was going to be achieved.

By comparison, formal monitoring and evaluation traditionally has not been part of DoD culture. That said, one respondent pointed out that “DoD – whether through the defence attaché in the country team or at the policy level – is always in monitoring mode. We don’t have regular monitoring as with outside contractor work orders, but we are in constant communication and have a finger on the pulse of what is happening”.²⁵ There are two layers of performance measurement by DoD: first, as suggested above, by the defence attaché and security assistance advisor in the embassy and, second, by the Combatant Command. The Office of Deputy Assistant Secretary of Defence for African Affairs, for example, regularly receives, reviews, and provides feedback on information concerning defence assistance projects from individual embassies and from AFRICOM – creating a ‘feedback loop’. At the same time, ODS/PS has established a ‘tiger team’ to look at SSR programme assessment, including M&E. The initiative is still in development stage and likely to roll out in early 2009.

The intention behind DoD M&E has predominantly been to measure performance and ensure accountability (that is making sure that public money is being used wisely), but the Department is currently trying to shift towards learning – “not something that DoD has been good at.”²⁶ DoD has also traditionally been resistant to outside evaluations. One problem that self-assessment has generated – one acknowledged by many within the Department – is ‘grade inflation’ (overstating results).²⁷ Hence, existing measurement practices are recognised as, at times, not having much value, as they provide little accurate data for appropriate resource allocation.

²³ Interview, confidential source, 22 September, 2008.

²⁴ Interview, US Department of State official, 27 August, 2008.

²⁵ Interview, US Department of Defense official, 09 September, 2008.

²⁶ *Ibid.*

²⁷ An official federal history of US-led reconstruction efforts in Iraq notes that ‘the Pentagon simply put out inflated measures of progress to cover up the failures’ in rebuilding the Iraqi police and army. James Glanz, ‘Official History Spotlights Iraq Rebuilding Blunders,’ *The New York Times*, December 13, 2008, p A1.

According to one DynCorp police advisor, there was only sporadic, *ad hoc* monitoring of police training programmes in Iraq due, first, to the fluid and volatile nature of the operating environment and, second, because the slow deployment of advisors meant they were always behind the curve. A second DynCorp advisor cited routine reports – ‘police station monthly reports’, ‘daily station’ or ‘daily situation’ reports, and weekly ‘storyboard’ or ‘good news’ stories. He noted, “all these forms have some merit but all of them lack the ability to demonstrate, in qualitative fashion, that real progress is being made.”²⁸ Many international police advisors in Iraq are also resistant if not hostile to rigorous supervision; not least due to command by inexperienced, junior military officers. International police advisors have been under the command of DoD for nearly five years; often this has meant that senior police officers on contract as trainers were under the command of officers with no experience in policing, who therefore had little understanding of the objectives and components of police.

Moreover, the metrics²⁹ that are being developed – such as the number of weapons issued to police, number of arrests made (but not number of convictions), numbers of bombs detonated, incidence of violence – can all be manipulated. For example, an incident of violence may be classified as ‘anti-Coalition’ when it may be a tribal dispute. Similarly, using arrests is problematic, as detaining large numbers of men who are later released does not measure anything significant.

The DoS and the DoD have published formal guidelines for 1207 project proposals, including a description of project monitoring and evaluation. DoD, however, has no direct responsibility for 1207 projects since funding is transferred to State and USAID (those organisations are responsible for M&E). DoD, ‘fully supports and requires M&E as a component of each 1207 proposal and hopes to start requiring implementers to build M&E into their 1206 [military assistance]³⁰ projects. Here, again, the relationship is indirect since implementers are military Service components (e.g., US Army) or contractors. Hopefully, adoption of the State-DoD-USAID policy paper on SSR will provide the impetus to conduct greater M & E’.³¹

As stabilisation programmes have increasingly come to dominate US foreign security assistance, the DoS and DoD Offices of the Inspector General³² – and, outside of the executive branch, the GAO – have an increasingly important role in reviewing programmes, systems, and issues related to security assistance, particularly those funded under the 1207/1210 programme:

‘The OIG and the GAO are two independent bodies that also drive evaluation within the DepartmentOIG also conducts reviews of specific programmes, grants, and contracts at the request of the Department. These reviews and evaluations provide the Department an objective assessment of programme performance and recommend specific actions to be taken in meeting the challenges ahead.’³³

²⁸ E-mail correspondence with author, confidential source, 14 December, 2008.

²⁹ The US Government uses ‘system of metrics’ interchangeably with ‘measurable indicators’.

³⁰ Section 1206 of the FY 2006 NDAA. Section 1206 provided the DoD with a funding stream ‘for non-traditional security assistance to train and equip foreign military forces in counter-terrorism, capacity building, stabilisation, reconstruction, and humanitarian relief. The provision was intended to enable combatant commanders to assist countries threatened with terrorist infiltration without re-programming already allocated funds or waiting until Foreign Military Financing (FMF) became available.’ Perito, *op cit*, p 3.

³¹ Interview, US Department of Defense official, 19 August, 2008.

³² USAID also has an OIG.

³³ Appendix B: Department of State Programme Evaluation Plan, <http://www.state.gov/s/d/rm/rls/dosstrat/2004/23510.htm>, accessed 21 August, 2008. Certain State Department SSR programmes have also been subjected to reviews by the White House Office of Management and Budget (OMB) Performance Assessment Rating Tool (PART). The AF Bureau has had to report on how well it is moving ahead on performance measures for its Security Assistance to Sub-Saharan Africa programme – for example, ‘the number of corps deployed on peacekeeping missions by African partners’ – to gauge the impact of US-supported training programmes, as well as broader metrics on conflict-management indicating progress from conflict to post-conflict situations (e.g., for border security, number of forces stood up). See Office of Management and Budget, <http://www.whitehouse.gov/omb/part/>, accessed August 27, 2008.

As the Congressional watchdog, the structure with GAO is specific to its evaluation and auditing mandate. GAO has 12 'engagement teams', which are broad portfolios – for example, development, trade, infrastructure. Reviews of security programmes fall under the responsibility of the International Affairs and Trade Team (IATT), which has a total of 140 staff. As GAO is an auditing office, there is no specific 'evaluation unit'; rather GAO analysts are interchangeable within the engagement team. A typical evaluation team consists of 2-3 people – usually an analyst-in-charge (AIC) and one or two supporting analysts – who are assigned an audit on an *ad hoc* basis. Assignments are issued by the directors during their monthly review of what audits are on docket, who is available, and with what expertise. Generally, a project (especially a one-off) lasts six months to one year, but may be longer on an exceptional basis.

All GAO evaluations – and those of individual agency and department Inspectors-General – are guided by the Generally Acceptable Government Accountability Standards (GAGAS), commonly referred to as 'the yellow book'.³⁴ GAGAS provides across the board guidance, but is not sector specific. In addition to written protocols on engagement with agencies being audited, GAO uses the Electronic Assistance Guide for Leading Engagements (EAGLE), which provides analysts with online access to policy and operational guidance for conducting assessments. The methodology for individual evaluations is designed by the analyst-in-charge and supporting analysts, but must be approved by the Applied Research and Methodology (ARM) unit of GAO, as well as by the directors, assistant directors, and stakeholders (who may be agency or department staff who provide expertise, for example, on weapons systems). This ensures that there is no duplication with evaluation efforts currently underway by the agency or department on the programme being reviewed. GAO has a team that covers strategic issues that identify criteria and key practices – for example, performance indicators, best practices, that programmes should have.

The purpose of GAO evaluations is to report to Congress on the effectiveness of programmes in meeting their intended outcomes. According to a GAO official, "it would look at whether the goals for a programme were overly ambitious or lax given the operational environment and the needs. In case goals have not been met, we would look at factors and challenges that the implementing Agency faces and highlight these. We may also be critical if the goals were not changed despite changes in the context."³⁵

The decision to evaluate a programme may be taken: i) if Congress enacts legislation specifically requesting a GAO review; ii) if a Committee or Sub-Committee Chair or Ranking Member requests a specific review; or iii) under Comptroller General Authority (CGA), when an issue is deemed of such importance that the GAO may decide to undertake an evaluation on its own. Reports are sent to the requesting Committees – or, for CGAs distributed to the Appropriations and other relevant Committees – and made public.

There is wide variance in the acceptance of agencies and departments to being audited by GAO. GAO has statutory authority to documents from all programmes funded by Congress. Technically, the GAO's statutory authority does not include speaking to those involved in policy and implementation, but, according to one GAO official, most agencies recognise the utility of being subjected to review, both in terms of improved programme performance and conveying the message of effectiveness to Congressional committees responsible for oversight and appropriations. Personality of those involved on both sides was cited by one GAO official interviewed as a possible facilitator or hindrance to good co-operation.³⁶ USAID was singled out in particular for its recognition of the benefits of working with GAO. At times, there is a split in attitude between the policy/headquarters level and field mission level. Ultimately, very sensitive programmes are liable to get greater pushback, making it more difficult for GAO to get access to information. Its standing has been dealt a blow by the current administration following a Supreme

³⁴ See, <http://www.gao.gov/govaud/ybk01.htm>

³⁵ Interview, GAO official, 18 August, 2008.

³⁶ *Ibid.*

Court ruling challenging the authority of the legislative arm via the GAO to investigate the executive, despite this being the intent of the GAO.³⁷

GAO rarely sets indicators. Instead, it evaluates the extent to which executing agencies and departments are tracking their own progress against the indicators mandated by legislation or the programme indicators set by the department. GAO will look at whether goals for a programme were overly ambitious or lax and that DoS/DoD have sufficient risk mitigation plans; it is understood that it will take longer to get things done in stabilisation situations, for example, if staff rotation is higher due to hardship: "In case goals have not been met, we would look at factors, challenges that an agency faces and highlight these. We may also criticise if goals have not changed despite the context."³⁸ GAO also follows up on how its recommendations are implemented by agencies. Informally, when GAO starts looking at a programme, it asks agencies for their metrics: "Maybe the agency doesn't have any when we start, but by the time of the audit, they've implemented metrics. Also, just being out there looking at programmes – and the possibility of us returning – has an impact".³⁹

Local Ownership

Local ownership is recognised as a critical component of sustainable SSR activities by those interviewed during this research. It is included as a principle of the draft policy statement on joint SSR by USAID, DoD and DoS. Yet, there are different views among the departments about what local ownership means and the extent to which programmes and projects will result in it. It is not clear whether these tensions are readily solvable. Local ownership in SSR is also viewed differently from USG engagements in other sectors. SSR programming is largely driven by and undertaken from the perspective of US national security, rather than development. Indeed, SSR assistance is often intended to primarily further US national or global security – for example, training SWAT-type police that can partner with US personnel on counter-narcotics missions. The US may not necessarily be developing the capacity of partner states to provide safety and security to their own citizens.

USAID is considered the biggest supporter of local ownership – not surprising given the agency's mandate for development, the field that has been the primary supporter for greater engagement with local partners. As one USAID official stated, "USAID is seeking to work itself out of the job." But USAID is by no means alone in supporting local ownership. DoS AF/RSA recognised that "we cannot just train and equip our partners' military or police forces – we need to have institutional reform, too – improving systems (training, management, logistics, etc.). Sustainability – statebuilding – is critical if partners are going to be given the opportunity to become safer, more stable, secure, and prosperous for government and civil society."⁴⁰ This was echoed by DoD's OSD/PS: "If you can't maintain the logistics chain, training, and troops, then the assistance effort has been wasted."⁴¹ However, it was also noted that support for the inclusion of local ownership as an outcome is not always present at the policy level – decision-makers may have neither the appetite or financial/human resources for longer interventions required to actually achieve local ownership. As a result, DoD projects tend to have short-term objectives not because longer-term objectives are overlooked. Nonetheless, there was consensus that most actors involved in SSR understand the importance of local ownership.

The GAO does not measure local ownership unless it is already a performance measurement of the agency being reviewed. Nonetheless, sustainability – which a GAO official defined as whether

³⁷ *Richard B. Cheney, Vice President of the United States et al., vs. United States District Court for the District of Columbia et al.* <http://supct.law.cornell.edu/supct/html/03-475.ZS.html>, accessed 12 December, 2008.

³⁸ Interview, GAO official, 22 August, 2008.

³⁹ *Ibid.*

⁴⁰ Interview, US Department of Defense official, 27 August, 2008.

⁴¹ Interview, US Department of Defense official, 21 August, 2008.

"local people can run the programme and local financing exists to support it"⁴² – has become a central feature of recommendations coming out of GAO audits. The office cited concrete examples of progress in Colombia, where US had provided helicopters, maintenance, and pilots as part of a programme supporting counter-narcotics. Monitoring the impact of its own report, GAO found that four to five years later, its recommendations on building in greater sustainability had been adopted, resulting in a transition towards a more nationally-led process. Similarly, in Afghanistan, the Afghanistan National Army (ANA) has also started to focus more on sustainability, shifting from reliance on external support to developing 'enablers' like its own airlift capacity. While GAO commented that this change was not conclusively attributable to its reports, DoD was clearly giving greater attention to sustainability in its implementation of SSR programmes supporting the ANA.

SSR: Same as other sectors?

The approach to performance measurement by different agencies and departments to SSR is generally the same as across other sectors, though this varies by degree from one to the next. For example, in the selection of evaluators, USAID's outside contractors (according to an internet-based scan) seek M&E experts with sector experience, such as on rule of law and governance. The GAO, when undertaking an assessment, by contrast draws on its pool of analysts in the IATT engagement team on an *ad hoc* basis. Each assessment undertaken by GAO has its own methodology specific to the nature of the programme being assessed, as well as the timeframe and questions sought by Congress. However, GAO staff also stated that the process for determining the methodology and the guidelines governing how the audit is conducted are the same whether it is for security assistance or another sector.

Performance measurement is obviously more difficult if undertaken in a war zone – but this is true whether the programme is food aid or security; there are likely few, if any, records, as well as issues of mobility, and language ability that complicate the ability to visit sites, to solicit views from stakeholders. These constraints may also hinder the use of certain methodologies, such as random sampling.

However, USAID and the GAO observed that the security sector can be harder to evaluate because of the sensitivity of the subject matter: "Whereas development people speak freely, security is a little more difficult than other areas because of classification of information, sharing of information, sanitisation of information."⁴³

3. Needs and challenges

The needs and challenges identified in this section reflect the views of the author, and draw upon thinking around best practice in M&E.

Presently, it is impossible to determine with any degree of certainty the percentage of US Government SSR work that gets evaluated. USAID evaluates 100 percent of their work, but they are the 'third player' next to DoD and DoS – neither of which have an organisational culture that has traditionally undertaken rigorous evaluation. Moreover, US engagement in stabilisation missions has tipped the balance among these three actors even further in the direction of military oversight of security assistance.

New SSR programmes, however, are now much more rigorous in including M&E requirements than in the past. The practice of M&E is growing. It may be handled in different ways by different

⁴² Interview, GAO official, 22 August, 2008.

⁴³ Interview, GAO official, 22 August, 2008.

departments and agencies, but it is happening nonetheless. Furthermore, it is being mandated in lots of new ways, as discussed above.

Yet, in order to encourage both routine inclusion and, equally important, execution of M&E in a manner that results in quality data that can be used to improve programme objectives, it must be handled in the right way. Otherwise, performance measurement will be regarded as just another data gathering exercise and not actually connected to an outcome. As one official observed, people generally will be more receptive to M&E if they perceive it to be useful. Willingness to adopt evidence-based programming depends, ideally, on providing this utility – which is not always easy. Indeed, there is no guarantee that by following a set of metrics – even one that has yielded success in the past – success will be repeated or sustained in the future. Ultimately, then, utility is a question of data – how information is gathered and disaggregated. Most agencies have neither sufficient dedicated financial resources nor staff to do this level of analysis. Furthermore, how the data is analysed and interpreted is important – ‘bad metrics can be a pseudo-science; they can give an aura to metrics that are not necessary relevant.’⁴⁴

So, while there is a need to continually refine indicators based on experience, analysis is equally important. This process is inevitably complicated because indicators need to be tailored to different contexts. The experience of those interviewed suggests that many people, both in the field and at headquarters, remain sceptical that metrics can adequately capture positive improvements in people’s lives. In the view of one police advisor with DynCorp, qualitative community surveys provide the best indicators for police, defence, and justice sector performance, though these instruments can be difficult to apply in ‘hotter’ environments. The second DynCorp advisor similarly advocated for qualitative M&E.

Likewise, the Haiti Stabilisation Initiative (HSI) has found year-on-year surveys by USAID a useful gauge of trends in Cite Soliel versus the nation as whole; though cautioned that it was both hard to attribute change to their programmatic impact and to determine subjectivity in the data due to how it was collected. HSI further commented that the process involved in determining how to collect data for MPICE, which they are using as a metrics framework, involved several false starts at the expense of time and resources. It was suggested that an advisory board of nationals familiar with the country-context could have provided guidance on what approaches might prove more effective.

As the HSI and Iraq experiences suggest, getting good quantitative indicators is often difficult in developing countries, so instead departments have had to rely on qualitative, descriptive indicators. These are more *ad hoc* – e.g., “how do you actually judge the performance of US trained battalions?” For example, it was noted that in 2001, Nigerian peacekeeping units deployed to Sierra Leone received substantial accolades for their performance in stabilising the country. Their performance could possibly have been attributed to the US military training effort, but the unit with best performance also had a good battalion commander. Therefore, one needs to be sensitive to the causal nature of the outcome and not attribute too much to the programme.

One source called for greater flexibility to combine both the social sciences approach of USAID and the simpler, ‘numbers on the ground’ approach of DoS and DoD. Noting that alone, either approach can be unsuited to measuring impact in post-conflict environments; it was suggested that each be applied over different timeframes of the project. The numbers approach might be appropriate during the first six months when the goal is to put ‘bodies on the ground’ and when the systems needed to capture the kind of data sets for statistical analysis are absent. However, once police or judges are trained and deployed, after 6-12 months, it may be appropriate to shift towards monitoring more refined indicators, especially as data collection systems for monitoring improve. To do this effectively, one needs to understand where one is in the process of rebuilding and reforming institutions, and which indicators are practical. However, the same source

⁴⁴ *Ibid.*

expressed scepticism that such a combined approach was possible since each agency or department has not yet understood the value of the other's approach – due largely to organisational culture.

Evidence-based programming is not without its sceptics – those who ask, “can you really gather this type of data in a meaningful way?” or “Are the linkages to conflict reduction realistic?” Many people within the USG continue to need convincing, in the view of several officials interviewed from both headquarters and country level.

In order for performance measurement to be not just useful, but recognised as such, metrics need to be built into management structures – agencies and departments need both expertise and resources. Otherwise, those required to implement M&E will see it as burdensome. Departments and agencies therefore require more resources – additional personnel specialised in M&E (including staff knowledgeable about contracting mechanisms, developing performance indicators and developing diligence), training capacity for existing programme staff, and above all budget lines for M&E systems – to be allocated. Ultimately, this requires willingness, particularly as this may divert resources from other programmes or functions. As an official commented, “It's really about *making a case* for evidence based-programming. Partly, this is an educational process – making sure that our superiors – including Congress and OMB – understand what is possible and what [we] are trying to do.”⁴⁵ Indeed, improving M&E systems is unlikely to result in improved programme performance if the demand for accountability and oversight is weak.

Military personnel pose a particular challenge because they are assigned for a limited period of time, and thus may need to focus on a shorter time frame that may not be appropriate to achieving overall high order, long-term goals. More effort needs to be made to delink the individual from success or failure and focus on overall outcome and its timeframe (that is to say, a focus on the big picture).

According to those interviewed in this research, encouraging more routine M&E is also linked to local ownership. It is not only USG departments in which the case needs to be made. Ultimately, the goal should be that the host country absorbs methodology – commission monitoring of activities and use data to make course corrections.

With regard to the usability and accessibility, one source noted that there was no consistent structure for evaluations even by USAID. As a practitioner, if one is trying to determine what works based on existing evaluations in the USAID Development Experience Clearinghouse database, it is hard to identify consistent themes. In the view of the respondent, it would be useful to have standard themes – programme design; programme implementation; progress towards goals, including assessment of monitoring and evaluation systems; sustainability, including availability of resources; and local government buy-in for after USG funding ends. This is not to argue for a standard format, but for greater discipline in content. Since evaluations should also identify failures, regular third-party evaluations and GAO assessments remain necessary. GAO, IG, and non-government reports are much more effective at revealing poor practices because they are intended to do so. Contractors, as well as agencies and departments, are often resistant to openly criticising their own methods or identifying what went wrong, as they perceive this jeopardising their chances of getting their contract, or budget appropriation, renewed.

Even where there is support for evidence-based programming, the availability of resources can determine the extent to which evaluations are used. Programmes simply may not have enough resources to fulfil recommendations and make necessary course corrections. While there are sufficient funding streams for military assistance, there is considerably less for police and for governance/oversight. Moreover, shortfalls and legal restrictions on the use of funds may prevent action even where problems and potential solutions have been identified.

⁴⁵ Interview, US Department of State official, 21 August, 2008.

The Cross-Departmental Nature of SSR

In the absence of an agreed policy framework for integrated SSR programming, where the USG is involved, each implementer is responsible for conducting its own M&E. However, there is growing recognition that, as the outcome may be greater than the sum of its parts, individual departments' programmes may benefit from reaching out to one another, especially when programme results are unsatisfactory. For example, if INL and USAID are working on national police and justice sectors, respectively, in isolation from each other, they may feel that their project is affected by factors outside of their responsibility – due to the inter-linkage and interdependence between police and justice. One respondent noted, "This is where the benefit of an integrated approach lies. Integration can't change who the implementing authority is, for example, getting DoS to work on military issues, but it can make sure that we are all linked up." Integration occurs, ideally, at the embassy/country team level, but it can be raised up to the policy level if beyond the ability (or willingness) of those on the ground to resolve.

The cross-departmental nature of SSR can also have a multiplier effect, particularly with senior decision-makers, when other stakeholders are also producing assessments saying the same thing. At the same time, overcoming institutional 'stove piping', due to organisational structures and cultures, and to funding mechanisms remains a challenge. This may be particularly true in stabilisation environments. If one has never worked in a stabilisation environment, one may not understand the importance of changing the overall environment, or the contribution of a single sector to the greater strategic goal.

Of course, the more actors there are, the more complicated M&E becomes due to the process of agreeing on performance measures, goals, plans, etc. This is true, not just of USG actors, but international partners as well. In Afghanistan, for example, the US is working alongside NATO, the UN, and bilateral partners. Each actor has different standards and means of implementation, especially when assistance is highly internationalised.

The cross-departmental nature of SSR also means that there is a wide range of different practitioners at different levels with varying degrees of understanding. Some may be fully aware of the OECD SSR definitions, others may not even be familiar with term SSR. Several DoS, DoD, and USAID staff interviewed are working together to try to build collaboration so as to ensure greater understanding of the linkages between their respective projects and programmes. There is an expectation that as inter-agency work proceeds, it will gradually become institutionalised; replacing the *ad hoc* process that currently exists in getting advice from other agencies.

4. Trends and opportunities

The trends and opportunities identified in this section reflect the view of the author, and draw upon thinking around best practice in M&E.

There are several trends at work within the USG concerning M&E, including – but not limited to – the past several years experience with stability operations.

- *Agreement on an inter-agency SSR policy:* The draft inter-agency policy on SSR has been signed by the Assistant Secretaries of DoS, DoD, and USAID. If formally adopted, the policy may promote evidence-based SSR / M&E. One official noted that both their department and the USG as a whole were "still in crawling stage" concerning SSR policy and implementation, and that as the field develops, existing M&E criteria are likely to need refining.⁴⁶ The draft policy document identifies as a 'next step' whether other M&E criteria than those currently

⁴⁶ Interview, US Department of State Official, 21 August, 2008.

used by individual agencies and departments are required for a whole-of-government approach to SSR programmes. Additionally, several respondents noted that a formal inter-agency approach would encourage – indeed, require – greater collaboration and co-ordination, which could potentially lead to joint evaluations, raising the level of examination beyond agency-specific programmes to a more macro/strategic level.

- *Stabilisation Operations in Afghanistan and Iraq*: The experience of the DoD in stabilisation missions, particularly in Afghanistan and Iraq, has resulted in a growing recognition that performance measurements are important for quantifying impact. This point has been reinforced through numerous evaluations of security assistance programmes in these two countries by both the Inspectors-General for DoS and DoD,⁴⁷ as well as the GAO. For both Afghanistan and Iraq, there are now monthly meetings between the Inspectors General in USAID, DoS, and DoD, and the GAO to try to not duplicate their audits. The experience in these two countries has also influenced DoS and USAID, resulting in increased co-operation between the three on the MPICE framework.

Much of the impetus at the policy level for enhancing monitoring and evaluation is the 'Guidance for the Employment of the Force', the top-level DoD operational guidance and planning document, which President Bush signed in April 2008. In addition to prioritising defence and SSR, the document requires more robust assessment: 'COCOMs [Combatant Commands] and Services [the US Army, Navy, Air Force, and the Marine Corps] will be required to assess how effective their exercises are in contributing to the achievement of DoD's strategic end states These assessments will affect decisions on which exercises 'make the cut'.⁴⁸

GAO noted that while its Comptroller General's Authority (CGA) is generally being used less today than in the past, when it is invoked as with Iraq, it is more likely due to *too much* interest from too many Congressional committees, rather than too little Congressional interest as in the past. Thus, the CGA was used to draw attention. If GAO is the invoking authority, it is better able to design and undertake an assessment in a manner that balances the needs of multiple interested committees. CGA is thus used to make the audit process more neutral. It was noted that US stabilisation programmes in Afghanistan are likely to fall under CGA soon. Another advantage of the GAO audit process is that it is also cross-cutting – for example, on counter-narcotics, it is examining the involvement of DoS, DoD, USAID, and DEA. A process that is further reinforcing the utility and merit of joint assessments.

A Special Inspector General for Iraq Reconstruction (SIGIR), established as a temporary federal agency in October 2004, provides quarterly reports to Congress, including on the progress of US support to developing the Iraqi security forces.⁴⁹ As a result of SIGIR, the IGs are "staying ahead of curve on Afghanistan."⁵⁰ Co-ordination among the departments is improving, because they know there is going to be a lot of auditing.

Nonetheless, the USG, especially the DoS, remains dependent on outside contractors to implement its SSR projects – contractors that are under pressure as a result of the above

⁴⁷ There are now, for instance, joint reports by the IGs of DOS/DOD on police in Afghanistan.

⁴⁸ Tim Hoffman, Guidance for the Employment of the Force (GEF), ODASD Partnership Strategy Worldwide Joint Training and Scheduling Conference, 17-21 September, 2007 via, www.dtic.mil/doctrine/training/wjts07_2feg_brief.ppt, accessed on 19 August, 2008.

⁴⁹ In response to the 'late focus on metrics' between contractors and the U.S. Government, SIGIR's *Iraq Reconstruction: Lessons in Programme and Project Management* (March 2007), Lesson 9 states, 'An effective monitoring and oversight plan needs to be in place within each agency from the outset of contingency operations. This will allow for early and direct feedback to programme managers, who can implement course corrections in operating practices and policies....Operations that involve multiple agencies, funding streams, and management systems require that the Congress take steps to standardise oversight and provide clear guidance on any reporting requirements involving multiple agencies.', p 19, http://www.sigir.mil/reports/pdf/Lessons_Learned_March21.pdf.

⁵⁰ Interview, US Department of Defense official, 09 September, 2008.

mentioned reports to improve their performance, but whose cultures are adverse to transparency, particularly if it exposes negligence or wrong-doing. Similarly, the subordination of civilian agencies to the military in stability operations represents a new paradigm that has complicated project oversight, subjecting it to competing interests and different standards.

A separate result of US experience in stability operations that could, potentially, have a bearing on improved monitoring and evaluation of security sector programming is the nascent Civilian Reserve Corps being set up under S/CRS. The Civilian Reserve Corps was first proposed by the 2006 National Security Strategy, and has since been referenced by the incoming US administration, to address the critical need for civilian expertise – including police and rule of law advisors – in post-conflict environments. The Corps is still in the process of development, but, in theory, it could be designed to include both M&E experts and a mandatory training on M&E requirements.

- *Measuring Progress in Conflict Environments (MPICE)*: Driven largely by the challenges the USG has faced in building new security institutions in Afghanistan and Iraq, the US Institute of Peace – in collaboration with a range of stakeholders including the Army Corps of Engineers, and the US Army Peacekeeping and Stability Operations Institute – has developed the ‘Metrics Framework for Assessing Conflict Transformation and Stabilisation’. The framework is part of the MPICE initiative, which was established in 2006 with funding from the Office of the Secretary of Defence, USAID, and the DoS Office of the Co-ordinator for Stabilisation and Reconstruction.

According to one of its primary architects, MPICE was created to fill a critical gap in measuring the effectiveness of performance in stability operations. In addition, it was to develop a systematic approach to identifying performance indicators, tracking progress, and making course-corrections. As noted above, this need has been identified by Combatant Commands, as well as by reports of the OIGs and GAO. By providing a tool for both baseline assessment of challenges and measuring progress, the intention of the MPICE framework ‘is to enable policymakers to establish realistic goals, bring adequate resources and authorities to bear, and focus their efforts strategically, and enhance prospects for attaining enduring peace’.⁵¹ It focuses on drivers of violent conflict and the capacity of national/local institutions to address them, and on outcomes rather than outputs in five core areas: political moderation and stable democracy, safe and secure environment, rule of law, sustainable economy, and social well-being, which encompass many goals of SSR.⁵²

MPICE has been designed specifically for situations of direct engagement by the US in stability operations. It has been field tested in Afghanistan, is being piloted in Haiti by the HSI, and rolled out in Georgia. It could, in theory, also be applied to steady-state contexts. However, as noted above, getting agencies and departments to think beyond their immediate sectoral priorities towards an overall strategic goal is even more difficult in non-stabilisation environments.

As with M&E systems generally, the willingness of senior managers to adopt MPICE – and its utility for influencing senior political actors – depends on how easily it can be understood. According to HSI, the framework is an important tool for determining whether there is progress towards the broad goal of stabilising Haiti, and across which sectors.

⁵¹ Metrics Framework for Assessing Conflict Transformation and Stabilisation, February 2008.

⁵² Examples of metrics include, in the area of security, ‘number of assassinations or attempted assassinations of opposition groups members attributed to state security forces’, and ‘Extent to which personnel rosters are inflated with phantom soldiers’. In the area of justice, ‘percentage of known prison population detained beyond the period specified in the law who have not had their case reviewed by an appropriate authority’, ‘average time from entry into system on serious crimes charges until seeing a lawyer’, and ‘boundaries between formal and informal dispute resolution mechanisms are clear and uncontested’.

These apparent benefits do however need to be tempered; while MPICE can show trends, it cannot prove direct causality and therefore it cannot be used as a method of leading decision-making. As a consequence, it does not, for example, provide a reliable guide as to whether to increase or cease funding for a particular activity. MPICE also needs to be easier to initiate than in Haiti. However, the desire for something 'off-the-shelf' needs to be reconciled with the need to account for differences in environments.

Beyond the USG, the UN Peacebuilding Support Office and UN Department of Peacekeeping's Best Practices section have expressed interest in the applicability of MPICE to UN peacebuilding and peacekeeping missions. At the field level in Haiti, the World Bank and several bilateral donors have also expressed interested in developing frameworks similar to MPICE. While multilateral and bilateral actor interest in developing MPICE-like frameworks is an indication of growing recognition of the utility of measuring impact, it also raises questions about whether data sets will be interoperable and, therefore, whether results will be reliable guides to progress.

Annex 1: Interviewees

Anonymous, DynCorp Contractor, (Middle East)

Anonymous, DynCorp Contractor, (Iraq)

David Becker, Director, Haiti Stabilisation Initiative, USAID (Port-au-Prince)

Mike Bittrick, Deputy Director, Regional and Security Affairs, Bureau of African Affairs, US Department of State

Miriam Carroll, Senior Analyst, International Affairs and Trade Team, Government Accountability Office

Mike Dziedzic, Senior Programme Officer, Centre for Post-Conflict Peace and Stability Operations, United State Institute of Peace

Joey Dasgupta, Senior Analyst, International Affairs and Trade Team, Government Accountability Office

Col. Gregory Hermsmeyer, Dep. Director, Int'l Capacity Building, OSD/Global Security, Affairs/Partnership Strategy, US Department of Defense

Kalkus Hynek, Assistant Director, International Affairs and Trade Team, Government Accountability Office

Robert Perito, Senior Programme Officer

Centre for Post-Conflict Peace and Stability Operations, United State Institute of Peace

Col. Dan Pike (ret.), Director, African Affairs, OSD/International Security Affairs, US Department of Defense

Elizabeth Repko, Senior Analyst, International Affairs and Trade Team, Government Accountability Office

Christina Rosati, Senior Co-ordinator, Office of Plans, Policy and Analysis, Bureau of Political-Military Affairs, US Department of State

Joseph Trincellito, International Consultant

Julie Werbel, Senior Security Sector Reform Advisor, Office of Democracy and Governance, US Agency for International Development