Measuring Stabilisation: Some observations from international experience

Andrew Rathmell

As with most areas of public policy, the desire to measure progress and assess impact has been a central feature of stabilisation efforts in recent years. At times, this has reached absurd proportions. For instance, at the height of the NATO effort in Afghanistan, a common problem was survey fatigue, with Afghan villagers being asked to respond to multiple waves of surveys by different organisations, often with no tangible projects subsequently being delivered. This desire is not of course new. Indeed when designing evaluative systems for the US effort in Iraq in 2003, I benefited greatly from consulting RAND colleagues who had helped design evaluation systems for US forces in Vietnam in the 1960s. So, while the techniques and environment may have evolved, the policy imperative to know what is happening and why remains consistent.

Nonetheless, for governments seeking to plan and evaluate complex stabilisation operations, there are a number of useful pointers from recent international experiences. In this note, I address four aspects: why measure; how to use measurement to inform decisions; how to measure; and some overall lessons as well as new directions in this discipline.

Why measure?

This may seem an obvious question but being clear about the goals of the exercise is important to ensure the measurement and evaluation process is fit for purpose. Lots of stakeholders will have an interest in measuring. Government officials and security force officers will want to know if they are delivering the plan; politicians will want to demonstrate success; opposition politicians and journalists will want to identify failures; international partners may tie future assistance to good performance; and stakeholders from citizens, businesspeople, and criminal groups will want to know whether to support or oppose the stabilisation efforts.

From the perspective of central government, I would posit three central purposes for the measurement function. The first, and most obvious, is to enable policy-makers to assess how *successful* their plans are being, so enabling actions to be adjusted to reinforce success or to address failures. This is just good project management. The other two purposes, though, are sometimes neglected and are equally vital. The second is communicating and engaging. *Engaging stakeholders* (across government as well as outside government) is critical to the successful delivery of any stabilisation operation. The overall planning process and the ongoing process of measurement and review are very valuable tools for ensuring such engagement. The third purpose, however, is arguably the most important as it determines the success of the overall effort. This is about using measurement to force planners and decision-makers to *think through the problem*. Teasing out a sound theory of change and understanding the causal logic of proposed interventions are steps that are often not done very rigorously. This can lead to misdirected, or even counter-productive, efforts and the privileging of activity over impact.

Using measurement to inform decisions

In order to illustrate a variety of ways in which a rigorous measurement and evaluation approach can make a difference to stabilisation efforts, a few stories from the recent past.

Iraq: The story of the US "surge" in Iraq in 2006/7, and the role this played in tamping down the Iraqi civil war, has been retold many times. The commitment of Iraqi Sunni tribes and coalition military and civil personnel in the field were critical. However, fundamental to the design and hence success of the strategy was a clear-eyed theory of change. Compared to previous coalition plans, which aimed at the long-term goal of building a "democratic, stable Iraq", the campaign plan directing the surge focused on achieving an "acceptable" level of violence in part through a balance of power between competing Iraqi factions. While less lofty a goal, this reflected a deeper understanding of the context, a realistic level of ambition, and hence drove practical plans across the US mission to achieve this objective, tied to clear metrics as signposts along the way.

Afghanistan: At the national and provincial levels, with the injection of additional military and civilian resources into the coalition starting in late 2009, there was considerable investment in a variety of measurement and evaluation mechanisms. While there were numerous challenges in making these systems integrated and effective, they did allow evidence-based decisions to be made that had tangible impacts on the campaign.

At the provincial level, notably in Helmand and Uruzgan, but also more widely, systems were put in place to systematically collect and analyse data on projects, perceptions, and stabilisation outcomes. Over time, this allowed decision-makers to understand which investments were most successful in achieving mission objectives. For instance, in Helmand, this process allowed the civil-military mission leaders to determine the relative effectiveness of costly programmes such as road-building, school-building, healthcare, or justice provision, in relation to the objective of increasing legitimacy of the provincial government. At the national level, detailed, independent evaluations of public perceptions of the Afghan National Security Forces (ANSF) enabled ISAF to understand that throwing resources at building the numerical capacity of the ANSF was at times actually counterproductive. In plenty of instances, ANSF were exacerbating the conflict through their corrupt and predatory behaviour. This insight enabled ISAF to shift its focus, with a much greater investment in anti-corruption activities.

Philippines: Stabilisation operations in Mindanao have been led by the Philippines authorities but with external, largely US, development and security support. Evaluations of USAID work there in the 2000s illustrated a common challenge when targeting development assistance in a stabilisation environment. Although the ostensible rationale for much of the development assistance was to support stabilisation, and projects were often informed by detailed conflict analyses, the bulk of programming was focused on long-term socio-economic development issues rather than on shorter-term interventions targeting drivers of instability or capacities for peace. As a result, it was hard to find any correlations between the development spend and patterns of violence. This is symptomatic of a wider, ongoing, tussle between development and stabilisation practitioners – how far should and can long-term development programming be adjusted to address shorter-term conflict drivers. While long-term development should not be ignored, without measurable impact on short-term conflicts, the utility of long-term socio-economic development in a stabilisation environment is questionable at best.

How to measure

There is a vast and growing industry devoted to the mechanics of how to design, operate, and use measurement and evaluation to improve public policy. Though military, civilian, and development agencies often have differing jargons and focus, the techniques and practices are similar and can be integrated for stabilisation operations. There are six crucial steps when setting out to design a stabilisation measurement process.

Define stability: While there are templated models of what causes instability and of what contributes to stability, every context is different. It is crucial therefore to use deep local knowledge and an explicit normative framework to define the stabilization objective that is being sought. Generic references to goals such as "good governance and sustainable development" will not suffice. The more precise the definition of the goal sought, and the more achievable this goal, the better.

Develop a bespoke Theory of Change: The hardest part of any planning and measuring process is to determine how proposed interventions are expected to lead to the desired outcome. How exactly will strengthening local security forces impact on conflict dynamics and public perceptions of the state? Will land reform contribute to or detract from stability both in the short and long-term? Often, at the outset, theories of change are not based on detailed knowledge of how causal linkages work, meaning that they need to be tested over time and reformulated as the situation becomes clearer.

Collect data: Data collection often consumes most time and effort which is why it is important to set the goals and theories of change clearly and narrowly so that policy-makers do not drown in unnecessary data. "Traditional" data collection techniques such as face to face surveys and focus groups, collecting administrative data (e.g. on numbers of trained police officers, schools built and staffed), and using intelligence and social research methods remain at the core of the process. Increasingly, tools such as social media analysis and big data analysis can add considerable value and depth to the data that is collected.

Analyse causal relationships: Extracting meaning from the data will generally involve a range of techniques, involving at least qualitative, quantitative, and geospatial analyses. It is always vital, and sometimes politically difficult, to ensure that the analysis is encouraged to rigorously test assumptions, to test theories of change, and to explore alternative hypotheses. The key with the analytical process is to explore how stabilisation interventions have affected conflict dynamics and whether other interventions may be more effective.

Ensure participation: Any stabilisation operation will involve multiple actors across national and local government, and probably include international agencies. The evaluation process can provide a very useful platform with which to bring these actors together to forge a shared understanding of the problem and to ensure coherence amongst the agencies responsible for delivering parts of the stabilisation intervention. Beyond the government, evaluation processes can also be useful in building wider support for the stabilisation process. For instance, in support of a widely agreed peace process, allowing citizens and NGOs to report on how they see progress, can be very valuable in providing local feedback and in maintaining confidence in the stabilisation and peacebuilding process.

Drive decisions: In many walks of life, evaluation reports can result merely in a study that sits on a shelf. To be of use, the measurement and evaluation process needs to be tied to a decision-making process. The evaluation report and process needs to be owned by a body capable of making decisions, and there needs to be some incentive to use the information with which to make decisions. This may be tied to public or parliamentary scrutiny, to budget cycles, or some form of independent evaluation oversight structure.

Lessons and new directions

To conclude, I'll propose three overall lessons and two new directions that practitioners should keep in mind.

Three lessons

Once it gets into the hands of evaluation and analysis professionals, the measurement and evaluation process can become mind-numbingly complex; with intricate diagrams and spreadsheets full of data. At a high level its important to keep in mind three core principles to make the process of use to decision-makers:

Keep it simple but intelligent: However complex and detailed the underlying analyses, focus on ensuring that the big questions are addressed and answered in simple ways. Is it working? Why is it working? What unexpected impacts are we having? What do we do differently? Build a process in which the practitioners and decision-makers can have intelligent, evidence-based debates over these issues so that the overall result is a more intelligent strategy.

Ensure accountability: The persistent challenge is to get the balance right between encouraging honest reporting and debate, which necessitates avoiding a blame culture, and ensuring that wrong decisions or implementation failures are rapidly corrected. Those who decide on and execute stabilisation strategies need to use measurement to be held accountable but if measurement processes are used to punish then honest reporting will not happen. The "no-fault" incident reporting system used in the airline industry is one approach that has gained popularity among organisations seeking to get the balance right.

Consciously manage trade-offs: Many stabilisation plans will try to paper over trade-offs or deny internal conflicts within the plan. For instance, in Afghanistan, did coalition efforts to target opium producers weaken insurgents by hitting their finances or did they strengthen insurgents by arousing popular anger among farmers? A common bureaucratic tendency is for any plan to pursue multiple objectives but a rigorous and honest evaluation process can help tease out where there may be trade-offs to be made and so encourage an honest debate among policy-makers about the relative importance of these priorities.

Two new directions

There are significant developments underway in thinking and practice in two fields that stabilisation planners and evaluators need to understand and make use of: complexity theory and behavioural science. Neither are new but in both cases there is a growing body of experience in applying these that can be of use.

Managing Complexity: It is no longer novel to say that large scale social change, of which stabilisation is an example, needs to be understood as a complex adaptive system that cannot be managed by "simple" linear project planning processes. There have been numerous efforts in recent years to work out practically what this means – from the US military's "planning by design", through USAID and DFID's "adaptive management" approaches. However, we are at the early stages of working out how to take full advantage of such insights and to build these into our stabilisation planning and evaluation systems. At a minimum, it is important to use techniques of adaptive management, putting resources into Learning as well as into Monitoring and Evaluation. More interestingly, stabilisation planners can look to work in the profession of economics, for example being done at the Santa Fe Institute and Oxford Martin School, which is seeking to explore how complexity thinking can shift paradigms for the design and measurement of large scale social interventions.

Using behavioural science: Stabilisation is, of course, all about behaviour change – persuading populations to support government rather than insurgents; persuading people not to engage in violence, etc. Most stabilisation planners therefore focus on analysing the target audiences and, in their theories of change, working out how to alter behaviours in the desired way. Hence, advances in the science and techniques of behaviour change need to be closely watched by stabilisation planners and evaluators. These advances range from the now well-established "nudge" theories, through the sort of fine-grained social network analysis increasingly used by both military and private sector actors, through the sort of micro-targeting that is being improved month by month by political and consumer marketeers, to deeper work on topics such as integrative complexity that is enabling a better understanding of individual proclivities to violence. Since the ultimate aim of stabilisation programmes is to shift the actions of selected populations to behave in the desired manner, advances in this area need to be exploited to allow for the creation of more effective stabilisation interventions and the measurement of change in the target populations.