

NOREF Policy Brief

Measuring peacebuilding: challenges, tools, actions

Svein Erik Stave

Executive summary

How can the effectiveness of peacebuilding operations in countries marked by conflict be better measured? This policy brief examines the steps needed to improve the measurement of peacebuilding work, highlights the technical and political problems this work faces, and makes recommendations for action by organisations in the field.

The experience of peacebuilding initiatives around the world has in recent years led to increased efforts to develop new and improved tools to measure their effects. Many projects are already underway, led by key civilian and military actors such as the United Nations itself to defence agencies, government departments, the World Bank, and NGOs.

These various efforts reflect both the mismatch between ambitions and results in Afghanistan and Iraq and longer-term concerns about the limitations of data, methodology and practices in the area of measuring peacebuilding. There is a stark contrast here with development goals, where monitoring procedures are well established and far more data are available.

The stakes are high, in that the United Nations alone currently spends more than \$7 billion every year on international peacekeeping and peacebuilding activities, and donors are increasingly pushing for improved documentation of the effects of this work. It is therefore timely to consider the needs and challenges of measuring peacebuilding, and what practical steps the relevant organisations need to take to develop their capacities at every level.

The challenges are many and varied. They include improving the quality of the strategic information that will be produced and used in decision-making, which in turn means developing the methodologies and procedures for collecting and processing data; recognising the technical limits presented by the difficulty of correlating particular peacebuilding activities to system-wide effects; addressing the tendency to focus more on achievements by acknowledging failures and risks, and achieving a better balance in reporting; being aware of the influence of assumptions and ideas that might influence conclusions in ways the evidence may not support; and, crucially, factoring in detailed attention to local contexts, thus avoiding dependency on "universal" standards or theories in reaching conclusions, which can make the measurement less connected to ground-level realities.

To meet these and other challenges related to measuring the effects of peace operations, there is a need to build capacity in the field within organisations involved in peacekeeping and peacebuilding, including the UN; promote organisational cultures and systems that take diversity, uncertainty and risk into account; and develop data-collection systems and databases that accord with peacebuilding perspectives and needs. A key activity ahead should be to integrate the new perspectives, methodological approaches, and guidelines in the field of peacebuilding monitoring and evaluation.

Svein Erik Stave is a researcher with the Fafo Institute for Applied International Studies in Oslo. He is the author of *Monitoring Peace Consolidation: United Nations Practitioners' Guide to Benchmarking* (United Nations 2010), and is involved in a number of ongoing international initiatives to develop methodologies and indicators related to measuring the effects of peacebuilding. Svein Erik Stave studied geography at the University of Oslo, and worked as a statistical advisor at Statistics Norway. His main research interests include the crafting of models and systems to measure complex and ambiguous concepts such as sustainable development, vulnerability, and peacebuilding. Among his current projects are establishing sustainable development indicators for China; measuring the costs of displacement for the World Bank; creating a disaster-risk profile for Ethiopia; and working on a project on how to measure Norway's contributions to international sustainable development.

Introduction

There has recently been an increase in efforts to develop new and improved tools to measure the effects of peacebuilding initiatives and of peacekeeping. These terms are normally used to describe separate objectives and different phases in the transition from conflict to sustainable peace, though in this brief peacekeeping is included in the category of peacebuilding.

The trigger of these efforts is both particular and shorter-term (the gap between high ambitions and disappointing results in cases such as Afghanistan and Iraq) and general and longer-term (the limited success of international peacebuilding efforts over the last fifty years).

The stakes are high: the United Nations alone currently spends more than \$7 billion every year on international peacekeeping and peacebuilding activities, and donors are increasingly pushing for improved documentation of the effects of this work. This pressure may also be related to a growing public critique in many donor countries of whether and how well such large investments in peacebuilding (as well as in development aid) work in practice.

In order to comply with the expanding requests and needs to measure the effects of peacebuilding, a number of key civilian and military actors are in the process of developing new tools and guidelines:

- The United Nations itself has published its first handbook on so-called Peace Consolidation Benchmarking, whose principles the organisation's peacebuilding practitioners¹ will be expected to apply
- The United States and French military have developed monitoring and evaluation tools under Nato's Multi National Experiment (MNE) programme²

- The UK's department for international development (DfID)³, the World Bank, the US Agency for International Development (USAID)⁴, and Safer World have outlined guidelines on conflict assessment and planning
- The Organisation for Economic Cooperation and Development (OECD's) Development Assistance Committee (DAC) has proposed a Draft Working Guidance on Conflict Prevention and Peacebuilding⁵.

In addition, various monitoring and evaluation manuals have been prepared by police forces, defence agencies and emergency actors; and research centres and consultancies have constructed their own tools. The latter include an ongoing initiative at the Crisis States Research Centre of the London School of Economics (LSE) on how to improve the "index methodologies" used to rank countries according to their level of peace or fragility.

Relating political and technical aspects

These multiple projects notwithstanding, the tasks associated with measuring the effects of peacebuilding – whether assessing a country's overall development in this area, or the results of a particular peacebuilding initiative or actor – remain significant. An important aspect of such tasks is methodological or technical: how to correlate observed system-wide changes to individual peacebuilding projects, and how to attribute precise objective measures to often ambiguous concepts such as peace and fragility?

This challenge has become more testing in the context of the paradigm of integrated peacebuilding – which includes, apart from security-related issues,

1 United Nations (2010), Monitoring Peace Consolidation: United Nations Practitioners' Guide to Benchmarking, New York, United Nations, http://www.un.org/peace/peacebuilding/pdf/Monitoring_Peace_Consolidation.pdf, accessed 9 May 2011.

2 USJFC (2010), Handbook for Joint Force Commanders: Assessing progress in environments involving irregular adversaries, http://mne.oslo.mil.no:8080/Multinatio/MNE6produkt/31FRJFCsAs/file/3.1%20FR%20JFCs%20Assessment%20of%20Progress%20Handbook_final.pdf, accessed 9 May 2011.

3 DFID (2002), Conducting conflict assessment: guidance notes, http://www.swisspeace.ch/typo3/fileadmin/user_upload/Media/Topics/Peacebuilding_Analysis_Impact/Resources/Goodhand_Jonathan_Conducting_Conflict_Assessments.pdf, accessed 19 May 2011.

4 USAID (2004), Office of conflict management and mitigation, "Conducting a conflict assessment: a framework for strategy and program development", http://www.usaid.gov/our_work/cross-cutting_programs/conflict/publications/docs/CMM_ConflAssessFrmwrk_8-17-04.pdf, accessed 19 May 2011.

5 OECD (2008), Guidance on Evaluating Conflict Prevention and Peacebuilding Activities, http://www.oecd.org/secure/pdfDocument/0,2834,en_21571361_34047972_39774574_1_1_1_1,00.pdf, accessed 9 May 2011.

socio-economic and political dimensions, as well as more general development-related ones. What can make it even more difficult is that political factors may influence what is being measured and how the results are reported, as well as how the latter are subsequently used in decision-making.

Political narratives and priorities will always influence what is to be measured in peacebuilding.

This policy brief outlines some of the main political and technical questions raised by the measurement of peacebuilding. These questions mainly derive from reviews of past and

current monitoring practices within the UN,⁶ and of existing monitoring and evaluation tools developed and used by other actors, including indices of peace and fragility.⁷ They can broadly be divided into political and technical types, though in practice these are often interrelated; for example, techniques of aggregation can be related to the ability to conduct overall political assessments of effects and strategies.

The problem of narrative

Perhaps the most fundamental flaw in existing monitoring and evaluation tools is their inability to recommend adjustments to the overall strategic idea or narrative of a peacebuilding initiative; for example, the balance between military and civilian efforts, or whether creating western-style institutions or legal systems is an effective instrument of peacebuilding.

This in turn reflects something important in the process of measuring the effects of peacebuilding, which is that political narratives and priorities will always influence what is to be measured. Any peacebuilding initiative has at its heart some narratives or pre-assumptions about what characterises a peaceful society and what premises must be fulfilled if peace is to be sustained. These

pre-assumptions can take various forms: “theories of change”, “pillars of peace”, or merely the worldviews or priorities of donors and peacebuilding actors. From there, they are usually transferred directly into measurement frameworks used to specify the defined components of peace. The pattern can be seen in many monitoring and evaluation initiatives that are based on the Logical Framework Approach (LFA).

A problem in this approach – namely, selecting indicators based on particular peacebuilding theories, narratives, perspectives or pre-assumptions – is that its use of measurements leads to consolidating (rather than challenging) the pre-assumptions. The clear danger is a kind of circularity which produces only the semblance rather than the reality of knowledge, and where measurements are used to promote the construction of “virtual realities” – namely, progress against a non-existing situation rather than what is happening on the ground.

The problem of basing measurements of peacebuilding on theoretical pre-assumptions is not that the latter exist (after all, any measurement tool must be based on some underlying theoretical concept or vision), but that the pre-assumptions go largely uncontested by the measurements. The results of monitoring and evaluation exercises thus tend to lead only to minor modifications of programmes and approaches within the limits of their original narratives, and fail to provide information which might usefully question the underlying “theory of change”.

This phenomenon is seen in the relatively small adjustments made in the mandates of peacekeeping operations from the UN Security Council over recent decades. The vast majority of benchmarks and indicators used to measure the effects of these UN peace operations derive directly from their mandates, normally in combination with the UN’s five general pillars of peacebuilding: 1) Peace and security, public safety; 2) Political processes, including electoral processes, reconciliation and conflict resolution; 3) Human rights, rule of law, transitional justice; 4) Economic revitalisation; and 5) Provision of basic services.⁸ The very lack of

6 United Nations (2010), *Monitoring Peace Consolidation: UN Practitioners’ Guide to Benchmarking*, Appendix B.

7 Javier Fabra Mata and Sebastian Ziaja (2009), *Users’ Guide on Measuring Fragility*, Bonn /Oslo, German Development Institute/Deutsches Institut für Entwicklungspolitik and UNDP Oslo Governance Centre, editors Jörg Faust and Joachim Nahem, [http://www.die-gdi.de/CMS-Homepage/openwebcms3.nsf/%28ynDK_contentByKey%29/ANES-7W89TW/\\$FILE/UNDP-DIE%202009%20Users%20Guide%20on%20Measuring%20Fragility.pdf](http://www.die-gdi.de/CMS-Homepage/openwebcms3.nsf/%28ynDK_contentByKey%29/ANES-7W89TW/$FILE/UNDP-DIE%202009%20Users%20Guide%20on%20Measuring%20Fragility.pdf), accessed 9 May 2011.

8 Report of the Secretary-General on peacebuilding in the immediate aftermath of conflict, 11 June 2009, A/63/881-S/2009/304, <http://reliefweb.int/node/315500>,

substantial changes here reflects the power of pre-assumptions to limit the work of measurement.

The contextual v the universal

When universal theories of change are used as a basis to measure the effects of peacebuilding, the consequence is that the selection of indicators employed in the measurement also tends to the universal – and more and more removed from any particular context.

It is evident that the vast majority of existing monitoring tools (including indices) provide universal sets of indicators to be measured generally independent of a given context. For the purpose of comparison between countries this has its advantages; but if the aim is to provide strategic information about peacebuilding in an individual country, such an approach is likely to be oversimplistic, incomprehensible, or misleading.

In general, universal indicators such as (for example) the Global Peace Index,⁹ the Bertelsmann Transformation Index,¹⁰ and the Failed States Index¹¹ may provide useful guidance in identifying sound contextual indicators. These universal indicators can be used to generate useful contextual indicators in two ways: a) selecting only the indicators that are most relevant to the context, and b) converting universal indicators into a local expression.

In practice, for example, if a universal indicator is “social interaction”, a contextually adapted indicator might then be the social interaction of members from the two main conflicting groups; or if a universal indicator is “corruption level”, a contextually adapted indicator might be government corruption related to one particular industry or destabilising economic activity (such as the diamond trade or opium production).

An ideal objective here would be to develop indicators unique to the local context, rather than just adapted to particular contexts from a universal set of indicators (or from a “theory of change”). The

decisions this would entail might include whether to use human-rights goals and western ideas of democracy (such as gender equity or legal systems) as instruments of peacebuilding when the local culture in a conflict area is based on other values and institutions.

The value of unique contextual indicators is that they are generally based on in-depth knowledge of a local conflict and culture, together with a creative understanding of the contextual signals that reflect the condition and development of peace in a society. The contextual signals could be the growth or decline in a particular economic activity which highlights people’s focus on longer-term investments, and/or which are associated with a peaceful society in the past; or it could be particular features of interaction among young people from different social groups in (for example) education institutions or public spaces. The knowledge of context can make a major difference.

Contextual indicators are generally based on in-depth knowledge of a local conflict and culture.

The indicators of unclarity

The selection of contextual indicators can be made in a quite technical way: by using local sources of information in combination with sound methodological guidelines. But in practice, the selection of indicators is very much a political process. This is particularly visible when dealing with wide and relatively ambiguous concepts like integrated peacebuilding, sustainable development, and human development, each of which involves different actors with different perspectives, priorities, and goals.

There are numerous examples of indicator sets that are the result of a political buy-in process rather than of methodological considerations. The problem of such a process is not that any of the selected indicators, or the perspectives on which they are based, are necessarily invalid; but rather that their validity and their relationships to each other are unclear with respect to peacebuilding. This makes aggregation and analysis difficult, and produces results of limited value. In addition they tend to reflect universal goals rather than contextual reality.

accessed 10 May 2011.

9 Global Peace Index, <http://www.visionofhumanity.org/gpi-data/#/2010/scor>, accessed 10 May 2011.

10 Bertelsmann Transformation Index, <http://www.bertelsmann-transformation-index.de/en/>, accessed 10 May 2011.

11 Failed States Index, <http://www.fundforpeace.org/global/?q=fsi>, accessed 10 May 2011.

The limits of benchmarking

An additional problem arises here with particular reference to benchmarking¹² in the way it is applied within the United Nations. The formal ambitions of many peacebuilding operations have often been based on political wishes rather than contextual reality – a tendency evident in several country experiences, including those of Burundi and Afghanistan, where significant gaps emerged between anticipated and actual gains in various measures of stability. This reflects in part the reality that peacebuilding is rarely a linear and predictable process; but in part too the problem lies in the fact that benchmarks seeking to address a range of issues (the stabilisation of the security situation, socio-economic realities, and the root causes of conflict) are often overly ambitious and tend to underestimate the time and effort needed to achieve durable peace.

There is a further twist here, evident in Burundi. The tendency to base benchmarks and indicators on ideal rather than most-likely scenarios discourages individuals and organisations from tracking these indicators and benchmarks – because the information that comes through will inevitably be less-than-ideal, and those responsible for processing it will see it as confirming their failure.¹³

The tendency to bias

The experience of existing and past practice in measuring peacebuilding offers a further warning. There is a clear tendency to select indicators that suggest progress towards the achievement of objectives stated in mandates and visions, but which neglect showing setbacks or failures. This tendency might not be intentional but instead the natural result of basing the selection of indicators on particular theories of change inherent in the mandates and goals of peacebuilding initiatives. A typical example would be the use of indicators that measure an expansion of national security forces and their ability to deal with a country's security

threats, but without at the same time measuring a potential increase in security threats.

There is a related tendency to select easily measurable indicators, generally representing aspects that can be controlled by particular peacebuilding activities (such as the number of police staff trained, of ex-combatants entering disarmament, demobilisation and reintegration [DDR], and of schools constructed). Again, there are both technical and political arguments for measuring such indicators; technically it could be argued that it is not possible to measure the effects of a specific peacebuilding activity on the overall peace situation in a country, as there are too many uncontrollable factors involved; politically, there is a will to document and justify positive results from activities in order to obtain support and funding for the activities and the visions on which they are based. A bias towards showing positive results according to the given mandates or goals can characterise both organisations as wholes and units and individuals within organisations. In the UN, a lack of formalised

There is a tendency to select indicators that suggest progress in achieving objectives, but which fail to show setbacks or failures.

reporting formats can contribute to a focus on progress rather than on failure and risk. The reviews of reporting practice within the organisation suggest that information is often polished according to stated visions along the reporting chain – from the field to the highest levels of the secretary-general and the Security Council. This practice also reflects the uncertainties and conflicting information gathered during monitoring exercises, which makes it possible (and often justifiable) to interpret information according to the reader's perspectives and wishes. Thus it represents another example of the relationship between technical and political challenges in measuring peacebuilding.

¹² Benchmarking refers to the measurement of progress against a defined set target or benchmark, such as a situation presumed to represent a peaceful society. It contrasts with monitoring, which measures general progress via a baseline, where the reference-point is the situation prior to an initiative.

¹³ United Nations (2010), *Monitoring Peace Consolidation: UN Practitioners' Guide to Benchmarking*, Appendix C.

The activities-effects conundrum

A key technical problem in measuring the effects of peacebuilding is how to correlate particular peacebuilding activities to system-wide effects. It has proved to be near-impossible (despite some approaches which apply this perspective) to establish such correlations in open systems with high levels of precision, especially as they need to incorporate all the activities of relevant actors as well as factors out of the actors' control.

If it proved possible to link activities and effects with a high degree of certainty, the result would be to provide invaluable knowledge and control in peacebuilding. In the absence of such a capacity, the use of concepts and tools that do posit strong correlations (such as Logical Framework diagrams) can contribute to concealing uncertainties and diverting attention from alternative approaches to peacebuilding monitoring and evaluation.

Some other tools have a different starting-point in attempting to correlate peacebuilding activities and effects, namely an effort to define the contributions of different activities or actors to observed system-wide effects. That is, instead of treating system-wide effects as the sum of all activities, these tools seek to identify to what extent and in what ways the various activities or actors impact on the observed situation of a country or region.

The *UN Practitioners' Guide to Benchmarking* goes even further in this direction, by recommending a focus solely on monitoring the condition of a country and not necessarily trying to define the effects of UN operations. The rationale here is that monitoring should primarily provide on-the-ground information and knowledge of trends that can inform the UN's strategic planning and peacebuilding activities – and not, or not primarily, document the UN's involvement (although the latter can be done within the same framework).

The approach where monitoring is sought by trying to correlate activities and effects, which incorporates biases on achievements and measures of what is measurable, incurs the risk of a “tragedy of the commons” – where all actors take credit for achievements, while none takes responsibility for a negative development of the overall situation in a country. The result is a dissonance where an

organisation's annual reports of a situation and assessment of its own role bears little relation to independent evaluations of the country's predicament.

The dilemmas of data

There are three further technical dilemmas in measuring peacebuilding. The first, and fundamental, one is how to collect, make available and use data. The collection of data in conflict areas is particularly difficult, and this can lead to tradeoffs with respect to data quality (tradeoffs which normally would not be accepted in non-conflict areas). This situation leads to the widespread use of non-statistical data-collection methods: for example, the use of expert groups and media monitoring, or of data collected by other institutions (statistical offices, NGOs, or research bodies).

This can have a dual negative result: the paucity of statistics-based and/or highly reliable data collected primarily to monitor peacebuilding in conflict areas; and the fact that the data taken from alternative sources (such as reports and informal interviews) is of uncertain quality, produced by doubtful collection procedures, and of largely undocumented (or not well documented) character.

The collection of data in conflict areas is particularly difficult, and can lead to tradeoffs with respect to data quality.

The use of data with relatively low reliability is a tradeoff that often has to be made in conflict environments. It is compounded by another frequent tradeoff, and the second technical dilemma: how to avoid using data that does not necessarily correspond in an optimal way to a theoretically preferred indicator, because reliable data on the preferred indicator does not exist. This situation leads often to the selection of indicators that are not optimally relevant for measuring peacebuilding (in part because many monitoring tools or indicator sets look very similar, whether they are designed to measure peacebuilding or other types of development). In this respect, the development community is far ahead of the peacebuilding community in gathering data for monitoring purposes, resulting in a profusion of development-related data as against a scarcity of peacebuilding-related data.

The third technical dilemma is that the tradition of monitoring practice is based quite heavily on the use of quantitative data. Comparatively few monitoring tools use qualitative information to complement or to compare with quantitative data on the same issue/variable; and even when qualitative data are used, the tendency is to regard them as the equivalent of quantitative data where the latter are unavailable. The deeper problem here is that most monitoring tools are structured to provide a single measure on each selected indicator, and cannot easily accommodate sometimes conflicting information.

This in turn touches upon one of the more fundamental limitations of existing monitoring tools: their focus on producing average or consensus measures rather than reflecting potentially conflicting information and a diversity of perspectives. This point is particularly relevant for decision-making in peacebuilding operations where narratives may differ and data quality is routinely uncertain.

The tools for progress

Even to list the many tasks which must be fulfilled for the effects of peacebuilding to be properly measured could encourage a certain pessimism. But this would be wrong, for the reality is mixed and in many respects positive. A considerable number of the ideas and changes presented in this policy brief are already being discussed and taken into account in initiatives now underway. In this spirit, four recommended actions could greatly improve the development of tools for monitoring and evaluation in the peacebuilding field:

Build capacity to conduct monitoring and evaluation within organisations such as the UN

A review of monitoring or benchmarking practice within the UN reveals a huge gap between the organisation's formal benchmarking ambitions and its field missions' capacity to carry these out. At present, benchmarking relies to a large degree on individuals' dedication, competence, and capacity (including the capacity to direct funding for this purpose). As a result, the understanding and implementation of benchmarking varies widely from country to country. A related issue is the need to give a higher priority to establishing

benchmarking routines within the UN, including the allocation of earmarked resources and the creation of benchmarking units and clear responsibilities.

Promote a monitoring culture that takes diversity, uncertainty and risk into account.

There is in current monitoring practice a tendency to deliver information that draws on a single peacebuilding perspective, and produces consensus and average outcomes while omitting inconsistencies, uncertainties and diversities related to the data being analysed. This could be addressed by promoting system-wide assessment as the core focus of peacebuilding monitoring, with contributions from particular activities and actors examined from this perspective. This should be complemented by encouraging a monitoring culture which seeks a conscious balance between achievements and setbacks or risks, and where the reporting of failure should be seen as just as vital for future strategic planning as the reporting of success.

Develop data collection and databases in accordance with peacebuilding needs

At present, little data is collected for the precise purpose of monitoring peacebuilding activities. This contrasts with the area of development goals, where large databases and data-collection systems have been established (eg, the Devinfo¹⁴ system established with the purpose of monitoring the Millennium Development Goals). There is, however, a good opportunity to link peacebuilding data with the Devinfo system or similar existing databases. If this is combined with work to produce quality data that is peacebuilding focused and contextually relevant, the impact on the monitoring of peacebuilding efforts could be very beneficial.

New tools promise to deliver better information and outcomes for all concerned with peacekeeping and peacebuilding initiatives.

¹⁴ DevInfo, the database system endorsed by the United Nations Development Group for monitoring human development, <http://www.devinfo.org/>, accessed 10 May 2011.

***Implement the new perspectives
and guidelines in the area of peacebuilding
monitoring and evaluation***

In the near future, it will be crucial to verify and integrate the new perspectives and guidelines in the area of monitoring and evaluation. This is particularly important in providing more useful and effective tools in managing and assessing peace operations. Although the emerging consensus is that the new monitoring systems for peace operations (including indicators) need to be contextually founded, this

approach requires higher skills and greater resources among field practitioners compared to the more mechanistic approaches that have dominated up to now. Examples of how the guidelines can be applied in reality need to be produced which can then feed into more formalized methodological procedures; at the same time new perspectives, qualitative aspects, and contextual relevance should be taken into account. These new tools promise to deliver better information and outcomes for all concerned with peacekeeping and peacebuilding initiatives.

Further reading

Charles T. Call (2008), "Knowing Peace When You See It: Setting Standards for Peacebuilding Success", *Civil Wars*, Volume 10, Number 2, June 2008, pp. 173-194(22)

Charles T. Call and Elizabeth M. Cousens (2007), "Ending Wars and Building Peace", International Peace Academy, New York, http://www.ceinicaragua.org/posguerra/library/general/ENDING_WARS_Building_Peace_IPA.pdf

Charles T. Call (2010), "Beyond the 'failed state': Toward conceptual alternatives", *European Journal of International Relations*, 20 April 2010, 1-24, <http://ejt.sagepub.com/content/early/2010/04/16/1354066109353137.full.pdf>

Francisco Gutiérrez et al, (2011), *Measuring Poor State Performance: Problems, Perspectives and Paths Ahead*, Crisis States Research Centre, London School of Economics and Political Science, London

Ken Menkhaus (2004), "Impact Assessment in Post-Conflict Peacebuilding: Challenges and Future Directions", Interpeace, Geneva, http://www.peacecenter.sciences-po.fr/pdf/French_cycle/Sem_4/Menkhaus_Challenges_in_peacebuilding.pdf