INTERNATIONAL FORUM FOR THE CHALLENGES OF PEACE OPERATIONS

Impact Evaluation and Assessment of UN Peace Operations: What is the State-of-the-Art?

MICHELE LIPNER
ANN LIVINGSTONE

Occasional Papers
No. 6
JUNE 2015

THE INTERNATIONAL FORUM FOR THE CHALLENGES OF PEACE OPERATIONS

The Challenges Forum is a strategic and dynamic platform for constructive dialogue among policymakers, practitioners and academics on key issues and developments in peace operations. The Forum contributes to shaping the debate by identifying critical challenges facing military, police and civilian peace operations, by promoting awareness of emerging issues and by generating recommendations and solutions for the consideration of the broader international peace operations community.

The Challenges Forum is a global endeavor, with its Partnership encompassing Partners from the Global South and North, major Troop and Police Contributing Countries as well as the five Permanent Members of the UN Security Council.

Impact Evaluation and Assessment of UN Peace Operations: What is the State-of-the-Art?

MICHELE LIPNER
ANN LIVINGSTONE

Occasional
Papers
No. 6
JUNE 2015

ISBN 978-91-87927-08-9

Also available at http://www.challengesforum.org

Suggested citation:
Michele Lipner, Ann Livingstone, Impact Evaluation and Assessment of UN Peace Operations: What is the State-of-the-Art?, Challenges Forum Occasional Papers 2015:6 (June 2015).

Editors:
Annette Leijenaar, Jibecke Joensson and Andrea Rabus.

Printed in Sweden 2015 by Taberg Media Group AB, Laholm.

Contents

Preface

With more than 123,000 civilian, military and police peacekeepers serving in United Nations (UN) missions around the world, peacekeeping is currently at an all-time high, both in terms of numbers and the breadth and width of mission mandates.[1] And as with any other business, with more investments and investors come more demands for being able to present results, as well as to ensure that there are effective mechanisms in place for capturing and integrating lessons learned in future plans and operations. It is therefore no surprise that increasingly UN Member States and the international community at large call for more accountability and transparency of operations in order to measure the return on their financial and human investments in today's missions.

Moreover, the dire challenges facing the UN in the Central African Republic, South Sudan and Mali, to name but a few, further underline the importance for the UN to be able to quickly and effectively learn from experience and for the organization to be flexible enough to adapt to ensure constant improvement in its operations. To have an effective system of impact evaluation and assessment of peace operations is indeed a prerequisite for being able to meet the requirements of contemporary peace operations.

Recognising the importance of finding ways to better measure success of peace operations, or the lack thereof, the International Forum for the Challenges of Peace Operations (Challenges Forum) looked closer at impact evaluation and assessment of peace operations (2012-2014). It was one of four work strands that contributed to an overall Challenges Forum effort that resulted in a Report on *Designing Mandates and Capabilities for Future Peace Operations*, containing 24 targeted recommendations, which was presented to the UN Secretary-General at the UN headquarters in New York on 26 January 2015. The present Occasional Paper is one of the publications coming out of this project and complements two other discipline-specific

[1] United Nations, Peacekeeping Face Sheet as of 31 March 2015, http://www.un.org/en/peacekeeping/resources/statistics/factsheet.shtml (accessed 4 June 2015).

occasional papers published by the Challenges Forum in 2014, namely, *Evaluating Integrated Peace Operations* by Dr Jeni Whalen and *Challenges with Assessing Impact in International Police Reform and Assistance* by Mr Mark Reber.[2]

The present paper suggests that there is a lack of a common understanding of how to effectively measure the impact of peace operations; that is, open questions about what we are measuring, for whom and for what purpose. The paper aims to bring attention to what is arguably the state-of-the-art specific to impact assessment and evaluation of peace operations. It finds that increasingly, greater attention is focused on the means by which impact of peace operations is, or can, be measured. The overall conclusion is that there neither is, nor will there be, one right way of assessing effectiveness and evaluating impact. Rather, in this complex landscape that is peace operations today, measuring impact cannot be based on a certain number of pre-assumptions. It will necessarily be dependent on a number of questions to which the answers will always be context specific. We might therefore have to settle with a reality that 'good enough' ways in which to measure outcomes and impact may in fact be the more realistic, and therefore also best, option. Focusing on success-factors rather than full blown success-stories or operations is preferable and more useful to inform policy-making and it will arguably better address concerns about expectations on peace operations that cannot be met.

I would like to thank Dr Ann Livingstone, former Vice President of the Pearson Centre of Canada, and Dr Michele Lipner, former Adviser to the Pearson Centre and sponsored by the Australian Civil-Military Centre, for writing this important paper. Ms Annette Leijenaar, Division Head, Conflict Management and Peacebuilding of the Institute for Security Studies in Pretoria, provided invaluable input throughout the process. I would also like to acknowledge their combined contribution in pursuing these questions within the larger context of the *Designing Mandates and Capabilities for Future Peace Operations* project. My appreciation also goes

[2] Can be accessed on the following web pages:
Evaluating Integrated Peace Operations: http://www.challengesforum.org/en/Reports--Publications/Test/Occasional-Paper-No-1/?retUrl=/Templates/Public/Pages/PublicReportList.aspx?id%3d962%26epslanguage%3den%26r%3d2
Challenges with Assessing Impact in International Police Reform and Assistance: http://www.challengesforum.org/en/Reports--Publications/Test/Occasional-Paper-No-2/?retUrl=/Templates/Public/Pages/PublicReportList.aspx?id%3d962%26epslanguage%3den%26r%3d2

to our Partner Organizations contributing to the project effort at the Jordan Institute of Diplomacy, the National Defence College of Nigeria, the Foreign and Commonwealth Office of United Kingdom, the Norwegian Institute of International Affairs, the Swedish National Police, the Center for International Peace Operations of Germany and the Folke Bernadotte Academy of Sweden. Finally, I would like thank colleagues in the Challenges Forum International Secretariat, supporting the overall effort and the finalization of this publication.

The present paper aims to inform and inspire new thinking on how to make best use of the scarce human, financial and political resources available to UN peace operations at a time when several UN High-Level Panels are preparing to present their findings and recommendations stemming from comprehensive reviews of peace operations, political missions, the implementation of Security Council 1325 and the peacebuilding architecture. It is my hope that the paper will contribute to an inclusive and intensified dialogue on what to bear in mind when designing national, regional and international implementation processes of these recommendations. By bringing further clarity to some of the complexities involved in peacekeeping, I hope that the paper will contribute to an enhanced confidence in, greater support for, and thus, results from, UN peacekeeping as it remains a critical tool for international peace and security in the 21st century.

Annika Hilding Norberg
Director, Challenges Forum

June 2015

Executive Summary

This paper provides a broad overview of current thinking on the issue of assessment and evaluation of peace operations including a review of promising evaluation and assessment methodologies, tools and guidance. This is an area that has only recently gained significant traction, as donors, Member States, troop and police contributing countries and other key stakeholders are increasingly expecting value for money, accountability and transparency in peace operations. Without greater diligence and attention to effectiveness and impact of these operations, there is limited ability to capture lessons, identify good practices, gauge success—or failure—and improve the impact of current and future missions.

Assessment and evaluation of peace operations is a daunting task. These operations are political, extremely complex, context specific and mired in multi-layered cause and effect relationships. Even a shared understanding of what comprises success, effectiveness, what works or what does not work is absent, as is agreement ultimately over 'what matters' in peace operations. Nevertheless, emergent good practices, tools and methodologies can be identified, particularly if one looks at a number of key questions. Specifically, in peace operations what is being assessed or evaluated and at what level? For whom are assessments being undertaken? Why are the assessments being undertaken and how; that is, what tools and methodologies are available to promote quality evaluation and assessment.

Lessons to be learned and good practices are highlighted and include the following:

• Peace operations are, at their very root, about political processes Mission success or failure will have political implications for the UN, AU regional organizations, Member States and the host country. Recommendations and/or conclusions resulting from assessments and evaluations may lead

to changes in funding levels, ongoing constituency support, national security and stability, regional security and stability and a host of other considerations. As missions are affected by domestic and international politics, so too is the decision-making process around questions related to what should be measured, for whom and for what purpose. While issues around supporting transition to sustainable peace from economic, governance, security sector or rule of law perspectives are central concerns, politics must be factored into the discussion when answering key questions of the 'what', the 'why', the 'for whom' and the 'how'.

• Evaluation and assessment should be integrated into program planning, budgeting and implementation.

• The more successful programs are those that link objectives across sectors, and are thus more cumulative in their impact. Building evaluation and/or assessment frameworks around these inter-linkages not only increases awareness of the contributory nature of activities, but also allows for better understanding of this cumulative impact.

• No one method or approach is likely to provide significant information from evaluations or assessments of peace operations. The key is to be flexible in approach and accommodate a broader range of methodological options.

• Due to context, 'good enough' data may truly be 'good enough'. Data collection in conflict-affected and fragile states can be difficult at best. Not only is the issue of insecurity a factor, but so too are other considerations such as access to respondents and potential risk to evaluators and respondents who participate in evaluations.

• Robust conflict analysis must be central to the conceptual design of any and all programming, assessments and evaluations.

• Lessons to be captured and the identification of good (and bad) practices are best solicited when evaluations and assessments are framed around both learning and accountability. Accountability should aim to meet the requirements of a broad stakeholder base.

• To be more effective, those who seek to evaluate and assess—whether policymaker or practitioner—should have a shared understanding of

overall goals and objectives. In reality, a shared understanding is often absent and decision-making around assessment and evaluation will vary depending on stakeholder requirements and expectations. This decision-making will impact directly on the methodological approaches and tools used, as well as on the questions asked.

Abbreviations

CPPB Conflict prevention and peacebuilding intervention

DAC OECD Development Assistance Committee

DCAF The Geneva Centre for the Democratic Control of Armed Forces

DFID Department for International Development

DRC Democratic Republic of the Congo

FIB Force Intervention Brigade

IE Impact Evaluation

IPI International Peace Institute

M23 The March 23 Movement

M&E Monitoring and Evaluation

MONUSCO United Nations Organization Stabilization Mission in the Democratic Republic of the Congo

MSC Most Significant Change

NOREF Norwegian Peacebuilding Resource Centre

OECD Organization for Economic Co-operation and Development

ODE Office of Development Effectiveness

OM Outcome Mapping

UN United Nations

OIOS UN Office of Internal Oversight Services

PCC Police-contributing countries

ROA Rapid Outcome Assessment

TBIE Theory-based Impact Evaluation

TCC Troop-countributing country

1. Introduction

This paper is part of a series of Challenges Forum commissioned papers on impact evaluation and assessment of United Nations (UN) peace operations. The objective is to provide a broad overview of current thinking and literature on the issue of impact evaluation and assessment of these operations, including an examination of the methodologies and tools employed. While this paper takes a more theoretical approach to the topic, additional commissioned papers focus on 'practice', specifically the tools, methodologies and approaches that are being used via a sectoral and comprehensive/integrated lens.

To provide an overview of current thinking in terms of guidance, methodological approaches and tools, we asked the central question: why is this important?

2. Background

International conflict is increasingly characterised by intra-state violence that has resulted in substantive changes in how responses to end conflict are generated and carried out. Conflicts in the 21st century are protracted, often taking place in failed or fragile states, typified by violence between state and non-state actors, as well as among non-state actors themselves. Overall, they lack adherence to respected and/or accepted laws of armed conflict; there is a greater blurring of distinction between combatants and non-combatants; there is little respect for or distinction of interstate borders; and the use of violence aimed at civilian populations, particularly directed towards women and children has increased. Targeting international and local aid organiztions and personnel has also become more common. Yet, many of the very institutions needed to monitor and assess the impact of conflict and to mitigate its effects on the population and infrastructure are fragile, nascent or absent, leading to more suffering for the affected populations.

The impact of contemporary conflict on peacekeeping cannot be ignored. While it is widely accepted that peacekeeping is only one tool in the UN toolkit to respond to violent conflict, it has become the most widely used technique 'to preserve the peace, however fragile, where fighting has been halted and to assist in implementing agreements achieved by the peacemakers.'[3] The characteristics of modern conflict serve as a reminder that conflict is not linear in its evolution and that mechanisms required to address conflict do not follow sequential processes for their resolution and peace-building efforts.

The consequence for the international community has been the expansion of peacekeeping roles and responsibilities. Missions are increasingly mandated to undertake tasks identified with institution and state-building. In some cases,

[3] United Nations, Department of Peacekeeping Operations/Department of Field Support, *United Nations Peacekeeping Operations: Principles and Guidelines* (New York 2008), p. 18.

missions are mandated to use all necessary force in order to create the conditions for peaceful resolution to conflicts. For example, they may deploy in environments where there is still a peace to be had, as reflected in the recently established and deployed MONUSCO[4] Force Intervention Brigade (FIB) in eastern Democratic Republic of the Congo (DRC) to deal with M23[5] and other non-state actors. Additionally, it is no longer sufficient to just bring peace. A full menu of options is now required to help rebuild a fragile state into one that can develop the security, governance and economic infrastructure required to secure peace over time. Mandates increasingly reflect the realities of the complicated environment in which missions are deployed. As a result, they often use a range of peacebuilding tasks that support the transition to a viable state where the government and its institutions are seen to be legitimate and responsive to the needs of its population. Executive authority mandated to the UN mission has also been used to support the transfer of state authority in a post-conflict environment. Mentoring and supporting the re-establishment of state institutions, such as the police and judiciary, have also become part and parcel of mandates. Traditional peacekeeping remains an option, but the complexity of conflict results in multidimensional operations that bear little resemblance to the first UN deployments in 1948 and 1956.

Consequences of these changes are evident in the increased need for resources necessary for implementing the mandate. The calls for additional funding, however, are often met with silence, as are requests for appropriate equipment necessary for the mission to be successful in meeting the mandated tasks. Additionally, the global economic environment is not conducive to quick economic responses, particularly when taxpayers increasingly want to know where their money is being used, how effectively and to what end that specifically affects them in their lives. As a result, there is an increased expectation from donors and states of value-for-money, accountability, transparency and results that can be successfully conveyed to populations who are ever more sceptical of large international organizations and their effectiveness.

In light of the complex nature of contemporary conflict and the challenges of international responses to conflict, it becomes imperative to identify frame-

[4] United Nations Organization Stabilization Mission in the Democratic Republic of the Congo.
[5] The March 23 Movement, also known as the Congolese Revolutionary Army.

works that incorporate comprehensive norms, standards and codes that will provide a basis for determining how to measure the effect of international peace interventions. Recent emphasis at the UN indicates growing awareness of the need to develop training frameworks that ensure adherence and compliance to common standards which 'produces peacekeepers with the necessary skill sets.'[6] These frameworks can only be successful when there is agreement on what constitutes good training and good practice, as well as measures of success. As far back as the 2000 Brahimi Report, there was emphasis on a 'managerial' approach that would reinforce the need for a common standard in order for troop and police contingents to be similarly trained for international deployment.[7] Unifying the national training of troop (TCC) and police (PCC) contributing countries with UN standards was viewed as critical if a mission was to be successful in meeting the obligations of the mandate.

Fourteen years after the Brahimi report, the record remains mixed regarding the effectiveness of international peace operations. One of the driving factors in the need for measures of effectiveness has been the significant increase in budgets allocated for peace operations. As de Carvalho and Aune note, this has led to 'growing scrutiny of how this money is spent [...] and a growing emphasis on monitoring, evaluation and assessment of the results of these operations.'[8] Donors expect transparency and accountability regarding the funds allocated to international peace operations and auditing is now a routine part of programming; designed and implemented by development experts, humanitarian agencies and broader civil society stakeholders.

Despite recent attempts at 'whole of government', 'integrated approaches' and 'multidimensional' peace operations, the ontological divide among stakeholders remains. The recent global economic crisis has further highlighted the need for assessment and evaluation as resources are stretched between domestic political agendas and the need to respond to and manage international conflict and crises. Little agreement exists among stakeholders as to what projects, strategic plans, programs or activities will lead to better

[6] Alberto Cutillo, *Deploying the Best: Enhancing Training for United Nations Peacekeepers*, Providing for Peacekeeping No. 5, International Peace Institute (IPI), (New York, August 2013), p. 1.
[7] United Nations, *Report of the Panel on United Nations Peace Operations* (The Brahimi Report), (New York, August 2000).
[8] Benjamin de Carvalho and Ingrid J. Aune, *Assessing Complex Peace Operations: Some Considerations of Methodology and Procedure*, NUPI Working Paper 782, Norwegian Institute of International Affairs (2010), p. 1.

outcomes and impacts that can be measured as successful. There is even less agreement on what the lexicon means. The reality is that the ontological divide among sectors (military, police, civilian, governmental, humanitarian, development, public and private sector) will remain and will only be mitigated if dialogue leads to a shared understanding of what matters, how that is assessed and what impact there is on the overall landscape.

Some would suggest that given the complexities inherent in modern peace operations, the context specific nature of many of these operations and the funding nuances of Member States, a fundamental question remains, namely; should we be measuring impact at all? Given the long-term nature of building institutions that support sustainable peace and the relatively short funding cycles related to domestic political processes (which means less likely long-term commitment to support, mentor and partner with the state most recently out of conflict), should we attempt to measure impact in the first instance? The response is simple. We no longer have a choice. With greater expectations for transparency, accountability and impact of peace operations, we must, as a community, commit to this endeavour no matter how daunting the challenge. Assessing and evaluating the impact of peace operations will continue to be an expectation and requirement in order to improve planning and
manage missions, as well as to assist in the determination of lessons captured and lessons learned, good practices and the creation of an overall learning environment that will benefit all stakeholders who work in the service of peace.

3. Current Thinking and Trends

If we are to maximise quality responses and effectiveness in current and future missions, we must be better able to gauge what comprises success—or failure, how to measure what works—or does not, what are the good practices—or not, and what are the lessons learned—and those still to be learned. As the background section emphasises, regardless of the complexities inherent in measuring and assessing the efficacy and impact of peace operations, this does not negate our collective responsibility to ensure that each operation is of added value and working towards achieving its goals and mandates. Furthermore, that lessons to be learned are identified and incorporated into future missions and that assessment and impact evaluation become central components of the mission planning process.

While it is increasingly acknowledged that evaluation and assessment methodologies, guidance and tools are required for peace operations, this recognition has not necessarily translated into the actual development of a common set of tools, methods or guidance to achieve that end. The community of practitioners and policymakers has only recently begun to seriously grapple in concrete terms with the question of how do you measure or assess the effectiveness and impact of peace operations. This is somewhat expected given the multi-layered, multi-dimensional and ever more complex operations that have evolved over the last 20 plus years. This evolution has also been accompanied by revised mission structures that have sought to adopt a more integrated, 'one UN approach', adding to the challenges of creating greater synergies, not only in structure, but also in how programming is integrated and implemented. Thus, it is of no real surprise that the discourse on measurement and assessment is only recently evolving to address these complexities. And in this discourse, there is a movement away from the accepted traditional development approaches in evaluation and assessment towards an exploration of those that are more tailored for peace operations with their particular nuances, relationships and challenges.

The literature on peace and security is rife with commentaries and approaches touching on the broader categories of peacebuilding, crisis management and conflict prevention. The literature on assessing and evaluating the impact of UN peace operations or peace and security operations more generally (e.g. through the African Union) is still somewhat sparse with notable exceptions.[9] This more targeted literature recognises that the community of practitioners and policy-makers are at a crossroads where the need to find meaningful tools, methodologies and good practice guidance around 'measuring what matters' has become an imperative.

The literature outlines the shortcomings of existing methods and tools used to consider success and/or failure of peace operations. Criticisms range from the lack of methodological rigour and suitability, faulty conceptual frameworks and the absence of commonality in approaches (and language), to the lack of a shared vision on what is to be achieved and to what purpose. Many existing cause and effect models of analysis are viewed as of limited use for assessing peace operations. As Meharg notes, the multidimensionality of conflict is not conducive to 'two-dimensional measuring models based upon a cause and effect system of analysis (the logical chain). The multidimensionality of conflicts is steeped in variance and unpredictability.'[10] Stern et al argue that new methodologies are not being applied to complex development interventions; in fact, some of the most useful approaches to causal inference are neither generally known nor applied in the evaluation of complex development programs (e.g. multiple causality and configurations; theory-based evaluations).[11]

Other authors highlight conceptual shortcomings and some reflect on the learning and accountability gap that exists in many evaluations and assessments,[12] noting the need for balance 'between the use of evaluations to ensure account-ability on the one hand and to advance real time learning and programme

[9] See for example, Paul. F. Diehl and Daniel Druckman, *Evaluating Peace Operations,* (Boulder: Lynne Rienner Publishers, 2010).
[10] Sarah Jane Meharg, Measuring What Matters in Peace Operations & Crisis Management, (Canada: McGill-Queen's University Press, August 2009), p. 60.
[11] Elliot Stern, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies and Barbara Befani, *Broadening the Range of Designs and Methods for Impact Evaluations*, Working Paper 38, Department for International Development (April 2012), p. 38.
[12] Organisation for Economic Co-operation and Development(OECD), *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility: Improving Learning for Results*, DAC Guidelines and References Series (OECD Publishing: 2012); Cheyanne Scharbatke-Church, *Evaluating Peacebuilding: Not Yet All It Could Be,* Berghof Handbook for Conflict Transformation, Section II: Analysing Conflict and Assessing Conflict Transformation (Berghof Foundation: 2011).

adjustments on the other.'[13] At the same time, when the discussion of accountability is raised, it inevitably leads to the obvious next question, namely, accountability to whom—the Member States? Donors? Mission leadership? The host country? TCCs? PCCs? The answer to this question (of which there may be multiple responses) will further inform the discussion regarding what is to be measured and how this is to be done.

Definitional challenges over key concepts (e.g. stabilisation, state-building, peacebuilding etc.) also abound. Menkhaus notes, '[there is] the challenge of analysing the value of the core concepts at the heart of peacebuilding projects —peace, reconciliation, conflict, justice, good governance. These are subjective notions with multiple layers of meaning and competing definitions.'[14] Additionally, the language used to discuss issues relating to evaluation and assessment is itself not applied in any common fashion. For example, terms central to assessment and evaluation of peace operations are inconsistently used or understood—whether reference is made to impact, outcomes, effectiveness, monitoring, evaluation or assessment. Without a basic shared understanding, or at some point a common use in terminology, collective understandings of what works—or does not—may prolong the discourse, and compromise moving the field of evaluation and assessment in peace operations forward.

At the same time, there is the need to consider the politics of peace operations. The reality is that peace operations are, at their very root, political processes. Mission success or failure, as identified in assessments, will have political implications for Member States and the host country. Recommendations and/or conclusions resulting from these assessments and evaluations may lead to changes in funding levels, ongoing constituency support, both regional and national security and stability, and a host of other considerations. As missions themselves are affected by domestic and international politics, so too is the decision-making process around questions related to what should be measured, for whom and for what purpose. While issues around supporting transition to sustainable peace and all that entails from economic, governance, security sector or rule of law are central concerns,

[13] The Geneva Centre for the Democratic Control of Armed Forces (DCAF), *Measuring the Impact of Peacekeeping Missions on Rule of Law and Security Institutions*, Report of the Expert Workshop, (Permanent Mission of Switzerland to the UN, New York: 12 March 2012), p. 8.
[14] Ken Menkhaus, *Impact Assessment in Post-Conflict Peacebuilding: Challenges and Future Directions*, (Interpeace, 2004), p. 6.

there is always the 'elephant in the room' that must be taken into account when answering those key questions of the what, the why, the for whom and the how.

Making sense of the discussion suggests a need to return to first principles. What should we be measuring or evaluating and at what level—outputs, outcomes, impact? For what purpose are we measuring and to what end? How do we measure in order to glean meaningful results—what are the current methodological approaches under consideration? 'How do we know when a peacebuilding project has actually built peace? What indicators are most appropriate? What evaluation tools are most useful in assessing the theories of peacebuilding on which projects [or missions] are based?'[15] These are not easy questions to answer.

Michael Quinn Patton, 2010: Utilization-Focused Evaluation

What Should We Measure and at What Level?

Peace operations are highly complex with multiple stakeholders from the military, police and civilian space providing manifold inputs into numerous activities, projects and programs. All these inputs are intended to achieve or to support the achievement of the overall strategic objectives as mandated at the highest levels of international or regional decision-making bodies. At the

same time, missions do not operate in isolation of other key stakeholders—most significantly the host country—who will likely be undertaking concomitant activities, projects and programs outside of the mission space, either in support of mission activities or separate from these activities.

It is within this web of activity that assessment and evaluation of peace operations take place. Yet, what is to be assessed or evaluated, and at what level, is often contingent on answering two fundamental questions: to what purpose and for whom? As the Organisation for Economic Co-operation and Development (OECD) states, 'every evaluation […] should begin with the question: What is this evaluation meant to ascertain and how will this information be used?'[16]

Answering the 'why' question requires looking at issues focused on overall evaluation objectives. Is an evaluation being carried out to determine if an activity, project, programme or mission is achieving its objectives and as intended (accountability). And/or is an evaluation being undertaken to 'provide evidence and improve knowledge of results and performance, which can help improve on going or future activities and increase understanding of what works, what does not, and why.'[17] Ideally, assessments and evaluations in the name of accountability are conducted to ensure that mission objectives are accountable to multiple constituencies, including the host country, its population, donors, mission leadership and Member States. Assessments and evaluations undertaken for the purposes of learning usually seek to identify lessons to be learned and good practices that can be applied to current or future peace operations.

To gain maximum benefit, evaluations and assessments should seek to provide a platform for both learning and accountability (in its broadest definition). However, in part due to funding issues, donor requirements, security concerns and manpower, there will be compromises made in terms of what will be evaluated. Generally, it is far more the case that evaluations are done to support learning, whereas 'accountability mechanisms in peacebuilding are almost exclusively upwards in nature' as the focus on accountability is not directed towards recipients of the assistance, but more often than not towards the donors that fund or the Member States that support the

[16] OECD, 2012, p. 40.
[17] OECD, 2012, pp. 41-42.

interventions.[18] As a result, accountability becomes more politicised as political agendas and expectations often come into play.

In addition to the 'for what purpose', the question of 'for whom' is of fundamental importance as the variety of stakeholders will have different requirements (political or non-political) for information. For example, for a member state, the requirement may be for a 'good news' story or information that promotes a particular political agenda at home or justifies continued funding. For the practitioner, the 'why' may be in order to identify what works and what does not—either for learning and/or accountability purposes. For the UN, the requirement may be to determine factors leading to the absence of conflict. For a TCC the 'why' may relate to questions pertaining to troop drawdown. For the host country, the purpose of an evaluation or assessment may be to determine popular satisfaction with the mission. For donors, the need may be to explain to taxpayers whether the money invested in a peace operation has been well spent.

Preferably, stakeholder requirements are, at the very least, complementary so that assessments or evaluations can accommodate these different expectations. In practice, this is not always the case. As a result, the questions posed in evaluations may not, in fact, be those that will necessarily yield the most robust findings, thus leading to only a modicum of either learning or accountability.

What to Measure?

Even when it is clear why an evaluation or assessment is being carried out and for whom, the next question of 'what is to be measured' is itself complicated. There continues to be significant debate and discussion over what is to be measured—or should be measured, or even more centrally, is measurable in the first instance. Meharg, as the title of her book suggests, essentially calls to 'measure what matters', reflecting that we are far better at identifying what does work rather than what does not work. Others have suggested not asking if something worked, but rather to ask if the activities made a difference.[19]

[18] Scharbatke-Church, 2011, p. 474.
[19] See for example, Stern et al., 2012.

Diehl and Druckman argue for a focus not only on the conditions of success, but also on the mechanisms responsible for the outcome. They note the lack of definitional clarity and absence of consensus on what constitutes 'peace operation' success in the first place. As they conclude, '[d]etermining what constitutes success or failure in peace operations is a prerequisite for building knowledge about the factors associated with those conditions.' They add 'most studies focus on the factors thought to produce success rather than devoting attention to the criteria used to assess that success.' And most importantly, in order to move forward, a clear definition of success is required; 'if we do not know what constitutes success, it will be difficult to ascertain what conditions produce that effect.'[20]

In all cases, there is a significant degree of subjectivity and variance in definition, whether speaking in terms of what works, what makes a difference, what matters or what constitutes success. Success is often in the eyes of the beholder, suggesting that what indeed matters is in fact partially (if not fully) dependent on for whom and for what purpose the evaluation/assessment is being conducted. For the TCC, success and the focus of interest in an evaluation, may be the elimination/reduction of insurgency activities and thus the end state required for troop withdrawal or a drawdown in troop numbers. For PCCs, success may be a functional court system or the establishment of community policing with a similar end state. Success through a civilian lens may be the provision of basic services, a functioning economy or governance structures. For the mission as a whole, it may be the absence of conflict, and the likelihood of sustainable peace, or stability—words that are themselves ambiguous. For the host country, what matters may be public support and what is deemed as successful is the establishment of a legitimate and stable government. Ideally, it would be of benefit to have at least a shared understanding of what matters. In the absence of shared or common understanding, careful consideration is required to construct the assessment and evaluation questions for determining 'success' for the different stakeholders involved.

Adding another level of complexity is the fact that there are multiple potential entry points for assessments and evaluations. Yet which are more likely to yield the type of information required in order to gauge positive (or negative) change and/or impact? Is it specific mission activities, projects

[20] Diehl and Druckman, 2010, p. 5.

or programs? Is there interest in one set of interventions in relation to one program, or a combination of interventions in relation to multiple programs? Should the focus be on the mission as a whole, taking a more comprehensive and integrated approach and looking at how the projects and programs therein achieve the overall strategic objective(s) of that mission? This is probably the most complex question as it goes to the heart of why it is so difficult to evaluate and assess peace operations. Can a program be assessed in the first instance in isolation from other key program areas? For example, can rule of law programs be evaluated or assessed without taking into consideration other program areas, such as security sector, governance or economic reform? As any number of practitioners have highlighted and as common sense dictates, the more successful programs are those that link objectives in one sector to those in another, thus bringing the sectoral objectives into alignment with overall mission objectives, as would be the objective of integrated missions. Bennett et al succinctly argue, 'the more successful initiatives are those that have linked objectives in one sector to those in another and hence have been able to follow through with tracing the cumulative effects of the various activities on conflict and peace.'[21] Yet as de Conig and Romita caution, 'system-wide evaluations pose particular challenges because the different actors involved in the intervention have different worldviews, mandates, theories of change, timeframes, organizational cultures, planning processes, methods of work, and approaches to measuring progress.'[22]

Because missions are so complex and the component parts often so interrelated, the question of what to measure appears to be an insurmountable behemoth. This can be even more complicated by the fact that the 'what' in evaluations often changes as goals and objectives change or are adjusted over the course of the mission due in part to political considerations. Thus, the 'what' is frequently a moving target. This can easily lead to the conclusion that because it is so hard, it cannot be done. However, as already stated, 'too hard' is no longer viewed as an acceptable fall-back position. No matter how difficult, methodological approaches and tools are being developed, adapted and refined to absorb and account for the complex relationships and agendas

[21] Jon Bennett Sara Pantuliano, Wendy Fenton, Anthony Vaux, Chris Barnett and Emery Brusset, *Aiding the Peace: A Multi-donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan 2005-2010* (United Kingdom: ITAD Ltd., December 2010), xviii f.
[22] Cedric de Coning and Paul Romita, *Monitoring and Evaluation of Peace Operations*, IPI, (November 2009), p. 7.

that not only are part of mission planning and implementation but of assessment and evaluation as well. These approaches increasingly use a multiplicity of tools in order to create a more robust picture of 'causal' or correlated relationships and processes that impact on mission progress, effectiveness and impact.

At What Level

Commentary regarding 'what to measure' invariably includes discussion on the level of analysis, commonly described as outputs, outcomes or impact. OECD definitions are most cited by policymakers and practitioners engaged in conflict management in fragile states. Additionally, the UN has built on these definitions by adding contextual relevance in relation to monitoring peace consolidation.[23] The definitions are mutually reinforcing and are provided in Box 1 (page 16).[24]

The level or levels at which evaluation and assessment occur depend on a number of factors, not least of which are those relating back to questions of what do you want to know and for what reasons. Additional factors may relate to funding and time constraints. Fundamentally, '[t]here is often lack of clarity within the literature on what should be measured and when. This relates to basic theoretical disagreements on the level of the results chain at which impact is located.'[25]

A number of authors highlight that evaluations are often focused on outputs rather than what are considered the more important questions of outcomes or impact.[26] They note that the lack of common definition and/or understanding of overarching objectives and goals discourage taking analysis beyond the level of outputs. Sherman writes, ' [...] in the absence of clarity on high-order goals, it often remains easier […] to focus on individual outputs rather than their collective outcome.'[27] Does current thinking see a role for measuring

[23] United Nations, *Monitoring Peace Consolidation: United Nations Practitioners' Guide to Benchmarking*, (2010), p. 19.
[24] The OECD definitions for outputs and outcomes and the partial definition of impact are based on: OECD, DAC Network on Development Evaluation, Glossary of Key Terms in Evaluation and Results Based Management, (Paris: 2002). They have been reproduced and expanded upon (impact) in the OECD 2012 publication, p. 13.
[25] Vincenza Scherrer, *Measuring the Impact of Peacebuilding Interventions on Rule of Law and Security Institutions*, SSR Paper 6, (DCAF: 2012), p. 48.
[26] See for example Diehl & Druckman, 2010; Meharg 2009; Scherrer, 2012.
[27] Jake Sherman, *Measuring Effectiveness in Peace-Building and State-Building*, Chapter 13 in Meharg (2009), p. 210.

Box 1: OECD and UN Benchmarking Definitions of Outputs, Outcomes and Impact

Level of Analysis	OECD Definition	UN Benchmarking Definition
Output	The products, capitals goods and services, which results from a conflict prevention and peacebuilding intervention.	The direct results of an input activity. Peace consolidation evaluations at the project, programme or sectoral levels assess whether anticipated outputs have been attained through the provision of inputs such as financial resources, technical assistance or training.
Outcome	The likely or achieved short-term and medium-term effects of an intervention's outputs.	The wider short- and medium-term effect (positive and negative) of the input activity. At the strategic level, peace consolidation has to focus on outcomes that are directly linked to reducing certain risks (including through the development of conflict management mechanisms) addressing critical conflict drivers and advancing towards desired goals through the implementation of mutual commitments.
Impact	[For peacebuilding]: Positive or negative, primary and secondary effects produced by an intervention, directly or indirectly, intended or unintended (OECD, 2002). Results that lie beyond immediate outcomes or sphere of an intervention and influence the intensity, shape or likelihood of a conflict.	The long-term direct and indirect effect (positive or negative) produced by operationalizing of the strategic vision. These should reflect the goal embedded in that vision. System-wide impacts are normally heavily affected (positively or negatively) by factors outside of strategic control, e.g. other social change actors and initiatives, interactions with neighboring countries and various unforeseen social and environmental changes.

outputs in these operations? Schumacher suggests that 'the benefit of using output indicators is that they are usually amenable to measurement and thus make it easy to affirm achievement.' The natural thought progression is that causal attribution can be ascertained by linking outputs to outcomes and the use of baseline data allows evaluation to infer progress in outcomes, to the progress in outputs by tracking their correlation through statistical analysis'.[28] However, this model assumes that attribution is possible to ascertain by isolating cause and effect relationships—assumptions that are difficult to achieve. Further, it is more likely that while outputs may be quantifiably measurable, they provide limited utility in relation to generating information regarding change in behaviour or actions. Nevertheless, acknowledging that evaluation within peace operations can be highly political, it is far 'safer' to focus on outputs, as they make no judgment on value or contribution to objectives except in the most rudimentary fashion.

Defined largely as the short and medium-term effects (positive and negative) of input activity, outcomes are more meaningful as they allow for information, and a level of assessment and evaluation in terms of answering the question 'has intervention/activity/program X made a difference in the short or medium term?' Moving further up the results chain, looking at the level of impact is undeniably a complicated proposition in peace operations. Measuring impact has largely been viewed as too problematic and too elusive, given its focus on higher level mission goals and its longer term horizon for yielding significance in relation to assessing whether change or progress is embedded or not. Scherrer notes that '[i]mpact has often been perceived as a particularly elusive level of the results chain where the contribution of an intervention cannot be proven. Furthermore, there is a tendency to perceive impact as being visible only several years after an intervention and therefore as too long to be measured effectively for the purpose of programming and policy.' Yet, drawing on the 'Collaborative Learning Project's Reflecting on Peace Practice', Scherrer adds that impact need not be assessed after the completion of an intervention or several years thereafter. Instead 'there is a growing view that impact can be measured in the more immediate term. This emerging approach offers opportunities for international actors that need to measure impact but cannot wait until the end of an intervention to review much needed information on what is working, what is not, and

[28] Joseph Schumacher, 'What to Measure in Peace Operations', *Measures of Effectiveness: Peace Operations and Beyond*, The Pearson Papers, 10(1): 45-59 (2007), p. 51.

why'.[29] In this fashion, attention is drawn to the fact that impact evaluation can take place both during project and program implementation as well as after. AusAID[30] argues that impact evaluation programs should not only be built into the design of an intervention, but in fact, be conducted throughout implementation, or at least during or when the intervention is complete. As the aid agency noted, 'impact occurs at multiple levels and timeframes—there can be short-term, intermediate and long-term changes resulting from an intervention. How and when impact occurs will differ depending on the type of intervention and the context.'[31]

There are still those who suggest that rather than focus on impact, which requires a shared understanding of the end state or even an identification of an end state, attention should be concentrated on assessing the relationship between drivers of conflict (the issues to be addressed by mission inputs) and outcomes. This approach looks at assessing whether or not a contribution has been significant. According to Brusset, this allows a move away from theory-based evaluation (methods that seek to reconstruct the logic of an intervention) towards conflict analysis based evaluation that focuses on the design of a conflict prevention and peacebuilding intervention (CPPB). Interventions are then examined by 'exploring three aspects that allow verification of the significance of the contribution', specifically relevance, extent and duration.[32]

Attribution versus Contribution

No discussion on evaluation and assessment in peace operations can ignore the questions of attribution versus contribution. Attribution links an effect directly to a cause. Contribution, on the other hand, does not assume a direct cause-effect relationship but looks at the contributions of an activity or series of activities to a particular end state. So, for example, rather than stating that a law and justice program in country X caused a reduction in crime (attribution), the argument would instead be that the law and justice program, along with other inputs/interventions, contributed to a reduction in crime. While perhaps a subtle difference, contribution acknowledges the

[29] Scherrer, 2012, p. 12.
[30] In September 2013, AusAID was merged into the Australian Department of Foreign Affairs and Trade. While no longer a separate government agency, it is assumed that guidance documents on aid effectiveness remain relevant.
[31] Australian Department of Foreign Affairs and Trade, Office of Development Effectiveness (ODE), *Impact Evaluation: A Discussion Paper for AusAID Practitioners* (September 2012), p. 2.
[32] E. Brusset, *Significance of Impact Assessment: A New Methodology*, Chapter 11 in Meharg, 2009, pp. 196-197.

complexities and multi-faceted nature of cause-effect relationships in peace operations (and peacebuilding more generally).

OECD notes, '[i]t can be particularly difficult to establish clear attribution and causality in settings that are complex and where changes for peace (or renewed violence) are often non-linear and unpredictable.'[33] Given this complexity, is it even possible to deconstruct the web of inter-related cause and effect relationships and determine causality? Should we instead consider the contributions of intervention(s) towards a desired outcome(s)? Scherrer states that, '…when it comes to measuring impact, there is a debate about the validity of attribution vs. contribution. Attribution is often promoted as the 'gold standard' because of its ability to demonstrate a direct causal link between an intervention and its impact. However, in complex post-conflict settings it is considered extremely difficult to isolate the effects of a particular peacebuilding intervention and thus to establish a causal link between the intervention and the observed outcomes and impacts.'[34] Brusset adds that 'the evaluation of impact in peacebuilding seeks to attribute impact to intervention but never succeeds. This is because the connections among activities, results and objectives are loose, and many other factors and agencies cut across with their own particular influence on the overall dynamics. The ability to attribute CPPB cause and effect relationships remains a myth.'[35]

However, this does not mean that there have not been efforts to attribute and identify causality. This process often involves adoption of the more 'scientific/experimental approach' utilising any number of methods including counterfactuals.[36] The question remains whether these efforts are of added value as standalone methods or do they instead risk being reductionist, thereby failing to capture the complex interdependencies and interrelationships of inputs into the broader peace and security goals and objectives? Stern notes that 'statistical and econometric models can have difficulties with multiple causality and struggle to capture the interactions among variables or represent irregular, complex paths […] it can be difficult

[33] OECD, 2012, p. 8.
[34] Scherrer, 2012, p. 8.
[35] Brusset in Meharg, 2009, p. 199.
[36] Counterfactual is 'a situation or condition which hypothetically may prevail for individuals, organisations, or groups were their no intervention, e.g. the war that would have occurred had a peacebuilding intervention not taken place'. See OECD (2012), p. 11.

for these methods to capture the influence of combinations of causal factors rather than of each causal factor as a free-standing agent.'[37] Menkhaus adds that 'assessment of peacebuilding in post-conflict settings is constrained by 'counterfactual' reasoning—namely, the impossibility of predicting what would have happened in the absence of the peacebuilding intervention. Evaluators are consistently forced to compare that which happened with that which would have happened, an exercise that is essentially speculative, not empirical, in nature.'[38]

While the debate over attribution and contribution continues, current thinking leans towards greater focus on the contributory nature of inventions to outcomes and impacts. Scherrer states that, '[f]ocusing on contribution as opposed to attribution recognises that there may be other factors that have also contributed to the observed impact. This is particularly relevant in post-conflict contexts as it takes into account the complexity of 'tracking causality' in the non-linear multi-agency contexts' within which peacebuilding support takes place.'[39] Thinking in terms of a 'causal package' recognises that an 'intervention plus other factors' is a far more meaningful way of looking at impact evaluation in more complex settings.[40]

How Should We Measure?

The question of how do we measure progress and effectiveness of peace operations is a discussion of methodological approaches and tools. There is no one current trend or best practice in relation to assessment and evaluation approaches or methodologies proposed for peace operations. As there is no one template for a peace operation, there is no 'one size fits all' approach to how impact or effectiveness is assessed. As Meharg suggests, 'there is no right or best approach to intervention and therefore no right approach to measuring success… there are only better ways to think, and that is what matters in measuring peace operations and crisis management.'[41] Diehl and Druckman note that 'even when analysts identify a conceptual definition of success, the operational measurement of that definition is often lacking or is

[37] Stern et al., 2012, p. 40.
[38] Menkhaus, 2004, p. 5.
[39] Scherrer, 2012, p. 8.
[40] Rephrased from Stern et al., 2012, p. 40.
[41] Meharg, 2009, p. 15.

suboptimal.' In fact, 'in a majority of extant works, there is an absence of indicators and often a lack of any conceptual specification of success behind them.'[42]

Stave raises the following concern shared by others as well.[43] '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'.[44] Why is theory of change important in this context?[45] Largely because a theory (or theories) of change implicitly defines what factors will create change. For peace operations, this means that how a mission is formed implicitly builds of certain theories regarding how change (sic peace) occurs.

Notwithstanding the challenges, there are a number of tools and methodologies available[46], some still evolving, and others based on methods embedded within the aid and development sector. Scherrer suggests a range of methodologies for both attribution (e.g. impact evaluation and theory-based impact evaluation) and contribution (e.g. contribution analysis, outcome mapping, rapid outcome assessment (see Box 3) and most significant change). To the former, the focus is on the use of counterfactuals and control groups, the latter on theory-based and participatory approaches. The Geneva Center for the Democratic Control of Armed Forces (DCAF) goes on to suggest that 'the UN should consider 'real world' approaches that do not necessarily require statistical counterfactuals and that may entail mixed methods.'[47]

[42] Diehl and Druckman, 2010, pp. 7; 12.
[43] See, for example Diehl and Druckman, 2010.
[44] Svein Erik Stave, *Measuring Peacebuilding: Challenges, Tools, Actions*, Norwegian Peacebuilding Resource Centre (NOREF), Policy Brief No. 2 (May 2011), p. 3.
[45] As defined by OECD, 2012, theory of change is 'the understanding of how a specific activity will result in achieving desired changes in a particular context—it is the logic that underlies action.' (p. 11) It is essentially the 'set of beliefs about how change happens and, as such, it explains why and how certain actions will produce the desired changes in a given context, at a given time.' (p. 31).
[46] See for example Stern et al., 2012, Chapter 4.
[47] DCAF, 2012, p. 3

Within the Australian aid program, there are calls for evaluation approaches that include realist evaluation, contribution analysis, general elimination methods, comparative case studies and process tracing, arguing that 'these approaches […] are more likely to be suitable for […] programs in complex or fragile situations or cases where an intervention is one of multiple causal factors leading to one or more outcomes.' Also highlighted is 'the main advantage of these approaches to impact evaluation is that they provide in-depth explanation of 'how', 'why' and 'for whom' an intervention has contributed to development change. By testing a theory or establishing the 'mechanisms' for change, they are also more likely to provide findings that can be applied in different contexts.'[48] Stern et al offer three main design approaches that are not yet widely used in impact evaluations, but appear to offer help in linking interventions with outcomes and impact: theory based approaches, case-based approaches and participatory approaches.[49]

Diehl and Druckman provide their own framework for evaluating peace operations. The framework does not speak directly to impact or outcome, but to success, with the component parts of the model built around goals, questions and indicators. The decision-making allows for identification of the primary goals of an operation to specification of appropriate measures of progress (quantitative and qualitative). It assesses progress towards attainment of the core goals of the mission. Importantly and as they noted, 'the template addresses the way in which possible indicators of success derive from practical questions asked about missions.'[50] The model is simple, elegant and intuitive.

Scherrer provides a broad-brush overview of evaluation approaches, reflecting on the utility and shortcomings of each. Although his work focuses more specifically on rule of law and security institutions in peacebuilding interventions, the overview has broader conceptual relevance and while not comprehensive, is illustrative of the choices available and their potential applicability.

[48] ODE, 2012, pp. 4-5.
[49] Stern et al., 2012.
[50] Diehl and Druckman, 2010, pp. 25-26.

Box 2: Illustrative Overview of Evaluation Approcahes*

Type	Approach	Application
Attribution	Scientific-experimental	Claims attribution use of counterfactual analysis.
Contribution	Theory-based	Supports contribution by testing assumptions at each level of theory of change; theory based evaluation seeks to identify and test causal pathways.
	Participatory	Support contribution by listening to perceptions of the beneficiaries of what initiatives have made a difference in their lives.
Non-causal	Action evaluation	Supports the collective definition of goals—therefore helps to identify jointly what impact should be measured.
	Goal-free evaluation	Examines the 'actual' impacts of an intervention by deliberately avoiding knowledge of the intended goals and objectives. Does not support attribution or contribution.
	Results-based evaluation	Seeks to measure impact to the extent that it focuses on that level of the results chain (i.e. with the use of indicators); it 'examines changes through time of multiple relationships between inputs and outputs. Results are not end states but variations in behaviour and performance during a process.' Does not support attribution or contribution.
	Utilisation-focused	Can address impact and depending on methods and the designated use of evaluation. Does not support attribution or contribution.

* Scheerer, 2012, p. 13.

Box 3: Overview of Methodologies for Measuring Impact*

Type	Methodology	Methods
Attribution	Impact Evaluation (IE)	Quantitative methods such as control groups (e.g. randomized control trials) and before/after comparisons, statistical modelling, econometrics
	Theory-based Impact Evaluation (TBIE)	Quantitative and qualitative methods such as control groups and before/after comparisons combined with theory of change approaches
Contribution	Contribution Analysis	Qualitative methods such as case studies, most significant change (MSC) stories, focus group discussion
	Outcome Mapping (OM)	Qualitative methods such as focus group discussion, workshops and use of 'progress markers'
	Rapid Outcome Assessment (ROA)	Draws on outcome mapping methodology, MSC technique and episode studies
	Most Significant Change (MSC)	Qualitative methods such as group discussions, interviews and workshops

* Modified from Scherrer, 2012, p. 18.

In practice, these approaches are applied using a variety of methodologies. The methodologies chosen for any one approach will also be dependent on the questions that need to be answered in an evaluation or assessment, which in turn is based on the purpose of the evaluation or assessment (i.e. learning or accountability). Box 3 provides illustrative examples of methodologies for measuring impact, categorised according to whether they are intended to support approaches that seek attribution or contribution.[51]

In practice, actual evaluations and assessments of peace operations have been limited. Yet, there are a number of illustrative examples of how some of these methods have been applied. For example, the UN Office of Oversight Services (OIOS) has carried out a number of mission reviews to 'evaluate the performance and achievement of results to determine the relevance, efficiency and effectiveness in terms of mandated objectives.'[52] These evaluations use a mix of qualitative and quantitative methods drawing on a number of data sources, including structured interviews, stakeholder surveys, population surveys and desktop reviews. Gilligan and Sergenti take a far more quantitative approach by using matching techniques on a sample of UN missions in order to evaluate if UN peacekeeping interventions have had positive effect.[53] Braithwaite draws on Diehl and Druckman's model to evaluate the Timor-Leste operation.[54] Autesserre carried out an extensive evaluation of international peacebuilding efforts in the DRC utilising rigorous qualitative analysis inclusive of comprehensive data collection and hundreds of interviews with multiple respondents spanning UN officials to victims of violence.[55]

As noted throughout this paper, answers to the questions 'for whom' and 'for what purpose' will inform the methodology, the design of the questions, as well as the analysis that will follow. If we are generally interested in the causal linkages within peace operations that impact on outcomes and can assist in both learning and accountability, it might be that whatever methodologies are used, they should be framed around the following questions:

[51] For further information on different methods used to establish causality, see Table 3.3 in Stern et al. (2012), p. 24.
[52] See https://oios.un.org/.
[53] Michael J. Gilligan and Ernest J. Sergenti, *Evaluating UN Peacekeeping with Matching to Improve Causal Inference* (New York University: October 2006).
[54] John Braithwaite, 'Evaluating the Timor-Leste Peace Operation', *Journal of International Peacekeeping,* vol. 16, (2012), pp. 282-305.
[55] Séverine Autesserre, *The Trouble with the Congo*, Cambridge Studies in International Relations, (Cambridge University Press: 2010).

1. To what extent can a specific (net) impact [or outcome] be attributed to the intervention?
2. Did the intervention(s) make a difference?
3. How has the intervention made a difference?
4. Will the intervention work elsewhere?[56]

Approaches are still evolving, as there is increasing realisation that outcomes and impact of UN peace operations are not easily quantified or explained in simple cause-effect links. Single method approaches are unlikely to yield the type of information required and nor is it likely that an either or approach—in relation to quantitative versus qualitative methods—will satisfy. What is gaining significant traction is a call for mixed methods rather than any one method that is either quantitative or qualitative. The use of multiple methods helps add validity to the findings and will raise the bar in inferring if not causality, at least significant contribution of specific interventions towards the achievement of goals or objectives. Indeed, by broadening the evaluation scope beyond the attribution goal-centred model, there can be greater discussion around 'plausible contribution' and/or how to create better synergy amongst those tools that seek to attribute and those that seek contribution. Perhaps it is best to remember that 'the real choice is not so much between empirical versus non-empirical methodologies as it is between thoughtful, rigorous, and pragmatic approaches to project evaluation versus simple-minded, bureaucratic, and dogmatic techniques.'[57]

Indicators

Even with the best methodological approaches and tools, the question of measurement is made all the more problematic as we seek to identify meaningful indicators to measure success or failure, what works or does not work, and/or to assess if goals and objectives are achieved. Peace operations often aim for more 'intangible' outcomes or impacts that are not readily amenable to measurement. Menkhaus adds, 'Inasmuch as peacebuilding is aimed at preventing recurrence of conflict, efforts to measure project impact are also confounded by the fact that the 'dependent variable' involves something that—if the project is successful—does not happen [i.e. a return to conflict]'.[58]

[56] Stern et al., 2012, p. 37.
[57] Menkhaus, 2004, p. 8.
[58] Menkhaus, 2004, p. 5.

What then are some of the considerations that should or could help inform the development of indicators? Drawing on themes emerging from the literature, a number of 'good practices' have been identified and include the following:

1. Evaluation should be framed around key thematic areas, such as, relevance, effectiveness, efficiency, sustainability, impact, and more recently the criteria of coherence and coordination/linkages.[59] From these thematic areas, indicators should flow.
2. Evaluations and indicators must be conflict and context sensitive. Conflict and context sensitive indicators are needed to help identify both positive and negative impacts of interven tions.
3. Indicators must be embedded within sound methodological approaches for measuring impact; indicators on their own 'are not a substitute for robust methodologies for measuring impact [...They are] useful for monitoring purposes but do not provide adequate toolsets for assessing impact because they are not able to assess attribution or plausible contribution.'[60]
4. The dilemma of choosing 'universal' or context specific indicators must be considered. As Stave cautions, '[w]hen 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'. He goes on to argue, 'the value of unique contextual indicators is that they are generally based on in-depth knowledge of local conflict and culture, together with a cre ative understanding of the contextual signals that reflect the condition and development of peace in a society.'[61] At the same time, universal indicators may be more appealing to Member States as they seek to justify and account for their spending on peace operations to their various constituencies.

[59] See OECD, 2012.
[60] DCAF, 2012, p. 12
[61] Stave, 2011, p. 4.

4. Lessons to be Learned/Good Practices

Based on the discussions above and in consideration of the literature on evaluation and assessment of peace operations—or more broadly the issue of peacebuilding—a number of 'good practice' guidance points are offered.

- Start early and stay focused. Evaluation should be integrated into program planning and budgeting.[62] As OECD argues, 'when evaluation and its requirements are an integral part of programming activities from the outset, it contributes to more effective programming and facilitates better evaluation'.[63] This also ensures that impact can be better assessed during implementation.

- Evaluation and assessment should be sensitive to the inter-linkages between activities, projects and programs in their analysis. The more successful programs are those that link objectives in one sector to another and are thus more cumulative in their impact. Building evaluation and/or assessment frameworks around these inter-linkages not only increases awareness of the contributory nature of activities, but also allows for better understanding of this cumulative impact.

- No one method or approach is likely to provide significant information from evaluation or assessments of peace operations. Instead, there is value in considering the use of mixed methods (qualitative and quantitative) as they will likely yield more robust information and provide a broader sense of understanding around what works, what is effective (and not) and what has made a difference. The key is to be flexible in approach and accommodate a broader range of methodological options.

[62] See OECD, 2012; Department for International Development (DFID), *Working Effectively in Conflict-affected and Fragile Situations*, Briefing Paper I: Monitoring and Evaluation (March 2010); United Kingdom Stabilisation Unit, *Stabilisation Issue Note: Monitoring and Evaluation* (2011).
[63] OECD, 2012, p. 29.

- Due to context, 'good enough' data may truly be 'good enough'.[64] The reality is that data collection in conflict-affected and fragile states can be difficult at best. Not only is the issue of insecurity a factor, but so too are other considerations such as access to respondents and potential risk to evaluators and respondents who participate in evaluations.

- Robust conflict analysis must be central to the conceptual design of any and all programming, assessments and evaluations. It not only provides the framework for identifying core goals, objectives and outcomes, but it also helps form and inform the monitoring and evaluation framework for the mission—inclusive of the component parts of that mission.[65] Context must be understood along the full spectrum of planning, implementation and evaluation.

- Similarly, there must be situational awareness of the drivers of conflict. As any number of authors argue, drivers of conflict are not well understood and donors often fail to address core peace drivers and conflict-mitigating factors. Within evaluations, conflict analyses will help 'assess the relevance and impact of the program […] assess the risks of negative effects of conflict on the evaluation design and process; and […] assess the risks of the evaluation exacerbating conflict.'[66]

- In addition to conflict analysis, 'all activities, whether explicitly aimed at peacebuilding or not, should be examined to assess their conflict sensitivity'.[67] Conflict sensitivity is a central principle that OECD advocates in terms of evaluation of peacebuilding activities in fragile environments and is defined as 'systematically taking into account both the positive and negative impacts of an intervention, in terms of conflict or peace dynamics,

[64] See for example, *Working Effectively in Conflict-added and Fragile Situations*, Briefing Paper I: Monitoring and Evaluation, A DFID Practice Paper, (March 2010).
[65] See for example OECD, 2012; Bennett et al., 2010; David Kilcullen, *Measuring Progress in Afghanistan,* unpublished essay, 2009.
[66] Marie Gaarder and Jeannie Annan, *Impact Evaluation of Conflict Prevention and Peacebuilding Interventions*, Policy Research Working Paper 6496, The World Bank, (June 2013), p. 11.
[67] OECD, 2012, p. 37.

on the contexts in which they are undertaken, and, conversely, the implications of these contexts for the design and implementation of interventions.'[68]

- The host country/beneficiaries should be involved in the evaluation and assessment process. This has two components; first, the host country, where possible and feasible, should be more involved in evaluation and assessment design and implementation. Second, far greater effort must be made in gathering and analysing information on public perception, sentiment and confidence regarding the mission and its components.

- Lessons to be captured and the identification of good (and bad) practices are best solicited when evaluations and assessments are framed around both learning and accountability. In this, accountability ideally seeks to achieve the requirements of a broad stakeholder base.

- A common language or at the least, common understanding, should be encouraged for key stakeholders engaged in peace operations. As de Conig and Romita note, the 'M&E systems of the major stakeholders and disciplines still lack a common vocabulary or approach. Without it, the actors that undertake development, security and political action find it difficult to develop common understanding of the context within which they operate and this has negative implications for their ability to develop coherent strategies, and for their ability to monitor and evaluate progress towards achieving such strategies.'[69]

- To be more effective, those who seek to evaluate and assess—whether policymaker or practitioner—should have a shared understanding of overall goals and objectives. In reality, shared understanding is often absent and decision-making around assessment and evaluation—particularly in peace operations where multiple stakeholders provide multiple inputs—will vary depending on stakeholder requirements and expectations. This decision-making will impact directly on the methodological approaches and tools used as well as the questions asked.

[68] Ibid., p. 11.
[69] de Coning and Romita, 2009, p. 16.

5. Conclusion

There are a number key messages emerging from this paper that can serve as general guideposts for action. Firstly, regardless of how difficult the task, assessment and evaluation of mission impact is no longer optional but an essential requirement. Secondly, there is no right approach to measurement but rather better ways of thinking. Thirdly, pre-assumptions built into programming and assessments or evaluations need to be made explicit, challenged and tested. Fourthly, in this incredibly complex landscape of multiple projects, programs, stakeholders, inputs and outputs in very diverse contexts, ultimately, good enough outcomes, achievements and impact may indeed be good enough, given the overarching intentions of peace operations. And most importantly, stepping back from the debates regarding success or failure, what worked or did not work, ultimately we should be asking, did our efforts make a difference?

More broadly, assessing effectiveness and evaluating impact in peace operations will continue to be challenging. The challenges come from multiple sources: the inherent complexities of peace operations; the significant role that context plays in these missions; the political agendas that drive both the missions as well as those who support and contribute to mission activities; methodological hurdles; and the reconciliation of multiple stakeholder requirements.

Yet amongst this complexity, peace operations, in whatever form, are a reality for the foreseeable future. All things being equal, it might well be that measuring what matters is ultimately a decision that is context specific, dependent on responses to the following questions:

- What do you want to know?
- To what end/for what purpose?
- For whom?

In this light, the value is to assist decision-makers and the community of practitioners to make better informed decisions about current and future programs; to provide tools and measures that can help gauge 'what works' in a current operation; from what works, extrapolate what could be useful in future operations; and provide a guidepost to good practice.

RECENT CHALLENGES FORUM OCCASIONAL PAPERS

2014:1 ***Challenges with Assessing Impact in International Police Reform and Assistance***, by Mark Andrew Reber (March 2014).

2014:2 ***Evaluating Integrated Peace Operations***, by Jeni Whalan (April 2014).

2014:3 ***Capacity-building and Development of Host State Police***: ***The role of International Police,*** by Marina Caparini (May 2014).

2014:4 ***Police in UN Peace Operations: Evolving Roles and Requirements,*** by William Durch (August 2014).

2015:5 ***Principles of International Police Command***, by Maureen Brown, William Durch and Henrik Stiernblad (May 2015).

For a complete list of our publications please go to www.challengesforum.org /en/Publications

Impact Evaluation and Assessment of UN Peace Operations: What is the State-of-the-Art?

This paper provides an overview of the state-of-the-art of assessing and evaluating the impact of peace operations. It is structured to provide a review of current thinking and trends on impact assessment and evaluation of these operations focusing attention on the critical questions of what should we be measuring and at what level; it includes a substantive discussion on the technical aspects of defining terms common to the discourse, but not commonly accepted in practice. The paper continues with a section attentive to the methodology for carrying out measuring impact and includes illustrative tables of methodologies and approaches currently used by policy-makers and practitioners to measure the impact of peace operations. The paper concludes with a discussion of lessons captured and good practices, offering a number of guidance points for assessing and evaluating the impact of peace operations.

MICHELE LIPNER

Dr Michele Lipner has over 20 years experience in the international aid sector, providing senior level management to UN and NGO programs, focused on disaster, humanitarian and development assistance and peacebuilding. More recently, she has also held positions and undertaken consultancies focusing on doctrine development and promotion of good practices in civil-military-police coordination in disaster response and conflict management, evaluation and assessment of peace operations, security sector reform and strengthening NGO effectiveness through sector professionalisation.

ANN LIVINGSTONE

Dr Ann Livingstone served for eleven years as the Vice-President for the Pearson Peacekeeping Centre, after a successful academic career that spanned 27 years focused on conflict analysis and post-conflict transitions. She is currently a Senior Consultant with NGOs, regional organisations and institutes focusing on doctrine development in rule of law and impact assessment and evaluation, as well facilitating needs assessments focused on strategic planning and building organisational infrastructure.

ISBN 978-91-87927-08-9