INTERNATIONAL FORUM FOR THE
CHALLENGES OF PEACE OPERATIONS

Evaluating Integrated Peace Operations

JENI WHALAN

Occasional
Papers
No. 2

APRIL 2014

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 aim is to shape the debate by promoting awareness of emerging issues, identifying key challenges facing military, police and civilian peace operations, generating practical recommendations and supporting their effective implementation. It is a global network of partner organizations from all continents of the world.

Evaluating Integrated Peace Operations

JENI WHALAN

Occasional Papers
No. 2
APRIL 2014

Copyright © 2014 International Forum for the Challenges of Peace Operations.

All rights reserved. This work may be reproduced, provided that no commercial use is made of it and the Challenges Forum is acknowledged.

ISBN 978-91-87927-00-3

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

Suggested citation:
Jeni Whalan, Evaluating Integrated Peace Operations, Challenges Forum Occasional Papers 2014:2 (April 2014).

Editors:
Ann Livingstone, Michele Lipner and Sharon Wiharta.

Printed in Sweden 2014 by Taberg Media Group AB, Laholm.

The views expressed in this paper are those of the author, and do not necessarily reflect the views, opinion or position of the Challenges Forum or its partner organizations.

Contents

Preface iii

Executive Summary vii

Abbreviations ix

1. Introduction 1

2. Understanding Integrated Operations 3

A Politicized Context 4

The Paucity of Integrated Evaluation 7

3. The Inherent Constraints on Evaluating Integrated Peace Operations 13

Conflicting Objectives and Contested Priorities 13

Attributing Responsibility and Blame Avoidance 16

Methodological Problems 24

4. Comparative Approaches to Evaluating Integrated Missions in the DRC 33

Joint Evaluation of Conflict Prevention and Peacebuildning, 2011 33

UN OIOS Programme Evalution of Performance and Achievement of Result: 2012 35

Autesserre: The Trouble with the Congo, 2010 36

5. Conclusions 39

An Agenda for Better Evaluation of Integrated Missions 39

A Useful Menu of Evaluative Methods 41

Preface

Assessment and evaluation of peace operations is increasingly recognized by policy makers and practitioners as an imperative in order to enhance current and future peace operations. At the same time the demands for better evaluation of integrated missions raise methodological problems.

As peace operations became a central component of the international toolbox for addressing conflict, it has become increasingly evident that we need to better understand how to measure success as a means for informing future decision-making. That has opened a dialogue for asking the hard questions of impact assessment and evaluation: what are we measuring; for whom are we measuring; and for what purpose.

This study by Dr Jeni Whalan explores the state of the field in relation to practices utilized, or those holding potential, for assessing impact of a mission through a more integrated and comprehensive lens. The paper focuses on three themes: first it identifies inherent constraints to evaluating integrated missions, thereby informing strategic thinking about the politics of evaluating such efforts. Second, a set of principles for improving assessment of such missions are developed trough review of emerging evaluation practices and finally the paper presents examples of good evaluative techniques to inform future policy development in this very complex area of enhancing effectiveness and impact of peace operations.

This paper is part of a larger Challenges Forum work strand in 2012-2014 focused on Impact Evaluation and Assessment in peace operations, co-led by the Pearson Centre and the Institute for Security Studies (ISS). The aim of this working group is to contribute to the debate on how we can improve and make peace operations more effective by measuring the effectiveness and impact. A central part of the working group's undertaking has been a comparative inventory linking tools, toolkits and reviews use to measure the effectiveness and impact of peace operations. I am grateful to Dr Ann Livingstone (Pearson Centre) and Ms Annette Leijenaar (ISS) for so ably leading the working group and to Dr Michele Lipner (supported by the Australian Civil-Military Centre) for managing the project. I would also like

to thank Challenges Forum Partner organizations for their contributions to the work strand. The Challenges Forum also gratefully acknowledges the generous support of the Folke Bernadotte Academy, which helped to make possible the commissioning of the present study on Evaluating Integrated Peace Operations.

Annika Hilding Norberg
Director, Challenges Forum

April 2014

Executive Summary

This paper examines the intersection between two prominent trends in peace and security operations over the past 10–15 years: the progressive integration of security and development objectives, and the increasing demand for comprehensive evaluation of policy interventions. It makes two substantive contributions to improved practice. First, through conceptual progress on the framework for analysing integrated operations, it identifies a set of inherent constraints on their evaluation, informing strategic thinking about the politics of evaluating such efforts. Second, by reviewing emerging evaluation practices, it develops a set of principles for improving the assessment of integrated missions, and presents examples of good evaluative practice to inform future policy development.

The starting point is an observation that evaluation is always conducted for some purpose, for some audience and with some intended effect. To review the state of the art in evaluative practice, it is essential to first understand the politics of evaluation, that is, the interests and incentives that underlie the different purposes for which evaluation is conducted. This political context is evident in both the accountability and the learning imperatives of evaluation, compounded in the case of integrated missions by the fundamentally political nature of their operations in conflict environments. No single method of evaluation can overcome these challenges. The paper finds that a much wider array of methods can be employed than are currently relied on in the policy community, and contributes to improved practice by compiling a set of innovative cases that, in some instances, look beyond the current agenda to identify a broader range of evaluation approaches. Case studies are used throughout to illustrate and extend the analysis, selected to focus on two of the more difficult contemporary conflict environments for integrated missions: Afghanistan and the Democratic Republic of the Congo (DRC).

An agenda for better evaluation of integrated missions

Based on its review of evaluative practices, the paper sketches a set of principles for improving the evaluation of integrated missions. While further analytical work is needed, these guiding principles are intended to bolster such efforts.

The first principle is that evaluative standards of success and failure for a particular mission should be drawn from the specific conflict context. Second, evaluation should strive for independence, balance methodological rigour with pragmatism and be understood as part of an ongoing debate rather than a final adjudication on an operation. Third, it should be acknowledged that comprehensive evaluation may involve disaggregating integrated missions into their component parts, but that this should also include evaluation of the kinds of cross-cutting goals that integration is intended to serve. Finally, assessing unintended consequences is essential to evaluating integrated missions, and should include the potential for negative impacts to result from integration itself.

A menu of evaluative practices

Finally, the paper highlights a number of emerging 'good practices' that are not widely used in the evaluation of integrated missions, but should be more often considered:

1. Independent local analysts.

2. Institutional auditors and investigations units.

3. Issue-specific quasi-experimental impact evaluation.

4. Stakeholder evaluations and public perceptions.

Abbreviations

DFS Department of Field Services

DPKO Department of Peacekeeping Operations

DRC Democratic Republic of the Congo

DSRSG/RC/HC Deputy Special Representative of the Secretary-General/ Resident Coordinator/ Humanitarian Coordinator

FARDC Armed Forces of the Democratic Republic of Congo (Forces Armées de la République Démocratique du Congo)

DAC OECD Development Assistance Committee

HC Humanitarian Coordinator

IMPP Integrated Missions Planning Process

M&E Monitoring and Evaluation

MINUSTAH United Nations Stabilization Mission in Haiti

MPICE Measuring Progress in Conflict Environments

MONUC United Nations Organization Mission in the Democratic Republic of the Congo

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

NGO Non-governmental Organization

OECD Organization for Economic Co-operation and Development

UN United Nations

OIOS UN Office of Internal Oversight Services

PKSOI US Army Peacekeeping and Stability Operations Institute

RAMSI Regional Assistance Mission to Solomon Islands

RC Resident Coordinator

SIGAR US Office of the Special Inspector General for Afghanistan Reconstruction

SRSG Special Representative of the Secretary-General

TLO The Liaison Office

UNAMID African Union-United Nations Hybrid Operation in Darfur

UNCT United Nations Country Team

UNHCR United Nations High Commissioner for Refugees

UNIFIL United Nations Interim Force in Lebanon

UNMIL United Nations Mission in Liberia

UNMIS United Nations Mission in the Sudan

UNOCI United Nations Operation in Côte d'Ivoire

USACE US Army Corps of Engineers

1. Introduction

This paper examines the intersection between two prominent trends in peace and security operations over the past 10–15 years: the progressive integration of security and development objectives, and the increasing demand for comprehensive evaluation of policy interventions and implementation. It focuses on 'integrated' operations: those missions that adopt a comprehensive approach to stabilizing and/or building peace in fragile and conflict-affected countries, incorporating military, civilian and police actors in the pursuit of complex security and development goals.

Current demands for better evaluation of integrated missions raise complex methodological problems. More difficult to navigate, however, are the highly politicized agendas in which both integration and evaluation are embedded. Military, police and development actors have in the past decade recognized the new complexity of the operating environments to which integrated missions are deployed. In the midst of these complex environments, military, civilian and police components have been asked to forge new partnerships with one another, coordinating their activities and aligning vastly different cultures and operational strategies, in many cases in the field and on the hoof. At each level, this complexity introduces more variables that make rigorous evaluation more difficult.

This paper makes two substantive contributions to improved practice. First, through conceptual progress on the framework for analysing integrated operations, it aims to identify a set of inherent constraints on their evaluation and thereby shape strategic thinking about the politics of evaluating such efforts. Second, by reviewing emerging evaluation practices, it develops a set of principles for improving the assessment of integrated missions, and presents a number of examples of good evaluative practices to inform future policy development in this area.

2. Understanding Integrated Operations

In a United Nations context, an integrated mission refers to 'a strategic partnership between a multidimensional United Nations peacekeeping operation and the UNCT, under the leadership of the SRSG and the DSRSG/RC/HC'.[1] This is known as the one UN' approach. The UN has pursued the integration agenda in its field operations since 1997, seeking to increase the efficiency and effectiveness of the UN system in fragile states and conflict-affected environments through common strategic and operational means.[2] This reform agenda has included promoting unified leadership of all country activities by the Special Representative of the Secretary-General,[3] pursuing interagency programming and compacts between the UN system and the national government,[4] and reforming headquarters practice through such means as Integrated Mission Task Forces and Integrated Mission Planning Processes.[5]

Beyond the UN, peace operations have been influenced by similar dynamics within regional organizations and national governments. Their embrace of civil-military cooperation and whole-of-government approaches has increased the significance of integrated missions around the world.[6] This paper focuses on the UN context, but draws where relevant on insights about integrated missions in other contexts. For the purposes of this paper, it is assumed that

[1] United Nations (UN), Department of Peacekeeping Operations (DPKO)/Department of Field Support (DFS), *United Nations Peacekeeping Operations Principles and Guidelines* (New York, 2008), p. 69.
[2] See UN, *Renewing the United Nations: A Programme for Reform*, Report of the Secretary-General, A/15/950, 14 July 1997.
[3] See UN, *Note from the Secretary-General: Guidance on Integrated Missions*, 17 January 2006.
[4] Seen for example in Liberia and the Democratic Republic of the Congo (DRC).
[5] See UN, *Integrated Missions Planning Process (IMPP): Guidelines Endorsed by the Secretary-General*, 13 June 2006; UN, *Delivering as One: Report of the Secretary-General's High-Level Panel*, 9 November 2006.
[6] See Robert Egnell, *Complex Peace Operations and Civil-Military Relations: Winning the Peace* (Routledge; Abingdon; New York, 2009); Stewart Patrick and Kaysie Brown, *Greater than the Sum of its Parts? Assessing 'Whole of Government' Approaches to Fragile States* (New York: International Peace Academy, 2007); Michael C. Williams, 'Empire Lite Revisited: NATO, the Comprehensive Approach and State-building in Afghanistan', *International Peacekeeping* vol. 18, no. 1 (2011); Jonathan Goodhand, 'Contested Boundaries: NGOs and Civil-Military Relations in Afghanistan', *Central Asian Survey*, vol. 32, no. 3 (2013).

the differences in evaluating UN and non-UN integrated missions are not sufficiently significant to warrant separate treatment. This assumption may require further investigation in subsequent work.

A Politicized Context

Evaluation is always conducted to some purpose, for a specific audience, and to achieve some intended effect. In other words, evaluation is always a political endeavour. Understanding the different interests and incentives that may be involved in efforts to assess integrated peace operations is therefore an important starting point for this review of evaluative practices.

The past decade or so has witnessed the emergence and consolidation of a consensus among public policy communities that monitoring and evaluation should be central to policymaking and implementation. There are two underlying drivers for this: the imperatives of accountability and learning.

First, integrated missions exist amid particularly complex accountability relationships. One consequence of breaking down the traditional 'silos' through which peace operations acted in the past is that their lines of accountability became blurred. Notwithstanding their other problems, a substantial benefit of silos in policy implementation is their hierarchical nature, which clarifies the exercise of authority and attribution of responsibility, at least when compared to the horizontal accountabilities of integration. In a UN context, changing to a less hierarchical accountability model compounded the extant difficulties of holding peace operations to account, most notably as they have multiple principals, including the Security Council, the UN Secretariat, the broader UN membership and troop contributing countries, and they require substantial autonomy to operate in the field.[7] Calls for more and better evaluation of UN peace operations can be understood as part of a wider effort to enhance their accountability, particularly given the dramatic increase in the UN's peacekeeping budget since the turn of the century (see Figure 1).

Second, evaluation also serves to identify the 'lessons' of policy implementation: to determine what worked, what did not, why and what

[7] See for example Darren G. Hawkins et al. (eds.), *Delegation and Agency in International Organizations* (Cambridge University Press: Cambridge, 2006).

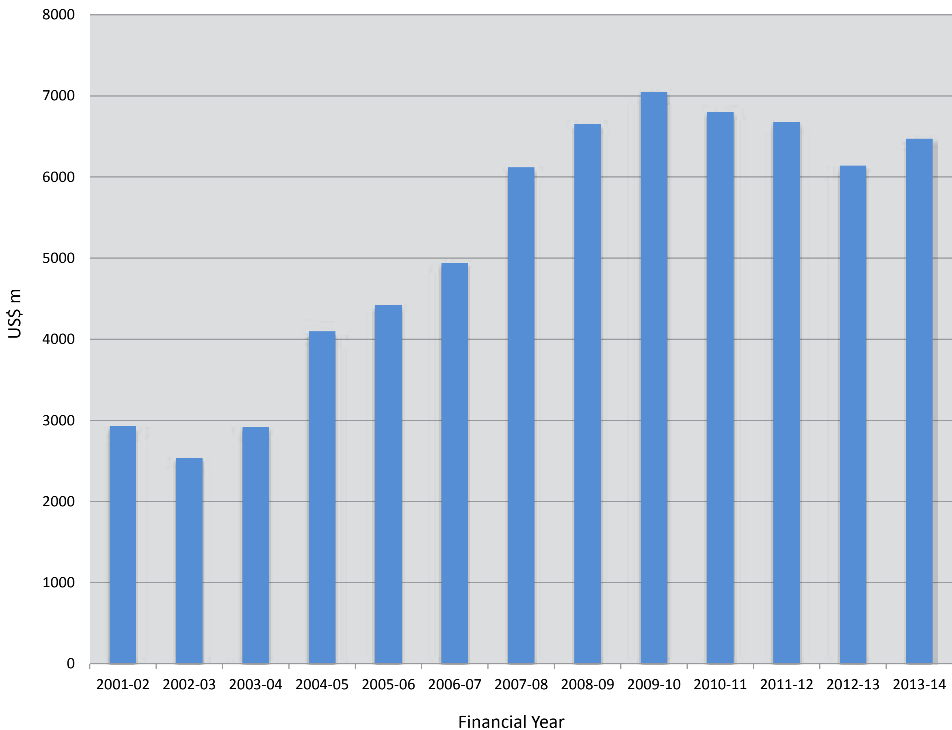

Fig 1. Evolution of Approved Peacekeeping Resources, 2001/02–2013/14. Budgets are set in US$ m at constant (2005) prices. *Source: UN documents A/58/705, A/60/696, A/61/786, A/C.5/61/18, A/C.5/62/23, A/C.5/63/23, A/C.5/64/15, A/C.5/65/15, A/C.5/66/14, A/C.5/68/21.*

reforms should be enacted accordingly. Given the frequently disappointing record of UN peace operations in achieving those cross-cutting tasks for which integration is deemed necessary, evaluation is encouraged as a means to improve policy, practice and outcomes. Of course, evaluations often combine learning and accountability.[8] Nonetheless, as section 2 discusses, learning and accountability may not always be mutually reinforcing.

Given that evaluation always has a particular purpose, audience and effect, we should not be surprised that evaluative practices are subject to considerable contestation. Nonetheless, there is a tendency among policy communities to assume that evaluation is a technocratic and procedural

[8] For example, the UK Department for International Development's (DFID) country programme evaluations are designed both to 'build accountability for the aid funds used and the results achieved in the countries selected, and generate learning that can inform future strategy and improve DFID's overall aid effectiveness'; see DFID, *Synthesis of Country Programme Evaluations Conducted in Fragile States,* Evaluation Report EV709 (February 2010), p. 1.

exercise. Instead, recognizing their multiple possible purposes, audiences and intended effects can help to explain why the conclusions and recommendations of evaluations can be contested, even if they have been conducted with appropriate rigour.

Of course, the political context of evaluation can also distort findings. In their leading book on the evaluation of peace operations, Paul Diehl and Daniel Druckman highlight the ease with which political concerns can derail evaluation:

> Foremost among the validity concerns is the temptation to interpret the information in a favorable way. Preferences for a positive spin serve to bolster the perceived value of the peace operation. Of course, it could also be the case that negative interpretations support a desire to exit from difficult missions.[9]

Recognition of these political effects of negative evaluation led Séverine Autesserre to warn that her study of peacebuilding operations in the DRC should not be read simply as a criticism of MONUC. First, she argued, her study is broader than the UN peacekeeping mission. It includes:

> other UN actors, as well as diplomats from various countries and international organizations, and many nongovernmental agencies' staff members. Reducing the analysis to a mere criticism of MONUC would thus miss one of the book's central arguments—the fact that the peacebuilding culture, as well as the understandings and actions it shapes, are spread across a variety of interveners.[10]

Second, and perhaps more importantly, Autesserre acknowledges explicitly the real-world implications of evaluation, warning potential audiences not simply to attach resource decisions to discussion of peacebuilding's failures:

> [At the time of] writing, MONUC's presence is one of the main reasons why the Congo has not (perhaps yet) slid back into a full-scale national and regional war. The Congolese population would suffer tremendously more if it did not benefit from the peacebuilding, development, and humanitarian aid delivered by various international actors. The policy implications of this book are therefore not that donors should stop financing aid programs in the Congo and in other conflict situations because the international intervention during the Congolese transition was a resounding failure. Rather, the goal of this book is to help policy makers further boost the positive aspects of international

[9] Paul F. Diehl and Daniel Druckman, *Evaluating Peace Operations* (Lynne Rienner: Boulder, CO, 2010), p. 81.
[10] Séverine Autesserre, *The Trouble with the Congo: Local Violence and the Failure of International Peacebuilding* (Cambridge University Press: Cambridge, 2010), p. 13. This observation also holds true in other missions.

> peacebuilding interventions, in particular by including bottom-up conflict-resolution programs in their initiatives.[11]

The problems with overlooking the political foundations of policy evaluation are exacerbated in the context of integrated peace operations, which are themselves highly politicized policy instruments.[12] Understanding the politics of evaluation can help to illuminate the challenges of performance evaluation and the attribution problem, discussed further below.

The Paucity of Integrated Evaluation

There is surprisingly widespread acknowledgement of a crisis in the evaluation of integrated efforts. For example, the UN's 2010 practitioner guide *Monitoring Peace Consolidation* found, 'very few examples of benchmarks focusing on system-wide effects, and virtually no examples of benchmarks on system-wide effects not directly related to particular mandated United Nations objectives and targets, e.g. benchmarks focused purely on contextual aspects of progress toward sustainable peace in a country'.[13]

Similarly, the Organization for Economic Co-operation and Development Development Assistance Committee's (OECD DAC) 2012 guidelines for evaluating peacebuilding identify a 'persistent evaluation gap' in peacebuilding and conflict prevention activities by its members, and 'little to no evaluation activity in settings of violent conflict'.[14] This evaluation gap distinguishes integrated efforts from those of their component parts. This is true whether it is the integration of actors (i.e. civil-military-police relations) or the integration of objectives (i.e. policy goals that demand comprehensive approaches, such as the rule of law, security sector reform, or broader statebuilding) that are being analysed. Nonetheless, attempts are being made to close this evaluation gap. Two examples are examined below: the Measuring Progress in Conflict Environments (MPICE) initiative, and Diehl and Druckman's Evaluating Peace Operations.

[11] Autessere, 2010, pp. 13–14.
[12] This is widely acknowledged as a defining feature of integrated efforts. See for example Organization for Economic Co-operation and Development (OCED), *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility: Improving Learning for Results*, DAC Guidelines and Reference Series (OECD Publishing: Paris, 2012).
[13] UN, *Monitoring Peace Consolidation: UN Practitioners' Guide to Benchmarking* (New York, 2010).
[14] OECD, 2012.

Evaluative framework 1: Measuring progress in conflict environments

The MPICE framework is a comprehensive metrics-based approach to measuring security and stability changes in the conflict environment of an integrated stabilization mission. Developed collaboratively by the United States Institute of Peace, the US Army Corps of Engineers (USACE), and the US Army Peacekeeping and Stability Operations Institute (PKSOI), MPICE is a highly structured framework designed for application in any environment. MPICE serves to highlight the incredible complexity involved in measuring every variable that might be relevant to the measurement of security and stability in conflict environments. It is designed to provide:

> a comprehensive capability for measuring progress during stabilization and reconstruction operations for subsequent integrated interagency and intergovernmental use. MPICE enables policymakers to establish a baseline before intervention and track progress toward stability and, ultimately, self-sustaining peace. The intention is to contribute to establishing realistic goals, focusing government efforts strategically, integrated inter-agency activities, and enhancing the prospects for attaining an enduring peace.[15]

The MPICE framework collects an impressive array of metrics for evaluating integrated missions. The set of MPICE tools includes a 'web-enabled tailoring wizard', which enables the method to be tailored to a specific conflict by selecting among more than 600 built-in measures.[16] Like any structured framework, this requires context-specific expertise to select between its metrics menu in order to design an appropriate evaluation framework, and to interpret the data collected. As an application of MPICE to stabilization evaluation in Haiti concluded:

> Measuring progress in a conflict environment is always a challenge, and even with a serious effort using sophisticated M&E methods, analytical techniques, and tools, including the MPICE framework, our program produced almost as many questions as it answered. We improved our efforts over each phase, and presumably, if we had had more than three collection phases…we would have had far more data to analyze and use for planning.[17]

[15] John Agoglia, Michael Dziedzic and Barbara Sotirin (eds.), *Measuring Progress in Conflict Environments (MPICE): A Metrics Framework* (United States Institute of Peace Press: Washington DC, 2010), p. xii.
[16] David C. Becker and Robert Grossman-Vermaas, 'Metrics for the Haiti Stabilization Initiative', *Prism,* vol. 2, no. 2 (2011), p. 149.
[17] Becker and Grossman-Vermaas, 2011, p. 158.

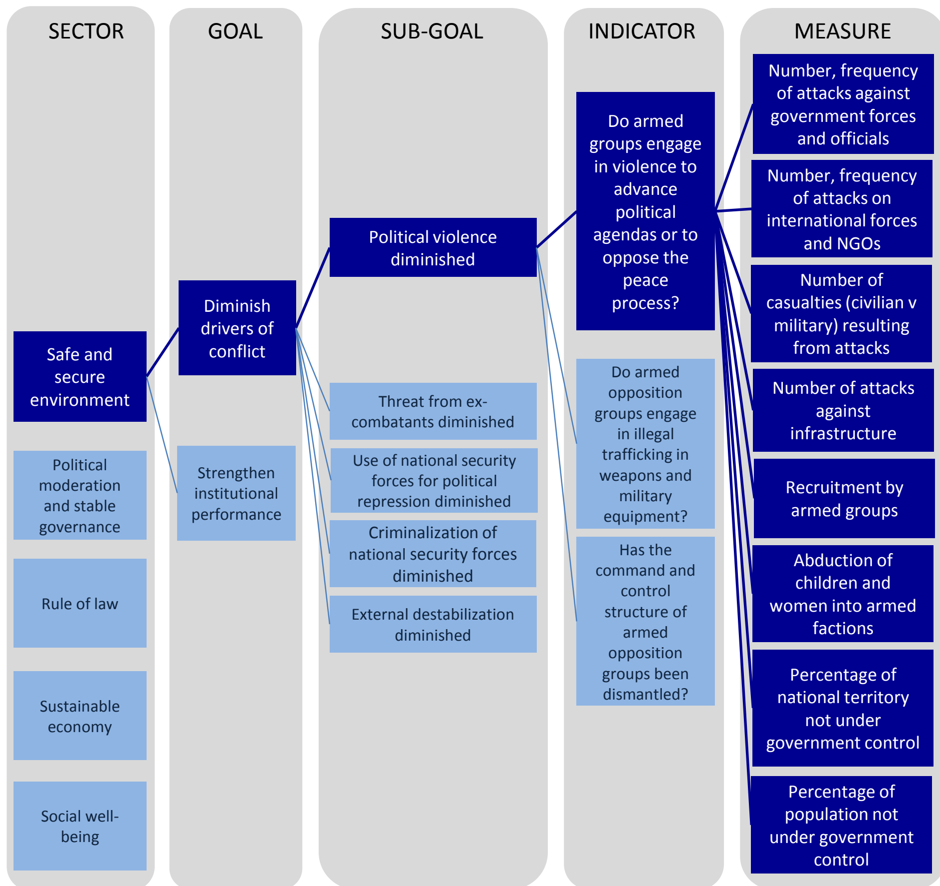

Fig. 2. The complexity of Measuring Progress in Conflict Environments (MPICE).

By aiming for broad applicability to any conflict setting, the MPICE framework requires considerable complexity to cover all possible indicators and metrics relevant to a specific case. For example, Figure 2 depicts the complexity of measuring just one indicator. The entire MPICE framework has 149 indicators.

This structured complexity is both a key strength and notable weakness of MPICE: it requires evaluators to have such a depth of expertise in a particular conflict setting that they should be able to develop an appropriate evaluation framework suited to the conditions without the formal methodology of MPICE metrics. The real strength of a structured framework is to enable comparison across cases, that is, a methodology of structured, focused comparison, in order to produce rigorous causal inferences.

Evaluative framework 2: Diehl and Druckman evaluating peace operations

For application to integrated peace operations (rather than the military stabilizations conducted in Iraq and Afghanistan, for which MPICE seems particularly relevant), the evaluation framework developed by Diehl and Druckman may be more suitable.[18] In *Evaluating Peace Operations*, Diehl and Druckman provide a 'decision-making template for assessing peace operation success that includes different goals or objectives on which operations may be judged, key questions to ask about the achievement of those goals and objectives, and operational indicators that may be used as evidence in answering those questions'.[19] It provides clear standards for assessing the effectiveness of peace operations, but retains more useful nuance than the MPICE framework. The approach bridges theory and policy, balancing the contribution of generalizable theorizing to better peace operations practice with the need for context-specific evaluation. As Diehl and Druckman note:

> Many studies of peace operation effectiveness have been based on single cases, creating problems of generalizability for any conclusions. More importantly… the standards for success were highly specific to the context and operation at hand. Such evaluations are less useful as lessons for future missions. Policy analyses of lessons learned are predicated on applying conclusions from one context to another. Case-specific standards or indicators inhibit the ability of policymakers to take what they learned from one operation and adapt policies to a different context. From a scholarly standpoint, researchers must be able to construct some common standards and indicators of success in order to compare performance across missions and draw generalizations. Case-specific benchmarks inhibit the empirical verification of theory-derived propositions about peace operations and thereby stifle the development of general knowledge and patterns.[20]

The framework evaluates peace operations against a set of core peacekeeping goals (conflict containment, conflict settlement and violence abatement), 'non-traditional' or broader peacekeeping goals (humanitarian assistance, human rights protection, disarmament, demobilization and reintegration, election supervision and democratization) and peacebuilding goals (local security, the rule of law, local governance and restoration/reconciliation). Evaluation proceeds according to a decision-making template that

[18] Diehl and Druckman, 2010.
[19] Diehl and Druckman, 2010, p. 2.
[20] Paul F. Diehl and Daniel Druckman, 'Peace Operation Success: The Evaluation Framework', *Journal of International Peacekeeping*, vol. 16, no. 3–4 (2012), p. 211.

operationalizes these goals into key questions for the evaluator to ask of a particular operation, and ways to measure progress on each.

Importantly, Diehl and Druckman also specify the benefits and limitations of particular measures of progress, making clear that evaluation can never be simply a technical exercise, but is one that demands an evaluator who approaches the task with continuous critical reflection: 'evaluators must weigh difficulties in information collection and measurement validity in deciding on evaluation instruments and making judgments'.[21]

The authors emphasize that 'binary judgments on peace operations are likely to be misleading', since they will usually be more successful on some aspects of a particular goal than on others, and that the multidimensional character of peace operation goals means that multiple indicators will be required to capture their outcomes comprehensively, to validate the assessment through triangulation and to accommodate the reality that data gaps will require the substitution of proxy indicators. Furthermore, 'indicators can include both quantitative data and qualitative information, and come from many sources including extant data collections, public sources, information provided by the sponsoring agency of the peace operation, and those that might be collected by the peace operation'.[22]

The framework's evaluative utility was demonstrated in 2012 by a collection of comparative evaluations of integrated peace and security missions in Cambodia, Côte d'Ivoire, Timor-Leste and Liberia.[23] The Bosnia case evaluated by Diehl and Druckman also demonstrates the value of the framework's approach.[24] More importantly, the comparative exercise highlighted the importance of conceptualizing evaluation as an ongoing, iterative and often contested exercise, one that is engaged in a continual search for the ideal of truth—not a technical process that can be undertaken in a mechanistic, uniform way.

[21] Diehl and Druckman, 2012, p. 20.
[22] Diehl and Druckman, 2012, p. 20.
[23] *Journal of International Peacekeeping*, vol. 16, no. 3–4 (2012)
[24] Diehl and Druckman, 2010, pp.174–202.

3. The Inherent Constraints on Evaluating Integrated Peace Operations

Substantial progress has been made in recent years on evaluative practices for assessing integrated missions. Nonetheless, there are some inherent difficulties in evaluating integrated missions that cannot be resolved through policy innovation alone. This section aims to provoke deeper thinking than is generally found in the otherwise rich body of policy literature on evaluation, and thereby speak to the strategic environment in which peace and security operations are embedded.

Conflicting Objectives and Contested Priorities

The United Nations has affirmed that integration is 'the guiding principle for all conflict and post-conflict situations where the UN has a Country Team and a multidimensional peacekeeping operation or political mission/office, whether or not these presences are structurally integrated'.[25] Intended principally 'to maximize the individual and collective impact of the UN response, concentrating on those activities required to consolidate peace', this integration imperative underlies much of the UN's recent reforms regarding peace operations.

Improving evaluation, however, demands a critical detachment from the normative agenda of integration. Evaluations of integrated missions should retain a degree of critical ambivalence about the value of integration if its effects are to be rigorously analysed. Evaluators must allow for the possibility that integration may not always be desirable. In some conditions, merging security, development, governance and humanitarian efforts results in internal contradictions between an operation's objectives. As Susanne P.

[25] UN, *Decision No. 2008/24: Integration*, Decisions of the Secretary-General, 26 June 2008.

Campbell and Anja T. Kaspersen note, UN guidance on peace operations largely ignores the potential for such contractions:

> Instead, it is largely assumed that the majority of the UN's activities implemented in or alongside multidimensional peace operations are compatible…integration reforms have to a large extent ignored other potential contradictions, such as those between stabilization and governance, or liberalization and institution building.[26]

Given the extent to which the integration imperative has also influenced the policy and practice of states and regional organizations, looking beyond UN operations is instructive. Afghanistan provides an exemplary case of a mission with deeply embedded tensions, including between counterterrorism and development, democracy and stability, and short- and long-term goals. In her book *When More is Less*, Astri Suhrke identifies a set of contradictions in what she calls the 'international project' in Afghanistan:

> the clash between the aim to create a locally owned liberal order and the heavy, intrusive means by which international assistance has attempted to enact it; that the rentier state created by extensive external assistance inhibits the aspirations for democratization; the rank contradiction of waging war and building peace simultaneously, clearly manifest in the civilian casualties which alienated rather than won 'hearts and minds'; and the imperative to achieve 'results' in the short-term, such as cooperating with a local strongman who could deliver stability, which nevertheless undermined longer-term objectives, such as the attainment of good local governance as a crucial link in building the legitimacy of the Afghan state.[27]

Roland Paris and Timothy D. Sisk come to a similar conclusion in their edited volume on the dilemmas of statebuilding:

> the larger challenge is not simply to 'do more'. The more crucial, but perplexing, challenge is to manage the difficult and deep contradictions of post-conflict peacebuilding, particularly those involved in efforts to strengthen or construct effective and legitimate state institutions as a foundation for security, human development, and other public goods within societies emerging from war.[28]

---

[26] Susanne P. Campbell and Anja T. Kaspersen, 'The UN's Reforms: Confronting Integration Barriers' *International Peacekeeping*, vol. 15, no. 4 (2008), p. 477.

[27] See Astri Suhrke, *When More is Less: The International Project in Afghanistan* (Columbia University Press: New York, 2011), pp. 15–18.

[28] Roland Paris and Timothy D. Sisk, 'Conclusion', in Paris and Sisk (eds.), *The Dilemmas of Statebuilding: Confronting the Contradictions of Postwar Peace Operations* (Routledge: London, 2009), p. 304.

Paris and Sisk identify five fundamental contradictions that underlie contemporary peace and security operations, each of which poses distinct challenges to integrated missions: (1) outside intervention is used to foster self-government; (2) international control is required to establish local ownership; (3) universal values are promoted as a remedy for local problems; (4) statebuilding requires both a clean break with the past and a reaffirmation of history; and (5) short-term imperatives often conflict with longer-term objectives.

In addition to the problems these tensions pose for policymakers, evaluations of peace operations must frequently grapple with contradictions in mission mandates and contested benchmarks criteria for success. Addressing such contradictions lends itself to nuanced analysis rather than binary evaluations of success and failure, or aggregated quantitative data. For example, to assess the effectiveness of security sector reform efforts, the impacts must be observed from the micro to the macro levels, for their short- and long-term effects, and for the distribution of costs and benefits. Security sector reform efforts in the DRC provide a useful illustration. At times, civilians have been made less secure by tactics that aim to achieve the larger peacebuilding goal of civilian protection. To use just one example, in an attempt to compel foreign forces to leave the eastern DRC, MONUC drew on the coercive capacity of the national military (the FARDC), which had the unintended consequence of prompting the displacement of civilians who fled in fear of FARDC brutality.[29]

The use of conflict sensitivity principles may help to address the need to evaluate integrated missions with contradictory goals, particularly the trade-offs and unintended consequences of interventions. Initially developed within the humanitarian aid sector,[30] the conflict sensitivity approach has gained substantial traction as a means of assessing and designing peacebuilding interventions.[31] Conflict sensitivity refers to 'a set of processes that help us recognize the unintended ways our work can contribute to

[29] Eirin Mobekk, 'Security Sector Reform and the UN Mission in the Democratic Republic of Congo: Protecting Civilians in the East' *International Peacekeeping,* vol. 16, no. 2 (2009), pp. 275–6.
[30] Mary B. Anderson, *Do No Harm: How Aid Can Support Peace—or War* (Lynne Rienner: Boulder, CO, 1999).
[31] See Sarah Brown et al., *Conflict Sensitivity Consortium Benchmarking Paper,* Conflict Sensitivity Consortium (2009); OECD, *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility: Improving Learning for Results*, DAV Guidelines and References Series (OECD Publishing: Paris, 2012); Rachel Goldwyn and Diana Chigas, *Monitoring and Evaluating Conflict Sensitivity: Methodological Challenges and Practical Solutions*, DFID Practice Product, (March 2013).

conflict',[32] which enables systematic review of 'both the positive and negative impacts of interventions, in terms of conflict or peace dynamics, on the contexts in which they are undertaken, and, conversely, the implications of these contexts for the design and implementation of interventions'.[33]

Attributing Responsibility and Blame Avoidance

Since evaluation is always political, a common obstacle to the rigorous assessment of a policy intervention is the desire by the agents involved to claim credit for successes and deflect blame for failures. This is considerably exacerbated in the case of integrated missions, where different components of the operation have distinct institutional identities, each with a firm stake in protecting their reputation—and often their supply of resources.

Interestingly, this problem appears to intensify when the components are drawn from the same institution, such as departments within a national government. Notably, the United Nations Secretariat has on a number of occasions issued quite dramatic *mea culpa* evaluations of its previous record,[34] a result perhaps of longer accountability chains and the UN Secretariat's reliance on, and latitude for, moral leadership.

The bureaucratic battles over responsibility, particularly the blame-shirking variant, contribute substantially to the reliance on procedural measures of outputs in integrated evaluations, such as the number of police trained by a security sector reform programme, rather than substantive assessment of outcomes, such as the competence of those police or, more significantly, the contribution of security sector reform efforts to peace consolidation. The negative consequences of accepting responsibility for poor outcomes lead departments to demand that evaluators focus on work completed (performance) instead of outcomes achieved (impact). This often results from pragmatic battles for reputation and resources among mission components, not from a principled rejection of rigorous evaluation. Nevertheless, such

[32] Goldwyn and Chigas, 2013, p. 8.
[33] OECD, 2012, p. 11.
[34] See for example UN, *Report of the Independent Inquiry into the actions of the United Nations during the 1994 genocide in Rwanda,* S/1999/1257, 16 December 1999; UN, *Report of the Secretary-General pursuant to General Assembly resolution 53/35: The fall of Srebrenica,* A/54/549, 15 November 1999.

organizational disputes can have dramatic implications for the quality of evaluation.

The relationship of performance evaluation to impact evaluation is a vexed one. Evaluating how well a peace operation has performed its duties and implemented its mandate is a necessary part of policy implementation and may be conducted, for example, as part of budgetary accountability. Performance evaluation is important because the conditions in conflict-affected states mean that it is possible—and indeed likely—that a peace operation can perform appropriately without measurable impact. Since peace operations are often responsible for the prevention of further deterioration in a conflict, assessing their performance by means of their impact on the conflict has evident flaws.[35] Nonetheless, such concerns can also obstruct the kinds of rigorous impact evaluation necessary for learning.

This underlines one of the central debates about peacekeeping policy in the UN: the plea by the Secretariat for peace operations to have clear, achievable mandates with the resources to match. After the succession of peacekeeping failures in the early to mid-1990s, the UN's Department of Peacekeeping Operations (and others arguing on its behalf) rightly protested the attribution of blame to it, rather than to member states and the conflict parties. If a peacekeeping operation is given an unrealistic mandate, is poorly resourced and is deployed to insecure environments populated by spoilers, so the argument goes, then the failure to resolve conflict cannot rightly be attributed to the poor performance of embattled peacekeepers. Instead, the responsibility should rest with the Security Council and the wider UN membership, and with the parties to the conflict.

The challenge for the evaluator is to insulate the discovery of outcomes and their causes from the politics of attribution. Two methods for doing so are worth exploring. The first is to embrace genuinely independent evaluation of integrated peace operations, that is, to design an evaluation in such a way that the political implications of its findings have minimal opportunity to influence the assessment itself. This runs counter to much of current practice, which sees donors, troop- and police-contributing countries, mandating organizations and mission personnel as 'stakeholders' that should

[35] The limited value of assuming that peace processes proceed in a sequenced fashion—prevention, peace-keeping, peacebuilding—is now well understood. As the Capstone Doctrine notes, 'Conflict prevention, peacemaking, peacekeeping and peace enforcement rarely occur in a linear or sequential way. Indeed, experience has shown that they should be seen as mutually reinforcing'.UN, *Principles and Guidelines,* 2008, p. 20.

be actively engaged in the design and implementation of an evaluation. Authorizing genuinely independent evaluation may also be unrealistic for those stakeholders that report to their own constituencies. This dilemma, true of any policy evaluation, is likely to be more intense for integrated missions, since their component parts may have competing interests as far as future resource allocation is concerned. The nature of this attribution dilemma can be imagined on a spectrum, extending from one end, where a mission component has legitimate reason to resist being unjustly blamed for the overall mission's ineffectiveness, to the other, where avoiding blame is an intentional act of bureaucratic politics. Insulating evaluation from these pressures holds substantial promise for enabling more rigorous evaluation capable of generating the kinds of causal insights necessary to promote learning.

Nonetheless, the value of independence makes it worthy of serious consideration at the strategic level. This would require an independent funding stream for evaluation to be built into an operation's design, and an independent evaluator to have the authority to access all relevant operational data. Examples include the World Bank's Independent Evaluation Group, and the US Office of the Special Inspector General for Afghanistan Reconstruction (SIGAR), discussed in Box 1. There is a need for further work on the models for and techniques of independent evaluation best suited to the demands of peace operations, at both the field and the headquarters level.

The second option for insulating evaluation from the perverse incentives of attribution is for the integrated mission (and its principals) to genuinely embrace the discovery of failures as opportunities to improve both learning and accountability. The notion that 'embracing failure' is central to innovation (i.e. policy improvement) has become fashionable in recent years in business and organizational thinking.[36] Notwithstanding the inevitable hype accompanying its rise, the concept of embracing failure has more to offer the evaluation of peace operations beyond its buzzword status, particularly regarding what it really takes to 'learn' from failure. Importantly, it is rare for an organization to really 'learn well' from its failures. The UN's efforts to encourage lesson learning in its integrated missions appear to fall into the trap common to many organizations: assuming that learning is

[36] See for example www.admittingfailure.com; Schumpeter, 'Fail often, fail well', *The Economist*, 14 April 2011.

Box 1: The US Office of the Special Inspector General for Afghanistan Reconstruction

The US Office of the Special Inspector General for Afghanistan Reconstruction (SIGAR) provides one model for independent evaluation within a national government setting. Created by the US Congress as an independent agency in 2008, SIGAR has three identified purposes:

1. To conduct or supervise independent and objective audits and investigations of US reconstruction operations in Afghanistan.
2. To provide independent and objective leadership and coordination of, and make recommendations on, the promotion of economy, efficiency and effectiveness in such operations, and prevent and detect waste, fraud and abuse.
3. To provide for an independent and objective means of informing the Secretaries of State and Defense about problems and deficiencies relating to the administration of such operations, and the need for and progress on corrective action.[1]

SIGAR's independent and objective oversight of US reconstruction activities in Afghanistan is conducted from its forward operations bases in Afghanistan and from its headquarters in Arlington, Virginia. Testament to its independence is the extent of SIGAR's reporting on the fraud and waste it has uncovered in US reconstruction activities.[2] SIGAR's most recent quarterly report noted that:

> This quarter alone, SIGAR investigations resulted in more that $63 million being frozen in bank accounts, two arrests, three sentencings, and more than $95,000 in fines and restitutions. To date, SIGAR investigations have led to 47 convictions and guilty pleas; more than $236 million in recoveries, savings, and contract monies protected; and 61 suspensions and 94 debarments of individuals and companies from receiving US-funded contracts. SIGAR has more than 300 ongoing criminal investigations in Afghanistan and the United States.[3]

---

[1] SIGAR Enabling Legislation, SEC. 1229 Special Inspector General for Afghanistan Reconstruction, *National Defense Authorization Act for Fiscal Year 2008*, Pub. L. No. 110-181 (28 January 2008).
[2] See www.sigar.mil for a comprehensive list of audits and investigations.
[3] SIGAR, *Quarterly Report to the United States Congress* (30 October 2013).

primarily about evaluation. Instead, 'the blame game' often obstructs both good evaluation and the necessary learning strategies that must follow.[37]

The following case study of evaluating a critical goal area for integrated missions—civilian protection—serves to highlight both the evaluative challenges posed by cross-cutting themes, and the ways in which contestation about attribution can play out.

Evaluating cross-cutting themes: Civilian protection

Cross-cutting themes present distinct challenges for the evaluation of integrated missions. Protecting civilians is the archetypal example in contemporary peace operations. It is precisely this goal type that integration aims to achieve: complex objectives that cannot be achieved by military, police or civilian actors alone. Yet this is what makes them so difficult to evaluate.

Isolating the reasons for the success or failure of shared mission goals such as civilian protection is more difficult when there is not just one policy intervention but many across the operation. For example, physical protection is pursued by peace operations through a number of military means, from the direct deterrence of violence against civilians through physical presence, to the indirect measures of security sector reform, which aim to enable the host country's armed forces to conduct physical deterrence, and, at times, coercive action against the armed groups responsible for civilian insecurity. In the DRC, the UN mission has used integrated, civil-military mechanisms to promote physical protection, including through Joint Protection Teams, Community Liaison Assistants and Community Alert Networks. The mission has also pursued a broader concept of protection by cooperating with humanitarian actors, including through the Protection Cluster, created in 2006 under UNHCR leadership and co-chaired by the UN mission.[38] Evaluating the success of integrated efforts to protect civilians thus requires isolating the outcomes and weighting of each type of strategy, as well as the myriad ways in which they interact—including the ways in which the

---

[37] See for example Amy C. Edmondson, 'Strategies for learning from failure', *Harvard Business Review*, vol. 89, no. 4 (2011), pp. 48–55.

[38] See J. Arthur Boutellis, 'From Crisis to Reform: Peacekeeping Strategies for the Protection of Civilians in the Democratic Republic of Congo', *Stability: International Journal of Security and Development*, vol. 2, no. 3 (2013).

different means of implementation may undermine one another. This is a very difficult task, made more so by the volatile security environments under observation.

To overcome this difficulty, the evaluator must make careful judgments about how to weight the effects of different actions. Good practice in such cases is to make those assumptions explicit; that is, to specify the theory of change from which those assumptions flow. The method of process-tracing is a valuable tool in such circumstances. Process-tracing 'attempts to trace the links between possible causes and observed outcomes', examining 'whether the causal process a theory hypothesizes or implies in a case is in fact evident in the sequences and values of the intervening variables in that case'.[39] The nature of cross-cutting goals means that many parts of an operation may be responsible for the success or failure of their achievement. The integrated character of both the mission and the goal makes it difficult to isolate the causal processes that explain outcomes. Good evaluation practice should include the possibility of 'equifinality', that is, the recognition that there are multiple pathways to a given outcome. A process-tracing method is particularly useful for locating and analysing evidence that can 'narrow the list of potential causes',[40] particularly in settings of 'multiple interaction effects, where it is difficult to explain outcomes in terms of two or three independent variables';[41] that is, the settings in which evaluation of integrated missions takes place.

Two further challenges await evaluation of cross-cutting goals such as civilian protection. First, progress on cross-cutting goals is often reported by many parts of the integrated mission, and often to different standards, for different audiences. Third, this difficulty is exacerbated by the politics of attribution, which can obstruct both the accountability and the learning objectives of evaluation. As J. Arthur Boutellis notes of civilian protection in the DRC:

> The latest in a series of cyclical crises in eastern DRC shows the limits of what has been a largely technical and UN-centric approach to physical protection, when the root causes of violence—whether at the local, national or regional levels—are often political. While peacekeepers may at times indulge in self-justification and preservation, they also fall victim to a system that too often

[39] Alexander L. George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences* (MIT Press: Cambridge, MA, 2005), p. 6.
[40] George and Bennett, 2005, p. 206.
[41] Peter Hall, in George and Bennett, 2005, p. 205–206.

> evaluates their performance based on the mission's own protection record and failures, rather than on whether it has enabled the host state to shoulder its primary responsibility to protect its own civilian population.[42]

These two challenges are evident in a 2013 UN Office of Internal Oversight Services (OIOS) review of civilian protection reporting by seven peacekeeping operations: the United Nations Interim Force in Lebanon (UNIFIL), the United Nations Organization Mission in the Democratic Republic of the Congo (MONUC)/United Nations Organization Stabilization Mission in the Democratic Republic of the Congo (MONUSCO), the United Nations Mission in Liberia (UNMIL), the United Nations Operation in Côte d'Ivoire (UNOCI), the United Nations Stabilization Mission in Haiti (MINUSTAH), the United Nations Mission in the Sudan (UNMIS) and the African Union-United Nations Hybrid Operation in Darfur (UNAMID).[43] The review did not assess the outcomes of civilian protection, but instead aimed 'to determine how missions with protection-of-civilians mandates reported on progress in the implementation of this task in their performance reports'.[44]

In performance evaluations, since nearly all components of a mission have some civilian protection responsibility, performance data tend to be dispersed throughout performance reports. As the OIOS notes, 'these activities are crosscutting in purpose but task-specific in nature. Consequently, dispersed reporting of civilian protection activities in a performance report may be inherently difficult to avoid'.[45] The 2013 OIOS review on the subject found that reporting on the protection of civilians was submitted under multiple frameworks, including 'military', 'peace and security', 'human dimension of sustainable peace', 'humanitarian assistance and human rights', 'civil society and human rights' and 'transitional process'.[46]

In addition to providing useful findings about the nature of protection reporting by peacekeeping operations, the review offers valuable insight into two common obstacles to the kind of evaluation necessary for evidence-based policymaking. First, it highlights the distinction between evaluations conducted for the purpose of ongoing strategic policymaking (in this case,

---

[42] Boutellis, 2013, p. 9.
[43] UN Office of Internal Oversight Services (OIOS), *Review of the Reporting by United Nations Peacekeeping Missions on the Protection of Civilians,* A/67/795, 15 March 2013.
[44] UN, OIOS, A/67/795, 2013, p. 6.
[45] UN, OIOS, A/67/795, 2013, p. 13.
[46] UN, OIOS, A/67/795, 2013, p. 13.

the Security Council's decisions about mandates), and those conducted for the purpose of accountability (in this case, budgetary performance review of the specific peacekeeping operation). A central finding of the OIOS review was that the number of civilian deaths reported by peacekeeping operations in their performance reports differed from that presented in the mission-specific reports of the UN Secretary-General for the same time period. Having identified this inconsistency, the OIOS sought an explanation from the UN's Department of Peacekeeping Operations/Department of Field Support (DPKO/DFS). The exchange is instructive.

OIOS found that the mission-specific reports of the Secretary-General regularly reported more civilian deaths than the mission's performance report.[47] A central recommendation of its review was that 'missions with civilian protection mandates include the consistent and quantified use of "conflict-related civilian deaths"...as indicators of achievement in their performance reports'.[48] In its comments on the review, DPKO/DFS emphasized its concern that this 'asserts causality where it may not exist'—specifically, that the number of civilian deaths 'may not be directly attributable to the actions of a peacekeeping mission, within the scope of its deployment', and as such does not reflect 'mission failure or success in implementing its protection-of-civilians mandate, nor are they the sole meaningful measurement of the impact of mission efforts in this regard'.[49] DPKO/DFS elaborated further on the issue of attributed responsibility, noting that protection responsibilities are born primarily by host states, and thus '[m]easuring protection-of-civilians performance on the basis of all conflict-related deaths and sexual violence/rapes that occur in a mission area also fails to acknowledge the host State's own successes or failures in protection'.[50] To the extent that DPKO/DFS raise a valid concern about evaluation, it appears to be an assumption about the interpretation of evaluations and the attribution of blame for failure or credit for success. DPKO/DFS appears anxious to avoid being blamed for civilian deaths, and hence failing to protect civilians, when that blame ought to be attributed to perpetrators and host states.

This suggests an example of bad evaluation practice. As OIOS rightly notes, the concern of DPKO/DFS 'is misplaced as the report explicitly states that

[47] UN, OIOS, A/67/795, 2013, p. 16.
[48] UN, OIOS, A/67/795, 2013, p. 21.
[49] UN, OIOS, A/67/795, 2013, p. 29
[50] UN, OIOS, A/67/795, 2013, p. 29

"an increase in civilian deaths does not necessarily mean a mission had been ineffective"'.[51] Indeed, any evaluation that measured mission performance on this basis would be flawed. This should not, however, prevent a rigorous, accurate assessment of the status of civilian protection. As OIOS notes, 'since both the mission-specific reports and the performance reports deal with protection-of-civilians issues, they must, despite their different purpose, be consistent, especially on the important issue of number of civilian deaths'.[52]

Methodological Problems

There are a number of methodological challenges associated with any international policy intervention, but the complexity of integrated efforts and the environments in which they occur make these substantially more difficult to resolve. The multidimensional nature of integrated missions means that they have a very large number of impact factors and observational variables, dispersed among military, civilian and police components. Furthermore, fragile and conflict-affected environments pose particular challenges for data collection. Integrated missions typically operate in countries with weak state capacity, which means that even the most basic population data are absent. If pre-conflict data were collected, conflict dynamics such as displacement make it unreliable. Furthermore, data collection during conflict is dangerous and time-consuming. These conditions combine to mean that the evaluation of integrated missions usually takes place without the kind of baseline data that enable impact to be assessed over time.

Nevertheless, evaluators of integrated missions have a range of innovative methodological tools available to them. This section discusses two key methodological challenges for the evaluation of integrated missions, and presents a series of practical case studies to illustrate the range of possible ways to overcome them in fragile, conflict-affected environments.

Impact evaluation and counterfactuals

As a methodological approach, impact evaluation refers to 'studies that measure impacts attributed to an intervention using experimental or quasi-experimental methods to either compare "treated" and "control" units

[51] UN, OIOS, A/67/795, 2013, p. 27
[52] UN, OIOS, A/67/795, 2013, p. 17

or compare different varieties of an intervention'.[53] This requires a valid counterfactual—usually a comparison group that was not affected by a particular policy intervention.

This kind of evaluative method is seldom used in integrated missions. Constructing valid counterfactuals in conflict-affected environments is difficult, often requiring creative solutions from evaluators. Furthermore, the denial of access to a policy intervention as required by a counterfactual raises ethical objections. Nevertheless, as the UK Department for International Development (DFID) notes, 'resources are often constrained and choices often have to be made. Programmes are not always introduced everywhere all at the same time because of limited capacity. Impact evaluations can exploit these existing constraints to learn from them about the impact of the programme'.[54]

Impact evaluation is best suited to the assessment of discrete policies with effects that can be measured at the level of individuals, households or communities.[55] Integrated missions pursue more diffuse goals with a broader range of beneficiaries, which are less amenable to the straightforward observation of outcomes. Reflecting on these challenges in development assistance, Martin Ravallion notes that impacts for interventions that have 'diffused, wide-spread benefits…are often harder to identify than for clearly assigned programs with well-defined beneficiaries, since one typically does not have the informational advantage of being able to observe non-participants (as the basis for inferring the counterfactual)'.[56] While these problems also exist at the level of a mission's component parts, they pose particular problems for integrated peace and security operations.

Counterfactual reasoning about the effectiveness of integrated missions can also fall well short of the rigour necessary for impact evaluation. As Paul Diehl notes, uncertainty about the standards for success can lead to a 'better than nothing' standard by which operations are almost always assessed

[53] Cyrus Samii, Annette N. Brown and Monika Kulma, 'Evaluating Stabilization Interventions', International Initiative for Impact Evaluation, *White Paper* (2012).
[54] Cited in Sabine Garbarino and Jeremy Holland, 'Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results', *Issues paper* (Governance and Social Development Resource Centre, March 2009), p. 3.
[55] See discussion in Samii, Brown and Kulma, 2012, p. 14.
[56] Martin Ravallion, 'Evaluation in the practice of development', *World Bank Policy Research Working Paper 4547* (2008), p. 6.

positively.[57] For example, one evaluation of the UN mission in the DRC concluded that, 'without MONUC's presence the situation would probably be significantly worse. Yet, MONUC has not adequately implemented its mandate of protecting civilians'.[58] Furthermore, classic counterfactual analysis is based on identifying what outcomes occur in the absence of a policy intervention. The relevant question for peace operations, however, is more often not whether 'doing nothing' would be better, but whether alternative policy implementation would be more effective. Indeed, the most common counterfactual implicit in peace operations evaluation is what would have happened with more intervention, not less. That is, since peace operations are so often under-resourced, discussions often consider what would happen with additional resources—more helicopters, for example—and more assertive interventions, such as a more 'robust' force posture. However, in her assessment of international operations in Afghanistan, Astri Suhrke cautions against the kinds of counterfactual reasoning that led analysts to embrace, from around 2007, the idea that the difficulties in Afghanistan were due to the 'light footprint' approach initially adopted by the international mission. According to this reasoning, an early 'window of opportunity' had been squandered, because the US-led operation had not deployed enough military and civilian resources to achieve the statebuilding and economic reconstruction objectives necessary to stabilize Afghanistan.[59] The counterfactual argument is most clearly stated by Ahmed Rashid, who argued that:

> In those critical days in the autumn of 2003, a few thousand more US troops on the ground, more money for reconstruction, and a speedier rebuilding of the Afghan army and police could easily have turned the tide against the Taliban and enhanced the support of the population for the government. It was a moment when even a little could have gone a long way.[60]

Instead, Suhrke argues, it is equally plausible that the reverse may have been true; that is, that a stronger international presence in the early post-invasion phase might simply have introduced at an earlier point the negative reactions, problems and strains that appeared in the second half of the decade. Making

[57] Paul F. Diehl in Daniel Druckman et. al., 'Evaluating Peacekeeping Missions', *Mershon International Studies Review*, vol. 41, no. 1 (1997), p. 153; see also Frances Stewart, 'Evaluating evaluation in a world of multiple goals, interests and models', in George K. Pitman, Osvaldo N. Feinstein and Gregory K. Ingram (eds.), *Evaluating Development Effectiveness* (Transaction Publishers: New Brunswick, NJ, 2005).
[58] Mobekk, 2009, p. 274.
[59] See Barnett R. Rubin, 'Saving Afghanistan', *Foreign Affairs*, vol. 86, no. 1 (2007); Seth G. Jones, 'Averting Failure in Afghanistan', *Survival*, vol. 48, no. 1 (2006); and the discussion in Suhrke, 2011, pp. 12–13.
[60] Cited in Suhrke, 2011, p. 13.

Box 2: Experimental Methods for Impact Evaluation

A recent study of evaluative methods used to assess stabilization operations demonstrates the potential for impact evaluation to contribute to improved policy, despite the inherent difficulties of counterfactual analysis in such insecure, volatile environments. Samii, Brown and Kulma conducted a comprehensive analysis of publicly available evaluations of US stabilization programmes over the past two decades.[1] They found that, although most evaluations discussed 'impact', genuine impact evaluation supported by adequate methodological rigour was rare. Instead, of the more than 50 evaluations studied, the focus was primarily 'on the process of the projects and the performance of the implementers'.[2] These kinds of evaluation—by far the most common in the context of the integrated missions investigated in this paper—can produce interesting qualitative information, particularly regarding the perceptions of those engaged in the evaluative exercise, but without being able to measure non-intervention (the counterfactual), this approach cannot attribute impact with validity.[3] Instead, Samii, Brown and Kulma. argue that: 'Experimental designs, although still in the majority, appear to be feasible less often for stabilization interventions. The careful and creative use of quasi-experimental methodologies will therefore be important for increasing the use of impact evaluation for learning about stabilization'.[4]

There is substantially greater scope for using such methods to evaluate integrated missions, particularly for targeted interventions such as ex-combatant reintegration programmes and community-level reconstruction.[5]

[1] Samii, Brown and Kulma, 2012.
[2] Samii, Brown and Kulma, 2012, p.3.
[3] Samii, Brown and Kulma, 2012, pp.3–4.
[4] Samii, Brown and Kulma, 2012, p.9.
[5] See Samii, Brown and Kulma, 2012 for case studies.

this argument, however, would shift the analysis to an examination of the shortcomings of the international project itself. At the time, this was not the focus of the dominant thinking in either academic or policy circles.[61]

This example serves to highlight the problems of poor counterfactual analysis, which bears little resemblance to the rigorous counterfactuals required by experimental methods of impact evaluation. These take the

[61] Suhrke, 2011, p. 13.

form of either a control group (the 'gold standard' of counterfactuals) or 'quasi-experimental' methods that require the evaluator to impose explicit assumptions about what would have happened in the absence of a policy intervention—an imperfect counterfactual that may nevertheless be 'good enough' for impact evaluation.[62]

Limits of self-reporting and the need for local data

Evaluations of integrated missions, as well as those of individual mission components, tend too often to rely on self-reporting by missions, and on perception data from mission staff. In part, this is a result of the volatility of operating environments, which requires deep contextual understanding to ensure evaluations are based on data collected at appropriate intervals—especially given the non-linear dynamics of change in these operational environments.

This section highlights two innovative approaches to collecting and analysing 'local' data, that is, data drawn from within the host society itself, including perception data. The first is a survey method used in the performance review and policy reform of the integrated Regional Assistance Mission to Solomon Islands (RAMSI); the second is a broader approach employed by a local Afghan NGO, The Liaison Office.

The People's Survey used by the Regional Assistance Mission to Solomon Islands

RAMSI is an integrated civil-military-police operation that since 2003 has sought to build stability and peace in this small Pacific Islands country, focusing first on the disarmament and demobilization of militia groups, and then on a comprehensive statebuilding programme with particular emphasis on the rule of law and building state capacity.

First piloted in 2006, the People's Survey is an independent annual survey that collects data about the opinions of Solomon Islanders on a wide range of issues, including business and employment, law and order, public accountability and access to services.[63] Qualitative and quantitative data

[62] See Samii, Brown and Kulma, 2012, p. 5.

[63] The complete collection of People's Surveys is available at http://www.ramsi.org/solomon-islands/peoples-survey.html.

are collected by local surveyors, who conduct surveys and focus group discussions in communities across the Solomon Islands.

The People's Survey provides an important means by which RAMSI can measure its progress against benchmarks and targets established in partnership with the Solomon Islands government. It also provides the Solomon Islands government with valuable data to inform national public policy.

The Liaison Office, Afghanistan

The Liaison Office (TLO) in Afghanistan uses an innovative research methodology to capture precisely the kinds of data too often missing from evaluations of interventions in fragile, insecure environments: rich, in-depth, context-specific information about security, governance, conflict dynamics and sources of instability relevant to analysis at the national, provincial and local community levels. These kinds of cross-cutting data are particularly valuable for integrated missions, the intended effects of which are often diffuse and thus difficult to measure.

TLO is an Afghan non-governmental organization focused on research, peacebuilding and livelihoods. In addition to its evaluative analysis, discussed further below, TLO provides a platform for dialogue among communities, and between the grassroots level and the Afghan government and international stakeholders. The organization also works to support the Afghan government's ability to carry out its responsibility to protect its citizens from violence by promoting peace and improving human security. TLO's approach has five key strengths that make it worthy of close examination and, where the conditions exist, emulation in other conflict environments.

Orientation to the host environment, not the intervention

TLO's primary focus is the host society—not the intervention itself. This overwhelming orientation to the local environment distinguishes TLO's analysis from that of typical impact evaluations of security, development and statebuilding interventions in Afghanistan. Typical evaluations of international efforts in Afghanistan, as elsewhere, put the international mission at the centre of enquiry, beginning by identifying its goals, telling a narrative of its engagement, and then proceeding to observe impact on the local environment, with varying degrees of methodological rigour. TLO's

research puts this last step first, thereby prioritizing rigorous observation and assessment of security, governance and development outcomes. That is, where most evaluations focus on the independent variable (the international interventions in Afghanistan), TLO focuses on the dependent variable (the security and stability of Afghanistan). This can help to insulate the observation of outcomes from the political pressures related to performance accountability. For integrated missions specifically, this approach helps to capture the impacts of joined-up efforts that often have diffuse, widespread effects and are thus less suited to the narrow impact evaluation of a particular programme or policy intervention.

Deep knowledge of the conflict context

Deep understanding of the engagement environment is an asset widely recognized as essential for rigorous evaluation of integrated missions. The organization is pragmatic in its approach to researching difficult, insecure environments where obtaining information is a dangerous enterprise. As an Afghan organization staffed by Afghan nationals who employ local field researchers, TLO gains better access to local communities than other evaluators usually can. The organization draws on its existing relations with stakeholders and communities to collect high quality data and to facilitate further access to key data sources. In addition to its more than 70 Afghan programme staff, TLO retains a small number of international staff with significant experience in Afghanistan to help ensure the validity and rigour of its research.

Low-profile security methodology

TLO is able to draw on its extensive experience with sensitive data collection to overcome the barriers of mistrust and insecurity that usually prevent international evaluators from obtaining crucial information about the grassroots conflict environment. TLO's evaluations draw substantial value from their embeddedness in the organization's broader peacebuilding agenda, enabling evaluators to use an action-research approach to avoid the unnecessary disturbance of local communities while also helping to ensure the security of researchers. Recruiting appropriately qualified field surveyors and field researchers who can also move with ease within the local communities being studied is crucial, so that data can be collected inconspicuously in hostile environments. This means field researchers usually come from the local communities under investigation, but must also

be known either to TLO or trusted members of its network. They travel in unmarked cars or the motorbikes common to the research area. While TLO does involve some expert international staff in its evaluation, their contributions are made largely from Kabul or remotely. Particular care is taken to limit the travel of international staff to provincial centres.

Multilevel analysis

The insecure environments in which TLO conducts evaluations demand a pragmatic approach to research, one flexible enough to accommodate the danger associated with the volatile security situation, but which can also guarantee sufficient methodological rigour to ensure valid results. Combined with the profile of its local field researchers, TLO's innovative multilayered methods enable its evaluations to triangulate data and assess significant causal mechanisms without the rigid requirements of experimental impact evaluation (see Box 2). TLO uses a combination of purposive or stratified sampling, where the selected subset of the population shares at least one common characteristic; cluster sampling, to study areas of specific interest, such as where displacement has occurred; judgment sampling, whereby the selection of key informants is based on the expert assessment of researchers; and convenience sampling, where research participants are selected from within the existing network of researchers, often in combination with a snowball technique to continuously identify new participants. In addition to these sampling strategies, data are triangulated by constructing questionnaires and interview formats to internally cross-check information (asking more than one question about one topic), which enables researchers to identify inconsistencies that may indicate poor data quality.

Research triangulation helps to overcome the problem of researcher bias. TLO often uses more than one researcher to collect the same kind of information. Furthermore, the field data collected by local expert researchers is verified by Kabul-based project staff, who conduct additional key informant interviews, monitor and compare findings with other sources of data, and conduct expert peer reviews.

Long-term engagement in the same environment

As a local NGO, TLO's analysis benefits from its continuous engagement and repeated evaluations in the same conflict environment. For example, TLO's 2012 report on Uruzgan province not only presented 'a comprehensive

assessment of the political, social, economic, and security situation' at the end of 2011, but was also able to draw on baseline data from TLO's 2010 Uruzgan report to measure changes in the 18-month period that followed the transfer of the Provincial Reconstruction Team from Dutch leadership. The value of this long-term engagement by a single evaluation team enables collection of the kind of time-series data that too often eludes impact evaluations of integrated missions.

4. Comparative Approaches to Evaluating Integrated Missions in the DRC

This illustrative summary of recent evaluations of peace operations in the Democratic Republic of the Congo serves to examine in one context many of the evaluative challenges discussed above. It presents a spectrum of evaluative methods frequently employed in such environments: a multi-donor joint evaluation, representative of the country programme evaluations common in the development community; a UN OIOS evaluation, representing the performance review method; and an in-depth qualitative study, representing the contributions that can be made by detailed academic work.

Joint Evaluation of Conflict Prevention and Peacebuilding, 2011

Eleven bilateral and multilateral donors and aid agencies operating in the eastern DRC jointly commissioned this evaluation.[64] The practice of joint commissioning can help to circumvent some political disincentives for rigorous evaluation. Nonetheless, the approach used appears to be a textbook case of blame-avoidance. Although the evaluation is based on a sample of projects implemented by specific donors, the results of those individual project evaluations are not presented. Since it assesses the aggregate effects of multiple donors, rather than the performance of any single donor, poorly performing agencies are shielded from the consequences of negative evaluation.

The evaluative framework itself disaggregates conflict prevention and peacebuilding efforts into five cross-cutting goal areas: sexual violence and humanitarian assistance; justice; security sector and demobilization; mining and regional aspects; and capacity-building. The report presents a

[64] Channel Research, *Joint Evaluation of Conflict Prevention and Peace-Building in the Democratic Republic of Congo*, Synthesis Report, vol. 1 (17 June 2011).

promising approach by explicitly declaring a hypothesis for each goal area, each related to an identified 'conflict driver'. Nonetheless, the approach falls short. Only one hypothesis is presented in each area when many are plausible, the hypotheses are overly broad and they are not rigorously tested in the evaluation. Instead, the discussion of hypotheses appears to refer more to a statement about the assumed conditions. For example, on the cross-cutting issue of sexual and gender-based violence, the hypothesis is that '[c]onflict prevention and peace-building require a strategic focus on the dynamics and the drivers of the conflict'.[65] It is not clear what standards the joint efforts would need to meet to qualify as successful, nor how the evaluators deal with the presumed variation among different donor efforts. Donor initiatives on sexual and gender-based violence are assessed in four areas. On coordination, the evaluation finds weaknesses, but no standards are defined, and the discussion relies on anecdotal examples rather than comparative data. On effectiveness, the evaluation finds that all projects reviewed were 'exceptionally good at delivering results',[66] but provides no data or other evidence sufficient to justify this conclusion. Instead, as is common in such policy evaluations, the reader is expected to take on trust the evaluation's accuracy. On efficiency, no standards or data are presented, despite the fact that this benchmark should be particularly amenable to quantitative measures, or at the very least a cost assessment. Finally, on impact, the evaluation examines the effects of humanitarian assistance on the conflict driver (state weakness), which includes more useful analysis of the interactions between donor efforts and the structural conditions of conflict, although no clear evaluation is made.

This evaluation presents one of the better approaches to the evaluation of multi-actor assistance to conflict environments. Nonetheless, the approach is in need of substantial improvement before it can be considered a case of good practice for the rigorous evaluation of integrated missions.

[65] Channel Research, 2011, p. 164.
[66] Channel Research, 2011, p. 77.

UN OIOS Programme Evaluation of Performance and Achievement of Results: 2012

The 2012 UN OIOS assessment of MONUSCO provides a typical example of performance evaluation.[67] It evaluates the mission's effectiveness at achieving the goals set out in Security Council mandates, based on a literature review (UN internal and public documents, plus external literature on the DRC), an electronic survey of a random sample of the views of 610 MONUSCO staff on the mission's achievements, and semi-structured interviews with MONUSCO staff, UN officials, DRC government officials, representatives of UN member states, representatives of civil society organizations and 'other stakeholders'. Finally, the evaluation's terms of reference and draft findings were reviewed by three 'internationally recognized peacekeeping experts'.[68]

The evaluation finds that MONUSCO 'contributed to improved security' in DRC, has 'deterred armed conflict', 'contributed to capacity-building' and 'reduced the capacity of remaining armed groups'. Furthermore, MONUSCO 'plays a critical enabling role of Government, the international community and civil society', including by supporting elections, facilitating humanitarian assistance and according particular attention to gender issues, conflict-related sexual violence and human rights. Nonetheless, the mission has made 'slow progress in security sector reform and establishing the rule of law', tasks which the evaluation notes are 'beyond the power of the Mission and the United Nations alone'. The report notes the limitations of reviewing the performance of peace operations according to the indicators specified in its budget documents:

> Like other peacekeeping missions, MONUSCO measures its achievement using indicators specified in its budget documents that are intended to measure progress towards the larger objectives specified in its mandate, including the improved protection of civilians, progress in combating impunity and improving human rights.…[But the] Mission's indicator set changes in line with mandate changes, includes indicators that yield only

[67] UN, OIOS, *Programme Evaluation of Performance and Achievement of Results: United Nations Peacekeeping Activities in the Democratic Republic of Congo*, A/66/741, 12 March 2012.
[68] UN, OIOS, A/66/741, 2012, pp. 7–8.

> a partial view of a larger situation and/or are influenced by factors beyond its control, cannot always be validated from external sources and does not measure the totality of its outcomes. As a result, *the set is an important but inherently imperfect means of measuring the Mission's impact*.[69]

To these limitations should be added the reliance on mission staff perceptions and the minimal inclusion of host society perceptions (see the sections above on experimental methods and The People's Survey for suggested approaches to correct this oversight).

Autesserre: The Trouble with the Congo, 2010

Séverine Autesserre adopts a different kind of evaluation strategy in her excellent monograph on international peacebuilding efforts in the DRC.[70] She begins, as is common in academic research, by observing an outcome: 'intense international peacebuilding efforts, including the largest peacekeeping mission in the world, have failed to build a sustainable peace in the Congo'.[71] Her conclusion is that international peacebuilders have systematically neglected the micro-level causes of peace process failure, that is, they have failed to build peace at the local level. Her explanation for this is that:

> a dominant international peacebuilding culture shaped the intervention in the Congo in a way that precluded action on local violence, ultimately dooming the international efforts.…In the Congo, this culture established the parameters of acceptable action. It shaped what international actors considered at all (usually excluding continued local conflict), what they viewed as possible (excluding local conflict resolution), and what they thought was the 'natural' course of action in a given situation (national and international action, in particular the organization of elections). It authorized and justified specific practices and policies while excluding others, notably grassroots peacebuilding. In sum, this culture made it possible for foreign interveners to ignore the micro-level tensions that often jeopardize macro-level settlements.[72]

Autesserre employs the kind of theoretical clarity and rigorous qualitative methods that evaluators of peace operations everywhere should aim for. This is a valuable contribution of scholarly evaluation, and one that

---

[69] UN, OIOS, A/66/741, 2012, p. 19 (emphasis added).
[70] Autesserre, 2010.
[71] Autesserre, 2010, p. 5.
[72] Autesserre, 2010, pp. 10–11.

policy evaluators could learn from. The phenomenon to be investigated—'culture'—is carefully defined, as are the ways in which it will be observed and analysed. Autesserre draws on existing theory to establish alternative explanations for the failure in the DRC, determines the limitations of those conventional explanations, and instead justifies the focus on an explanation underlying others; that is, that 'although material constraints, lack of national interest, and organizational constraints and interests did play roles in preventing international action on local conflict…these constraints and interests were not given, pre-existing, and objective. They were rather constituted by the dominant international peacebuilding culture'.[73]

Methodologically, the evaluation is based on painstaking, comprehensive data collection, conducted in more than 330 interviews with UN officials, Western and African diplomats, staff members of international and nongovernmental organizations, victims of violence, foreign observers, and political, military, diplomatic, and civil society actors in the DRC, France, Belgium and the United States. Most interviews were over two hours in duration.[74]

Since Autesserre sought to observe the 'culture' of international peacebuilders, these international actors feature heavily as sources. Nonetheless, as she notes, the book 'also gives voice to Congolese actors, to show how they received and interpreted this international action, and why they welcomed or fought it'. By contrast, many policy evaluations of peace operations rely on surveys and interviews with mission staff to gather data about effectiveness; that is, they ask the implementers about the success of their implementation. While such perceptions are not without value, they are not the primary means by which the effects of a peace operation should be observed and explained. Autesserre's own argument makes this plain: if she is correct and the culture of international peacebuilders shaped how they understood the world and what they perceived to be the appropriate action in ways that undermined important pathways to peace,[75] then evaluators need to treat critically the data gathered from within this culture.

Of course, the rigour of Autesserre's analysis required a book-length single case evaluation. This is not an approach suited to the demands of

[73] Autesserre, 2010, p. 23.
[74] Autesserre, 2010, p. 34.
[75] Autesserre, 2010, p. 29.

policymaking, which needs timely, accessible and clear inputs to guide decision makers. Furthermore, rigour does not guarantee accuracy, a point that appears to have been overlooked in much of the evaluative literature. The judgments, assumptions and data of an evaluation may be contestable, so that different evaluators may come to different conclusions. Nevertheless, the strength of academic literature such as *The Trouble with the Congo* is that it makes these judgments, assumptions and data explicit. By contrast, much of the evaluation conducted within the policy world, that is, commissioned or conducted by donors, implementing organizations, or the missions themselves, fails to make transparent the data and analytical decisions on which conclusions are based.

5. Conclusions

This paper has reviewed the foundations, challenges and practices of integrated mission evaluation. The analysis leads to two types of conclusions: first, a set of principles to guide the improvement of such evaluations at the strategic level; and, second, a collection of emerging 'good practices' that are not widely used in the evaluation of integrated missions, but should be more often considered in the selection of evaluative method.

An Agenda for Better Evaluation of Integrated Missions

Four thematic principles for evaluating integrated missions can be discerned from this review. While further analytical work is needed, it is hoped that identifying these guiding principles here will bolster such efforts in the future.

1. Generate evaluation standards of success and failure for a mission from the intervention context

Structured, focused comparative evaluation is necessary for theorizing effectiveness—a very important task best undertaken by academics using scholarly rigour.[76] Policymakers in governments, international organizations and NGOs should actively encourage this kind of work through funding mechanisms and by granting researchers access to collect new data and to use existing data not yet in the public domain. Nonetheless, the evaluation questions that motivate this paper are best addressed through more accessible

[76] For excellent examples of this approach see Michael W. Doyle and Nicholas Sambanis, *Making War and Building Peace: United Nations Peace Operations* (Princeton University Press: Princeton, NJ, 2006); Virginia Page Fortna, *Does Peacekeeping Work? Shaping Belligerents' Choices After Civil War* (Princeton University Press: Princeton, NJ, 2008); and Diehl and Druckman, 2010.

methods that can enable timely findings about the effectiveness of a particular mission. To this end, evaluative frameworks should be informed by the best thinking about operational effectiveness gleaned from academic studies and policy practice, but designed according to the specific goals, context and policy needs of a given operation.

2. Evaluation should be independent, balance methodological rigour with pragmatism, and be understood as part of an ongoing debate rather than a final adjudication

The gold standard of evaluation regarding integrated missions should have three core characteristics. First, it should be independent, insulating evaluation from the political pressures likely to distort analysis and findings. Second, it should prioritize methodological rigour within the constraints of insecure, volatile conflict environments; that is, evaluators should accept that real-world conditions may necessitate pragmatic compromises, while nonetheless striving to meet the ideals of scientific enquiry wherever possible. Third, the broader community of practice should approach integrated mission evaluation not as a technical exercise that produces a definitive conclusion about an operation's effectiveness, but instead as an iterative process aiming to edge ever closer to the truth, which may be subject to contested judgments and conflicting evidence, and require multiple evaluations using different methods. There is a need for further research and analysis on the models and techniques of independent evaluation best suited to the demands of peace operations, including at both the field and the headquarters levels, which should prioritize the assessment of impact not performance.

3. Comprehensive evaluation may involve disaggregation

The evaluation of integrated missions should combine two distinct levels of analysis: an assessment of the full operation and its cross-cutting themes and, crucially, disaggregated evaluation of each component of a mission and its discrete goal areas. This can help to overcome the prioritization challenge that too often undermines efforts to determine overall mission effectiveness. Evaluating integrated missions on a spectrum of success and failure requires the analyst to prioritize the operation's goals (success in what) and stakeholders (success for whom), but prioritization criteria are likely to be

highly contested. Instead, as I have argued elsewhere:

> it is the disaggregated assessment of peace operations—across goal areas and over time—that holds most value for scholars wishing to identify and understand causal processes and for policymakers seeking practical guidance in the design and implementation of peacekeeping. Questions of prioritization and effect still occur at this lower level, but considering a single goal area limits such contestation to a workable degree.[77]

4. Assessing unintended consequences is essential for evaluations of integrated missions

There is a strong case for arguing that all evaluations of policy interventions should assess their unintended effects. This must be a priority for integrated missions, since it is often precisely at the points at which military, police and civilian activities intersect that the full spectrum of a mission's effects can be observed. Evaluative practices for assessing unintended consequences can usefully build on the principles of conflict sensitivity, the primary purpose of which is to call attention to the potential for negative outcomes despite good intentions.[78]

A Useful Menu of Evaluative Methods

Finally, the paper highlights a small, selective array of emerging 'good practices' that are not widely used in the evaluation of integrated missions, but should be more often considered. The following are worth noting:

Independent local analysts, such as The Liaison Office in Afghanistan;

Institutional auditors and independent investigations units, such as the UN Office of Internal Oversight Services or, in a national context, the US Office of the Special Inspector General for Afghanistan Reconstruction;

Issue-specific quasi-experimental impact evaluation using rigorous social science methods; and

Stakeholder evaluations and public perceptions using methods similar to those of the People's Survey in the Solomon Islands.

[77] Jeni Whalan, 'Evaluating Peace Operations: the Case of Cambodia', *Journal of International Peacekeeping*, vol. 16, no. 3–4 (2012), p. 52.

[78] See also Chiyuki Aoi, Cedric de Coning and Ramesh Thakur, *Unintended Consequences of Peacekeeping Operations* (United Nations University Press: Tokyo; New York, 2007); Whalan, 2012, p. 54.

RECENT CHALLENGES FORUM OCCASIONAL PAPERS

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

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

Evaluating Integrated Peace Operations

This paper reviews the state of the art on evaluating integrated peace operations. It examines the intersection between two prominent trends that have characterized peace operations over the past 10–15 years: the progressive integration of security and development objectives, and the increasing demand for comprehensive evaluation of policy interventions.

It identifies a set of inherent constraints to evaluating integrated operations, develops principles for improving evaluation, and presents examples of good evaluative technique to inform future policy development. Case studies are used throughout to illustrate and extend the analysis. Finally, the paper highlights a number of emerging 'good practices' that are not widely used in the evaluation of integrated missions, but should more often be employed.

JENI WHALAN

Dr. Jeni Whalan is a Lecturer in International Security and Development at the University of New South Wales Australia, and a Research Associate at the Global Economic Governance Programme, University of Oxford. She is the author of How Peace Operations Work: Power, Legitimacy, and Effectiveness (Oxford University Press, 2013) and a founding member of the Security Council Analysis Network (SCAN). Her research focuses on peacekeeping, the United Nations, fragile states, policy evaluation and international security governance. She holds a DPhil and MPhil in International Relations from the University of Oxford, where she was as a Rhodes Scholar, a Wai Seng Senior Research Scholar and a Wingate Scholar. She has served as a foreign policy advisor in the Australian Government's Department of Prime Minister & Cabinet and Department of Defence, and as a consultant to the International Peace Institute, the International Forum for the Challenges of Peace Operations, and the UNESCO Education for All Global Monitoring Report.

ISBN 978-91-87927-00-3