Assessing Complex Peace Operations

Some Considerations of Methodology and Procedure

Benjamin de Carvalho and Ingrid J. Aune

Norsk Utenrikspolitisk Institutt
Norwegian Institute of International Affairs

Multinational Experiment 6

NUPI Working Paper 782

Publisher: Norwegian Institute of International Affairs

Visiting address: C.J. Hambros plass 2d
Address: P.O. Box 8159 Dep.
NO-0033 Oslo, Norway
Internet: www.nupi.no
E-mail: info@nupi.no
Fax: [+ 47] 22 36 21 82
Tel: [+ 47] 22 99 40 00

Assessing Complex Peace Operations

Some Considerations of Methodology and Procedure

Benjamin de Carvalho and Ingrid J. Aune

This report is part of the Norwegian engagement in the Multinational Experiment 6 (MNE-6). The project is financed by the Norwegian Ministry of Defence and is managed by the Chief of Defence through the Innovation, Network Capabilities and Information Infrastructure Command (INI). Besides NUPI, the Norwegian Defence Research Establishment (FFI) and the Defence Staff College are also engaged in the programme. Further information can be found at: http://mne.oslo.mil.no

As international peacekeeping and peacebuilding operations have increased both in depth, scope, and sheer number, so have also the budgets allocated to peacekeeping – both by the United Nations (UN) and groups of states (UN 2000, UN 2005, UN 2010)[1]. Following this dramatic increase in budgets has also been a growing scrutiny of the how this money is spent, including stronger and stricter auditing measures and a growing emphasis on monitoring, evaluation and assessment of the results of these operations.[2] However, few studies of the overall effects of this shift towards greater accountability have been published, and the literature addressing this shift is scant (a notable exception is the recent Diehl and Druckman 2010).

The mixed records of past international operations have largely given rise to the demand for more clarity on what constitutes effectiveness and success in international peace operations. The general desire to better understand and minimize the risk of negative impacts from international interference has also contributed to this demand for increased monitoring and evaluation (de Coning and Romita 2009: 2). But while evaluations are well-established within the humanitarian and development segments where they play a major role in efforts to assess effectiveness and performance (Borton 2009: 166), this is not the case with international peacekeeping and peace building operations. However, things are changing. As the budgets of international peace operations have increased, so has the budget for assessing them[3]. Since the 1990s, donors have steadily increased their demands for information from the organizations they fund (Borton 2009:161, 165).

While we can see the initial spurs of international peacebuilding already before the 1990s[4], the agenda really took off in the late 1990s, when peacebuilding became an intrinsic part of most international peace operations (see Paris and Sisk 2007). While the early period of implementing peacebuilding mandates in peace operations saw emphasis on the implementation phase, the field of peacebuilding evolved during the 1990s to involve a focus on "lessons learned". This focus, from being more ad hoc initially, contributed to the professionalization of the field of peacebuilding. However, while arguably a field in its own right (see Sending 2009), the methodologies underpinning the evaluation of peacebuilding and the methods and procedures

1 The annual cost of UN peacekeeping amounts to around 7,2 billion US dollars a year, but has yet to face drastic budget cuts during the financial crisis (Gowan 2010).
2 See for example Meharg (2009), Borton (2009), Power (1997).
3 The actual amount of money allocated to assessment of peace operations may be difficult to reveal as the costs of reporting often are embedded within program costs (Borton 2009: 166).
4 The UN mission in Congo in the 1960s, which was designed to oversee the departure of Belgian colonial troops, got caught up in the civil war (Paris and Sisk 2007: 2).

informing them are largely borrowed from related fields, most notably that of development.

But this focus on lessons learned and evaluations is not solely a consequence of the professionalization of the field of peacebuilding. In fact, it can be argued to be as much the product of donors' increased focus on accountability. The 2008-2009 annual budget for UN peacekeeping was approximately USD 7,1 billion compared with USD1,5 billion a decade ago (UN 2009). As peacebuilding activities increased in scope and depth with costs increasing accordingly, donors demanded to see results and came to be more reluctant to fund activities which did not have a clear impact[5]. The field of peacebuilding can thus be said to have moved away from focusing mostly on the effective implementation of measures and programs, to address the effectiveness and impact of these. The UN Security Council, for instance. focuses increasingly on indicators of progress and figures in its renewal of mandates, while the UN family as a whole is implementing systems for increasing the accountability of peacebuilding activities through a focus on joint planning and indicators of progress (IMPP).

This focus on effectiveness and impact must also be understood in conjunction with the introduction of greater measures of new public management in the accountability systems of international organizations (See Power 1997).

Auditing has increased following the programmatic commitments to the reform of public sector administrations since the mid 1980s. Accordingly, evaluation programs from the development sphere were introduced to the field of humanitarian programs and operations. During the 1980s the United Kingdom's Official Development Assistance (ODA) evaluated its funding of responses to the African food emergency in the mid-1980s, and the Netherland evaluated its response to the famine in Somalia (Borton 2009: 161). While the impact of new public management should not be overplayed, it must be noted that some authors today speak of a "shift from the welfare state to the regulatory or evaluative state" (See Carter, Klein and Day 1992). Sarah Meharg makes a similar point when she argues that "Along with the development of [the] public services contributing to the global common goals of peace and security has come the privatization of these services. The coupling of the private and public sectors in these social capital endeavors has produced a robust peace economy." (2009: 3)[6].

5 For further considerations on impact assessment (of humanitarian interventions), see Harpviken et al. 2003.

6 See also Call (2008)

The present paper addresses this shift and seeks to provide a summative and readily available overview of the field as well as raising a few issues pertinent to the further development of the field. Our underlying argument here is that the shift towards more complex and quantitative methods for assessing the effects of peace operations, as well as the budgets devoted to assessments of different kinds, may lead the focus away from understanding the operation itself, its goals and performance. Just as in the field of development, assessment of effectiveness, audit and evaluation efforts in international peace operations have increased both in scope and depth. As we point out, the setting in which peace operations are assessed is a crucial element that needs to be taken into account by any evaluation. As a consequence, an overreliance on quantitative measures of effectiveness may hinder an effective assessment of operations. As causality is difficult – or impossible, depending on methodological perspectives – to establish, relying on models which set causality as a premise for the evaluation itself (*ex ante*) will often lead evaluators to oversee key developments and processes in the field. As such, we make the case for relying more on thicker qualitative descriptions which emphasize narrative rather than presupposed causality. *Understanding* peace operations, we argue, will tell us more than *measuring* them.

First, we provide an overview of different types of evaluations, as well as the theoretical frameworks and perspectives underpinning them. Our aim here is not to be exhaustive, but rather to offer a schematic typology. From there we move on to discuss some of the challenges to undertaking collaborative evaluations across organizations and actors representing different segments. Finally, we discuss some of the challenges specific to evaluating peace operations in a volatile context.

I. Types of Evaluations: A Typology

Evaluators can choose from a plethora of perspectives and methods. The goal of evaluation of peace operations is to assess whether they have been successful according to certain criteria, but there is no agreement among scholars and practitioners as to which criteria should be decisive (See Druckman and Stern 1997). The following paragraphs show how the choice of evaluation perspective itself influences what constitutes effectiveness in a peace operation, as well as the criteria of success.

The choice of method for evaluation will generally depend on what is to be assessed. In assessing peace operations[7], we can group what is to be assessed in five categories. Firstly, an assessment or evaluation can assess *policies, outcomes* and *impacts*. The criteria of effectiveness here is whether outcomes are defined and met (Hansen 1988). Secondly, evaluations can address internal processes. Here, the criteria of effectiveness is whether internal processes work or not. Third, evaluations can also address the *symbolic effects* of policies or actions. In this case, the criteria of effectiveness is not whether the policies or actions in questions produced the desired or planned outcome, but whether they were *perceived* as effective or not (Reitan 1993, Gaertner & Ramnarayan 1983: 97). A fourth category that can be evaluated with regard to peace operations is *funding.* The central element of effectiveness here is whether funding has been allocated in accordance with priorities, and if these priorities were good ones. Finally, we can speak of system-wide evaluations of the impact of a peace operation. What is evaluated here is not a specific policy or program, but the totality of them. The criteria of effectiveness here will be whether the system in question can be seen to progress or not.

Evaluations must also be planned at an early stage. Otherwise, they run the risk of being ad hoc, and having a character of 'going through the motions' rather than a purposeful evaluation. Donors, policy makers and executioners must also early in the evaluation agree upon what type of benchmarks to evaluate in order to make the evaluation as relevant as possible. A good evaluation rests either on the common benchmarks agreed upon, but may equally rest on the operation's abil-

[7] The tools of evaluation are borrowed from related fields, and have found their way from organizational theory and public administration to development and peace operations. Michael Power (1997) discusses how the phenomena of auditing, originally being constrained to financial auditing, has been promoted to a variety of areas since the late 1980s. We see a similar development when it comes to evaluation.

ity to report according to (any donor's) established benchmarks where these are not considered most crucial on the ground[8]. The type of evaluation pertinent to a specific process will largely depend on the time at which the evaluation is undertaken. Before an operation, evaluation undertaken will generally involve *baseline studies* or *ex-ante evaluations*, while during the operation itself, assessments will generally involve *Real-Time Evaluations (RTEs[9]), formative evaluations*, or *mid-term evaluations.* After the end of an operation, evaluations will typically be *summative evaluations* or *ex-post evaluations.*

1. Evaluating Outcomes and Impacts

Evaluating outcomes and/or impacts of policies or a specific program is the most common and widespread form of evaluation in peace operations. They can involve both evaluations of specific programs (evaluations of a set of time bound interventions, marshalled to attain specific objectives) or a cluster of related programs (a set of related activities, projects and/or programs). This form of evaluation takes the starting point of activities and organizations as discrete activities and processes. This does not entail that these evaluations do not recognize the fact that different processes may be related and entangled, but the aim of the evaluation is one specific outcome or impact or a specific set of them. Such evaluations can involve cost-benefit analyses[10], but do not have to. They also tend to rely largely on the use of statistical analysis. The point being that a process or policy must be seen as having been effective if the outcomes or impacts have been met (see Reitan 1993: 252-256) .

The general understanding underlying these evaluations is that of organizations as closed, rational systems, whose activity is effective if it meets its set objectives. In this perspective, the effectiveness of an international operation is determined by (i) defining its objectives[11] and (ii) seeing whether they have been achieved or not (Hansen 1988). But such a view is not unproblematic. Inferring from goal attainment to the effectiveness of an organization overlooks the difficult analysis of cause and effect[12]However, since donor organizations began to use the Logical Framework Approach (LFA) in the 1980s, the use of derived

8 The latter may even constrain the effectiveness of an operation, as the situation for the executioner of any donor's policy as both goals *and* process are defined elsewhere, and the evaluation process becomes a tool to satisfy donors rather than a tool to increase the effectiveness of the peace efforts.

9 A real-time evaluation is one carried out whilst a programme is in full implementation and almost simultaneously feeds back its findings to the programme for immediate use. (UNICEF 2003)

10 Cost-benefit analyses are concerned with the worthwileness of a program – whether the outcomes might be considered worth the peace operation or not. (See Mohr 1995: 2 ff,.)

11 Defining the objectives or even a clear mandate of a peace operation has proven to be difficult: in some operations, as the Multinational Force in _Beirut, the coalition partners have even operated with different mandates (Diehl 1997).

12 For further discussion on causality see for example Mohr's *Impact Analysis* (1995),

models such as linear models of causality have become widespread (Borton 2009: 164). Originally intended as a tool to help planners and managers design and manage projects and programs, these methods have been converted into audit tools. Most humanitarian programs need to prepare a Log Frame at the outset of a program and then report against it (Borton 2009: 164).

The challenges this type of evaluation or assessment meet is that the outcomes and impacts of projects or policies are often difficult to measure. It is often also difficult to clearly isolate the effect of a specific policy or action (see Mohr 1995: 274 ff.). As such, it can be difficult to decide whether an outcome is the result of the activity in question or even map out the various impacts; the reason for this being that it can be difficult to point at clear causalities. Finally, it could be asked if this type of evaluations effectuated have tended to rely too much on statistical analysis, and whether thicker qualitative descriptions might not have given a better account of the causal mechanisms involved.

2 Evaluating Internal Process(es)

These are evaluations of the internal dynamics of implementing organizations. Evaluations of peace operations routinely involved assessing whether internal processes have functioned properly, and if they have delivered effectively in accordance with set aims. The criteria of effectiveness in such evaluations or assessments is whether internal processes *themselves* work, rather than whether they have produced a specific outcome or had a desired impact. The indicators evaluators will use will consist more of description of processes rather than documented effects. The evaluator will focus on how appropriate the processes within the organization are, and will strive to map the extent and significance of the symptoms of inefficiencies (Bolman and Deal 1991). These evaluations thus often stipulate a clear causal link between process and output, which may be problematic. Furthermore, it presupposes that the activities in operation are suitable and effective in relation to that which is to be achieved. Since the process indicators describe the organization's efforts rather than results, this type of evaluation may contribute to an undesirable goal displacement, where the results become uninteresting and irrelevant (Reitan 1993: 257).

It could be argued that research has neglected organizational performance as a determinant of peacebuilding success. Turid Lægreid explores how joint evaluations have facilitated the development of common policies and standards in multi-national operations. As the international peace operations are fairly complex in their nature, the learning from joint evaluation will “depend on the ability to elaborate

common ground rules and the inclusiveness of the process." (Lægreid 2009:30).

While this type of evaluation may be easier to undertake than evaluating outcomes, the problem encountered is that they often assume a link between process and outcome (Gartner and Ramnarayan 1983: 100). Such a fundamental assumption can often lead evaluators to emphasize process at the expense of outcomes, depicting a rosy situation where there may in fact be a number of problems with the causal link assumed. Also, these evaluations do not take into account coherence between processes, nor the coordination with other actors and processes. And while these limitations in and of themselves do not constitute grounds for not evaluating internal evaluations, one has to be mindful of them using these evaluations to make inferences about the impact of processes.

3. Symbolic Effect(s)

As peacekeeping and peacebuilding agendas have increased in scope and ambition, evaluations of their effects have had to follow and evaluate not only their material impact, but also how their impact interact with their environment, what we can call their symbolic effects. For the effect and success of international peace operations increasingly hinges upon other factors than their purely material accomplishments. Questions pertaining to "hearts and minds" and the legitimacy of an operation are becoming more salient as key objectives to operations. The criteria of effectiveness here does not depend on whether the operation "actually" achieves what it is set out to do, but on whether one can master signals and symbols. The effectiveness of the operation is thus less a matter of substance than of perception. Following Gartner and Ramnarayan's analysis, an effective peace operation would be "one that is able to fashion accounts of itself and its activities in ways in which [external and internal] constituencies find acceptable" (Gaertner and Ramnarayan 1983: 97).[13]

It is important to note that such evaluations in themselves can also have an important symbolic effect internally in terms of morale, as they address questions of whether what is being done is right, and whether it is important. Hansen (1988) makes this point, arguing that while this type of evaluations are little concerned with actual change, they may create an understanding about the importance of the activities undertaken (See also Gaertner and Ramnarayan 1983).

[13] Gartner and Ramnarayan (1983: 97, 105-106) present a framework to assess effectiveness by examining organizational outputs and internal processes as well as the linkages between them, nd find this crucial in order to deal with the differing views on what part of an organization's actions are most important – these are viewed differently from an external and an internal perspective.

The challenge of such evaluations, however, although also their strength, namely that the focus is less on substance and actual change than on perceptions. Such evaluations will thus give few indications as to a peace operation's actual ability to create peace on the ground. Accordingly, attaching too much importance to such evaluations can easily lead to misguided resource allocation if successes lead to continued activity at the expense of other activities. Furthermore, it is important to keep in mind that impressions can often be misleading. The strong focus on perceptions can also adversely affect the quality of outcomes. While an evaluation may portray an operation as relatively successful, such view of "effectiveness based on outcomes" may lead to the continuation of less than optimal modes of organization and policies (see Scott 1987).

4. Funding perspective

Effectiveness may also be seen as the peace operations ability to attract funding or political support for its activities (see Yutchtman and Seashore 1967). Assuming limited budgets and scarce resources, an operation must be perceived as effective as long as it attracts financial and political support from donors and governments. While such a perspective will seldom be used in assessing an operation, it is nevertheless important to keep in mind, as it may be one aspect which actors themselves understand as crucial to the continued success of an operation.

5. System-Wide Evaluations

One of the most important types of evaluations of peace operations – and perhaps the least undertaken as it is the most complex – is system-wide evaluations. The aim of these is to evaluate the response by the whole system to a particular disaster event or complex emergency, or of the system-wide impact of activities. The key feature in this type of assessment is the response by the whole system (higher order impact) to the operation. Generally, these evaluations involve aggregating findings from a series of evaluations in order to focus on the system as a whole and will involve a mapping of the various stakeholders' criteria of effectiveness.

While these evaluations are complex to undertake at the practical level, they also involve difficulties of a more theoretical character. For instance, how does one reconcile multiple (and often conflicting) criteria of effectiveness? Is an operation effective if most outcomes and impacts are met? The answer generally given, is that these evaluations have to be informed by a clear theoretically founded ”story” or ”angle” in order to interpret findings, and some stakeholder's perspectives

will be given a stronger emphasis than other. Although they do not propose a universal criteria against which to evaluate missions or peace operations, Druckman and Stern (1997: 163) argue that it *should be* possible to develop a criteria "that allow a mission to be judged against any objective, whether a particular analyst values that objective or not".

The challenges system-wide evaluations pose are many. Due to their complexity, they must be undertaken by a multitude of actors if one can hope to somewhat grasp the whole system. One perspective will not do it. Actors involved also have different timeframes, methodologies and methods, as well as measures of effectiveness. Assessing peace operations may therefore become more of a mapping-exercise, than an evaluation: the point being precisely that there is a multitude of actors and stakeholders involved – with differing views on what factors are essential for the success of a peace operation; thereby the factors of effectiveness (See Reitan 1993, Druckman and Stern 1997: 157). There is also no clear answer as to which actors should be involved. While donors might focus on (specific) outcomes and impacts, local capacities might be more concerned with the quality of these results and the process leading up to them[14]. Finally then, how does one weight different or contradictory perspectives, as well as take into account the self-interest of other evaluators?

[14] Reitan (1993) discusses a similar perspective.

II. Practical Challenge to Evaluations: The Problem of Terminology

Undertaking evaluations or assessments of international peace operations[15] poses a number of challenges. These, if they cannot be overcome, must at the very least be addressed by any evaluation aiming at giving an accurate understanding of the effects of an operation. These challenges comprise a host of different issues[16]. Different actors operate with different terminologies. All actors involved have different organizational cultures and identities which often can hinder effective cooperation and understanding. Evaluations often have different aims, scopes and objectives and as such it may be difficult to rely on them in order to compare findings or undertake more comprehensive evaluations.

Timeframes may also differ. While some evaluations take as their point of departure a short-term program or process, others may evaluate longer term or higher end goals or impacts. Finally, methods differ widely, and the underlying methodologies may vary to the point of making the (necessary) cooperation between evaluating actors impossible. The paragraphs below aim at giving a brief overview of these issues, and provide some reflections on how to overcome them.

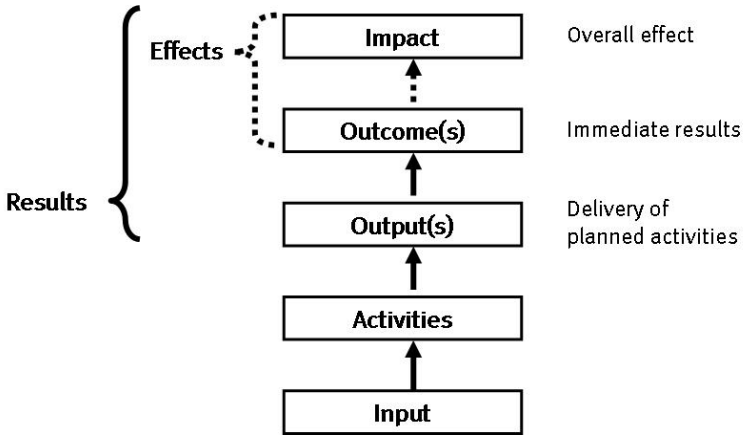
Before proceeding further with the importance of common terminologies, an important distinction must be highlighted, namely the difference between evaluations and monitoring. While evaluation can be taken to refer to the ”systematic assessment of policies, programs or institutions with respect to their conception and implementation as well as the impact and utilization of their results” (Rossi et al. 1999), monitoring generally refers to ”a continuing function that uses systematic collection of data on specified indicators” (OECD). Terminologies also diverge widely when it comes to describing stages of an operation as well as hierarchies of aims.

By way of illustration, consider the following examples of terminologies used by OECD DAC and military experts:

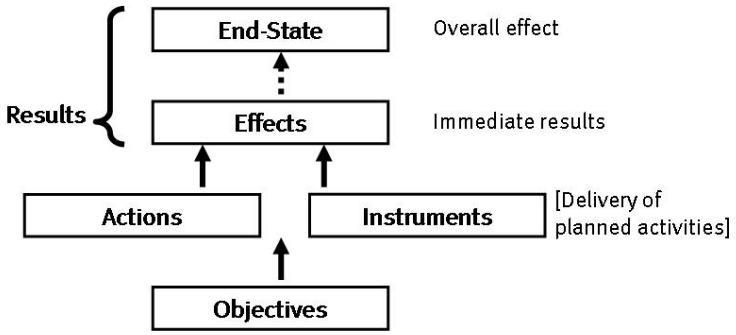
---

15 It is important to distinguish between evaluations of peace*keeping* and peace*building*. Peacekeeping generally involves less tasks and has more straightforward mandates which also makes the activity more easy to grasp for evaluators. Peacebuilding, on the other hand, is more complex and involves a set of (often) interrelated tasks and goals which can make it less palpable to evaluators and their methods.

16 For further considerations on peacekeeping and how its impact should be evaluated, see Druckman and Stern (1997).



OECD DAC terminology



Military terminology as employed in the Multinational Experiment series.

Attempting to sideline the two would result in the following table:

	OECD DAC		MNE
Final Result	Impact		End State
Indirect Effect	Effects		Outcomes
Result			Outputs
Activity	Actions	Instruments	Activities
			Input
Goals	Objectives		Theory of Change?

Terminology is a key issue when dealing with evaluation and monitoring, as there is no consensus on terminology, and many actors refer to the same processes with different names. This problem is more acute than many realize, and it is not limited to different terminologies employed by civilian and military segments, but also within these segments (civilian-civilian and military-military). But overcoming the challenges of different terminologies is not necessarily an easy matter. Different segments have different identities to which both stereotypes and prejudices are attached. In addition to this, dialogue between segments may be hindered by an often widespread reluctance to share information with each other. Civilians in peace operations often complain about militaries not understanding the aims of the operation, while militaries often regard civilians with a certain scepticism.

The lack of common terminology is also endemic to the methodologies for assessing, monitoring and evaluating an operation, as witnessed by the following comparison:

OECD DAC Criteria for Evaluating Development Assistance		**ALNAP/OECD adaptation for conflict prevention and peace-building**
1. Relevance	⟷	1. Relevance/Appropriatedness
		2. Connectedness
		3. Coherence
		4. Coverage
2. Efficiency	⟷	5. Efficiency
3. Effectiveness	⟷	6. Effectiveness
4. Impact	⟷	7. Impact
5. Sustainability		

The challenge of different terminologies poses two fundamental questions. The first question concerns whether or not the challenge can realistically be overcome. The second is of a more normative character and concerns whether this challenge should be overcome, or if the monitoring and evaluation communities are best served with a myriad of terms reflecting their different methodologies and foci.

Standardization of terminologies with a view to develop a common terminology is a favored alternative by many, including the OECD DAC[17]. Such an alternative would ensure that all involved in evalua-

[17] The Working Party on Aid effectiveness (WP-EFF), one of the nine subsidiary bodies of the OECD's Development Assistance Committee (DAC) has established itself as the international partnership on aid effectiveness (OECD 2010a). The WP-EFF has organised three High Level Meetings on Aid Effectiveness in Rome (2003), Paris (2005) and Accra (2008), with the fourth coming up in South Korea (November 2011) (OECD 2010b). The Rome High-Level Forum on Harmonization was devoted to the harmonization of development efforts and "to streamline donor procedures and practices". (Aid Harmonization and Alignment 2010).

tion work employ the same terms. It requires of course that the common terminology will be made readily available to all users, both donors, evaluators and actors in the field. However, such an alternative also has a number of obvious and less obvious flaws. Firstly, keeping up these standards would be impractical, as there is no single body that could ensure that the terminology is not only upheld, but also that it evolves. More importantly, perhaps, is however the fact that the sheer complexity of such an endeavour may make it difficult to achieve in practice. For how should one proceed in order to develop a common terminology across segments and organizations which addresses the needs of all actors involved and all situations?

Translation of terminologies is another possibility. Such a solution would entail a glossary of correspondence between terms which could be applied when using evaluations in a comparative manner. However, such an endeavour would run against severe obstacles as well. Firstly, it would be time consuming, and it would be difficult to learn all terminologies involved. Further more, even identical terminologies do not guarantee that the various actors taking part in a peace operation perceive the terminology in the same manner[18]. More importantly, however, is the fact that such an endeavour would come up against the fact that terminologies are often inconsistent and inconsistently used – even within one organization or segment.

As such, the fallback option for overcoming the challenges posed by diverging terminologies probably lies in being aware of the multitude of terminologies used in the field, and interrogating their meaning. Open-mindedness and resisting standardization as such is probably the most realistic and effective means of overcoming the challenges of different terminologies. Actors engaged in evaluations across sectors or organizations should thus show a willingness to cooperate, and to ask questions such as “what is meant by that?” Such an approach to resolving the challenge of different terminologies also has the advantage of increasing understanding of different methodologies and working procedures. For different terminologies is not only a question of simply calling things by different names. They can also be symptomatic of widely differing methodologies and ways to address the problems evaluations pose.

18 The shortcomings of translation of terminologies and wording are thoroughly discussed within the field of internasjonal law, see Ruud and Ulstein 2002: 36 ff.

III. The Problem of Knowing: Considerations of Methodology and Procedure

Evaluations involve a host of methodological challenges that must be addressed. The first – and probably the most important – of these is the problem of causality. For, while a program or a cluster of programs may have set out to have a specific impact, the fact that this specific aim happens does not necessarily mean that the programs have been effective. A specific outcome may have been produced by something else than the program itself – or may even have been produced because the program has failed. Making causal claims about the social world is a difficult – if not impossible – task. For while we may believe that activity A (eg. Anti-hate propaganda on the radio) will lead to outcome B (less violence) happens does not mean that it was because of activity A. In the case in point, outcome B (less violence) may very well have been the result of another activity C (work with traditional leaders). Thus, it is important to focus evaluations on addressing the question of "what happened?" rather than "did A cause B?" or "has B occurred?".[19]

The second methodological issue concerns the choice between qualitative and quantitative approaches. Some actors favour the latter at the expense of the former. One of the reasons for this may be a tendency to believe that quantitative approaches are more easily prone to generalizations and provide more "secure" knowledge. On the other hand, those favouring qualitative approaches often make the case that these give evaluators a "deeper" knowledge of the processes in play, through thicker descriptions. Quantitative approaches also tend to favour more generic evaluation tools and methods, often relying on surveys. The problem with these tools is that they rely heavily on prior understandings and operationalizations of higher end impacts, and leave less room for adapting the process to the field. The questions evaluators must ask themselves here is whether generic tools can be used, or if more qualitative approaches should be favoured – possibly used to complement – in order to provide thicker descriptions. The crucial issue presenting itself here is of course whether evaluations of impacts (eg. "democracy") is feasible or if one should concentrate on outcomes and effects instead – as these are easier and less controver-

19 For further discussion on multiple outcomes see Mohr 1995: 274 ff.)

sial to operationalize. For, when operating with impacts or so-called "end states", when can we say that these aims have been reached? As with the example of "democracy" above, can we expect evaluators to agree on what constitutes the end state for democracy when there is no scholarly or practical consensus on what democracy involves in the first place?

This leads us over to the third issue, namely that of understanding the conflict and its context. Can one expect evaluators – as well as donors who read evaluations – to understand all the facets of a specific conflict? Firstly, very few people involved have such an understanding, and professional evaluators often run from one conflict to another. The pressing question here is whether generic understandings of conflicts will prevail over the local knowledge necessary. This again raises two interrelated issues, which will be dealt with in more detail below, namely the status of the so-called "theories of change" underlying evaluations, as well as their baseline studies. The central question here is whether theories of change are robust enough. As there is little scholarly agreement as to what historically has triggered successful transitions from violent conflict to peaceful democratic development, can we expect evaluators and policy planners – who often lack knowledge of the specific conflict – to have such an understanding. In practice, evaluation of peacebuilding often relies on theories of change which at best are educated guesses about peaceful development.

As conflicts in which the international community intervenes have often been going on for years, there is in general little information about the economic, social and political situation of the country readily available. This in turn makes setting a baseline against which to evaluate or measure an arduous task. The overall initial understanding of context and starting point of an operation is crucial, as getting it wrong will make it more difficult to evaluate.

IV. Challenges to Evaluating Peace Operations

These general problems evaluators face are exacerbated when evaluating peace operations, as the setting is more volatile and difficult to control. The security situation may also make access difficult for evaluators, while the information necessary in order to conduct an evaluation may often be lacking. Baseline studies may be difficult to undertake and material may be difficult to gather. The wide-ranging quantitative studies generally necessary for setting a baseline may not be available and surveys may be equally difficult to undertake with any reliability. And while progress may be assessed, how does one decide on what constitutes success?

The first issue that need to be addressed when undertaking an evaluation of the progress made by a peace operation is to have a clear understanding of the point of departure against which to measure. It order to know where one stands in a process, it is crucial to know where one was when starting. The problem, as emphasized above, is that in conflict settings such information is scarce and not readily available. Information available may not always be compatible with the type of evaluation one wishes to undertake. Furthermore, there may be time constraints involved which may make a baseline study more difficult. The environment may also not allow for the necessary wide-ranging study. And while such a study may be possible, information and figures may be difficult to gather. For instance, many baseline studies rely on official figures for assessing the situation. However, in conflict situations, where infrastructure may have broken down for years, the provision of services and the economy are often informal. Thus, while updated information from informal sources may be scarce, official figures will most certainly be outdated.

As mentioned above, there is also no consensus on what methods are best suited for a specific type of evaluation. In our present case, one could ask if the wide-ranging quantitative methods borrowed from the development field are suitable for assessing conflict and peace operations. These methods for evaluating impact are often developed for assessing long-term impacts. Are they suitable for evaluating more medium-term developments in peace operations? Furthermore, how does one operationalize “peaceful development”, for instance? “Economic growth” is a much easier dependent variable to operationalize. There is also little agreement on what type of indicators to use, and where there is agreement, data may be difficult to gather. While capi-

tals are often fairly safe, travelling across a country to gather the necessary data in order to have a picture of the situation as a whole may not be feasible. These difficulties render data inherently difficult and expensive to gather. The pressing question of course being what price one is willing to pay for evaluating an operation, and at what point this price may take resources away from the operation itself. If one adds to this the fact that many evaluations are undertaken more as a matter of routine than for any specific purpose, or that the conclusions regarding whether a specific program, a cluster of programs or the operation as a whole will continue and how has often been reached before an evaluation is undertaken or completed, the future for evaluating the effects of peace operations – especially the broad peacebuilding agendas – is rather gloomy. Maybe the evaluation agenda borrowed from the development field is overly ambitious and does not take into account the inherent difficulties of assessing and evaluating in conflict settings. Maybe one should focus more on evaluating across a more narrow scope (eg. death rate) rather than including figures which at best are educated guesses. At any rate, as long as there is no clear aim for what to do with an evaluation and as long as their findings are unsure, one should be careful not to evaluate too much, knowing that their findings may be highly unreliable.

For the key issue here is what happens to evaluations. To what extent are they useful, and to what extent are they used to improve ongoing processes? Quite often, they are used to legitimize policies already decided upon or by a specific organization or department as a bulwark against policy-makers and donors eager to control what from the field may be seen as too much. de Coning and Romita (2009) make the case for a deficit in evaluations of peace operations, arguing that “monitoring impact and evaluating results should receive at least the same attention as oft studied areas like planning, coordination and integration”. Meharg (2009: 13) makes the case for more deeply ingrainin evaluations and measurings of success in peace operations. But does *more* assessment lead to *better* assessment?

To the extent that evaluations measure the success of peace operations, the process of evaluating efforts also involves a clarification of the purpose, aim and objective of the operation (Meharg 2009:4). Assessment is therefore important not only for measuring success, but also for defining goals, which may not always have been clear at the onset. The question here of course is whether evaluations and assessments are the correct tool for clarifying aims and goals of peacekeeping operations. Given these caveats, do we evaluate too much? A focus on evaluation can come at the expense of activities and processes. The question could also be turned on its head. Do we evaluate too little, and do we have too little information about the effectiveness of activities?

IV. Concluding Perspectives

Peacebuilding is a young field. Understanding the road to peaceful development requires a deep understanding of specific contexts, as well as broad knowledge of comparable cases. The pressing question to answer with regards to evaluations of peacekeeping and peacebuilding is whether such comparative knowledge is available – do we have enough cases against which to compare? – and whether operations are organized in such a way that knowledge of the specific case at hand is available. Given these limitations, it is a pertinent question to ask whether expectations with regards to the aims and impacts of peace operations are set too high. The question goes for what to expect of evaluations as well. Does a focus on evaluating effectiveness contribute to raise the bar for what is realistically to be expected of peace operations, and does a focus on too rigid methods of assessing effectiveness give the impression that peace operations can actually be measured precisely?

As evaluations are time and resource intensive, overly focusing on evaluations when not much can be hoped to be gained through them may come to compromise the resources of the operation. Resources spent on evaluation will not be spent on the operation. As the situation is today, agencies involved in peace operations are prone to "reporting overload." Actors involved in the operations frequently receive their funding from many agencies and are already overburdened with reporting to them all (see Borton 2009: 166).

This makes it all the more important to ensure that the evaluation is undertaken for a specific purpose; namely maintaining international peace and security. As large-scale quantitative survey evaluations are difficult to undertake in conflict settings and as they are all the more resource intensive, there may be – as far as evaluating peace operations is concerned – a case for relying more on thicker qualitative description which emphasize narrative rather than presupposed causality. Understanding peace operations may in the end prove more useful than measuring them.

References

Aid Harmonization and Alignment 2010, *Rome High-Level Forum (2003),* [online]. Available at: http://www.aidharmonization.org/secondary-pages/editable?key=106

Bolman, L.G. and T.E. Deal 1991, "Leadership and management effectiveness: A multi-frame, multi-sector analysis", *Human Resource Management*, Volume 30, Issue 4, pp. 509–534, Winter 1991

Borton, J. 2009, "Trends and Challenges in Measuring Effectiveness in the Humanitarian System", pp. 159-168 in S. Meharg (ed.), *Measuring What Matters in Peace Operations and Crisis Management,* Montreal and Kingston: Pearson Peacekeeping Centre.

Call, C.T. 2008, "Knowing Peace When You See It: Setting Standards for Peacebuilding Success", *Civil Wars,* 10, 2. June 2008, pp. 173-94.

Carter, N., R. Klein and P. Day 1992, *How Organizations Measure Success. The Use of Performance Indicators in Government,* Routledge: London.

De Coning, C. and P. Romita, *rapporteurs,* 2009, "Monitoring and Evaluation of Peace Operations," New York: International Peace Institute.

Diehl, P. 1993, "International Peacekeeping" in Druckman, D. and P.C. Stern 1997, "Evaluating Peacekeeping Missions", *Mershon International Studies Review*, vol. 41, No. 1 pp.151-165. Blackwell Publishing: Oxford

Diehl, P. and D. Druckman (2010) *Evaluating Peace Operations*. Boulder, CO: Lynne Rienner

Druckman, D. and P. C. Stern 1997, "Evaluating Peacekeeping Missions", *Mershon International Studies Review*, vol. 41, No. 1 pp.151-165. Blackwell Publishing: Oxford

Hansen, H.F. 1988, "Program og prosjekteffektivitet. Illustrert ved diskussion af paradigmer for evaluering af moderniseringsprogrammet for den offentlige sektor", *Arbejdspapir nr.1*., København: Center for Offentlig Organisation og Styring.

Harpviken, K.B., A. S. Millard, K. E. Kjellman and B. A. Skåra 2003, "Measures for Mines: Approaches to Impact Assessment in Humanitarian Mine Action", *Third World Quarterly,* Vol 24, No 5, pp 889-908, 2003.

Gartner, G.H. & Ramnarayan 1983, "Organizational Effectiveness: An Alternative Perspective", *The Academy of Management Review*, Vol 8, pp.97-107.

Gowan, R. 2010, "UN peace operations: the case for strategic investment", *Policy Briefing August 2010,* Center for International Peace Operations (ZIF).

Gutner, T. and A. Thompson 2008, "Explaining IO performance: An initial Framework", San Fransisco, California, March 25, 2008 (November version).

Lægreid, T. 2009, "Joint Evaluations and Learning in Complex Emergencies. Lessons from the Humanitarian and Development Domains", Oslo: NUPI.

Meharg, S. J., (ed.), 2009, *Measuring What Matters in Peace Operations and Crisis Management*, Montreal and Kingston: Pearson Peacekeeping Centre.

Mohr, L. B. 1995, *Impact Analysis for program Evaluation*, Second edition, London: Sage

OECD 2010a, "Working Party on Aid Effectiveness – How it works", [online] Development Co-operation Directorate (DCD-DAC). Available at: http://www.oecd.org/document/35/0,3343,en_2649_3236398_43382307_1_1_1_1,00.html>.

OECD 2010b, "Working Party on Aid Effectiveness – Overview", [online] Development Co-operation Directorate (DCD-DAC). Available at: http://www.oecd.org/document/4/0,3343,en_2649_3236398_43414212_1_1_1_1,00.html

Paris, R. and T. D. Sisk 2007, "Managing Contradictions: The Inherent Dilemmas of Postwar Statebuilding", *Research Partnership on Postwar Statebuilding.* New York: International Peace Academy

Power, M. 1997, *The Audit Society: Rituals of Verification,* Oxford: Oxford University Press.

Reitan, T. C. 1993, "Bill.mrk.: Har organisasjon, søker effektivitet. Perspektiver på evaluering og organisatorisk effektivitet", *Norsk statsvitenskapelig tidsskrift*, nr 4: 249-270

Rossi, P. H., Freeman, H. E. and Lipsey, M. W. (1999) *Evaluation: A Systematic Approach (6th edition)*. Sage, Newbury Park, CA.

Ruud, M. and G. Ulfstein 2002, *Innføring i Folkerett*, Oslo: Universitetsforlaget.

Scott, W. R. (1987): *Organizations: Rational, Natural and Open Systems.* Englewood Cliffs: Prentice Hall.

Seashore S.E. & Yutchtman, E. 1967, "Factorial Analysis of organizational performance", *Administrative Science Quaterly*, Vol. 12.

Sending, O. J. 2009, *Why Peacebuilders Fail to Secure Ownership and be Sensitive to Context*, Working Paper 755, Oslo: NUPI

UNICEF / Sandison, P. 2003, "Evaluation report", Global: Desk Review of Real-time Evaluation Experience.

United Nations 2000, "United Nations Peacekeeping Operations. Background Note: 20 March 2000", DPI/1634 Rev.14.

United Nations 2005, “United Nations Peacekeeping Operations. Background Note: 31 March 2005”, DPI/1634/Rev.46 – April 2005.
United Nations 2009, “Approved Resources for Peacekeeping operations for the Period from 1 July 2008 to 30 June 2009”, UN doc A/C.5/63/23 - May 2009.
United Nations 2010, “United Nations Peacekeeping Operations. Background Note: 30 September 2010”, DPI/1634/Rev.113 — October 2010.