NATO
OTAN

Headquarters Supreme Allied Commander

Transformation

Innovation in Operations Assessment

Recent Developments in Measuring Results in Conflict Environments

Edited by:
Andrew Williams
James Bexfield
Fabrizio Fitzgerald Farina
Johannes de Nijs

Innovation in Operations Assessment

Recent Developments in Measuring Results in Conflict Environments

Innovation in Operations Assessment

Recent Developments in Measuring Results in Conflict Environments

Edited by:
Andrew Williams
James Bexfield
Fabrizio Fitzgerald Farina
Johannes de Nijs

Capability Engineering and Innovation Division
Headquarters Supreme Allied Commander Transformation
Norfolk, Virginia
United States

Capability Engineering and Innovation Division
HQ SACT, Suite 100
7857 Blandy Road
Norfolk, VA 23551
United States
http://www.act.nato.int

This book is also available on line at http://www.act.nato.int

ISBN 9789284501915

Published by:
NATO Communications and Information Agency,
Oude Waalsdorperweg 61,
2597 AK The Hague, Netherlands

Contents

Foreword iii
Preface iv
Acknowledgements vi
Contributors vii

PART 1: OPERATIONS ASSESSMENT, CHALLENGES IN THE LIGHT OF HISTORY

Chapter 1 3
The Rationale, Challenges and Opportunities in Operations Assessments

Chapter 2 16
Learning from the Vietnam-Era Strategic Assessment Failure

PART 2: CURRENT ISSUES IN OPERATIONS ASSESSMENT

Chapter 3 39
Operations Assessment at ISAF: Changing Paradigms

Chapter 4 68
Public Opinion Polling in Operations Assessment with Examples from Afghanistan

Chapter 5 86
Time Series Analysis in a Theatre Strategic Headquarter

PART 3: EXTENDING BEYOND THE MILITARY DOMAIN

Chapter 6 117
Civilian and Military Evaluation and Assessment: Synergies and Differences

Chapter 7 145
A Review of Joint Evaluation: Opportunities for NATO?

Chapter 8 169
Engaging Local Actors in Operations Assessment

Chapter 9 185
The Use of Private Industry in Conflict Analysis and Assessment

PART 4: NEW APPROACHES, IDEAS AND CHALLENGES

Chapter 10 196
Complexity Theory: Implications for Measuring Peacebuilding Progress

Chapter 11 221
Operations Assessment and Planning for Transition Stages

Chapter 12 243
Operations Assessment: Decision-Based Evidence Making?

Chapter 13 259
Everyday Peace Indicators: An Alternative Form of Assessment

Foreword

It is with great pleasure that I introduce this inaugural volume of Allied Command Transformation's new Innovation in Capability Development series. The aim of this series is to capture innovative ideas and concepts that support NATO capability development, in a high quality peer-reviewed edited volume made publicly available to encourage wide use and stimulate open debate.

This first volume is on the concepts and processes used to conduct operations assessments, the function that measures the progress of military operations with resulting insights and recommendations supporting planning and operational decisions. With the complexity and duration of current missions and the intangible nature of "success," it is critical to ensure that we continually assess our progress and make necessary adjustments to our actions. We observe through the excellent work in this volume that, while there are many challenges to this endeavour, there are also promising solutions on the horizon.

Headquarters Supreme Allied Commander Transformation has been instrumental in improving operations assessment capability in NATO, including the popular NATO Operations Assessment Handbook, contributions on assessment in the Comprehensive Operations Planning Directive, and developing and executing the first official NATO training course for the NATO School Oberammergau. This volume, using these three products as a base, contains many new concepts and ideas that can help both the researchers and practitioners of assessments in their activities. It captures emerging insights and lessons learned on how to conduct high quality operations assessments developed by both the ACT staff, and other recognised experts.

I believe this volume is a significant contribution to the development of operations assessment in NATO and our Alliance nations, and I look forward to the new thinking that this volume will encourage.

General Jean-Paul Paloméros, French Airforce
Supreme Allied Commander Transformation

Preface

It is an enormous honour to introduce this first volume in the *Innovation in Capability Development* publication series—*Innovation in Operations Assessment: Recent Developments in Measuring Results in Conflict Environments*.

Innovation in capability development is about using existing ideas to do things differently and do them better. While in the business world this notion is commonplace, in the government and international organisation sectors too much reliance on organisational formalities and established ways of working and problem solving often stifles innovation and transformation. A core task of my Capability Engineering and Innovation Division at HQ SACT is to drive innovation in NATO by overcoming organisational constraints, challenging established ways of working and thinking, and taking a long term perspective on what change is possible. There are many tools for accomplishing innovation, including studies, concept development and experimentation, networking with experts from nations, academia and industry, and collaborative online venues such as the Innovation Hub. I'm very pleased to add this edited volume as another vector for innovation in NATO.

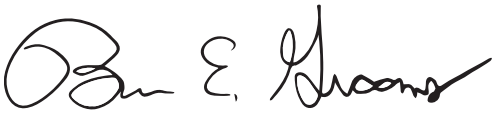
So how does this volume innovate in the world of operations assessment? First, our current thinking is challenged and our assumptions questioned. The articles by de Coning, Mac Ginty and Pennell question the underlying paradigms of our engineering style approaches to measuring progress. They do not reject current methods, but they do cause us to reflect on how our current processes may be improved. The chapters by Schroden, Gaul and Jesse, and Bexfield, highlight specific problems with current military systems of assessment and propose solutions.

Second, the scope of operations assessment is expanded. Traditionally conceived as a military-centric activity, Brusset, Kahlmeyer, Farina, and Muller demonstrate the importance of including a much wider spectrum of actors in the process, from "locals" to private companies to international development agencies. They remind us of the potential value of collaborating with our civilian partners as there is much common ground with the military, despite the differences.

Third, we encourage long term thinking. Many of the ideas presented in this volume cannot be incorporated directly into NATO policy and doctrine at the current time, for a variety of organisational, practical, and political reasons. Yet ideas need time to mature, evolve and be debated, and a public release of these high quality peer reviewed articles is a valuable stepping stone along this process. The chapters by Williams, Bell and Pennell, and Connable remind us of the long arc of development of operations assessment.

We are expanding our horizons in the area of operations assessment. Innovation requires outside input and fresh perspectives to allow us to think about the world in a new light. The best way to achieve this is sharing ideas through a diverse network of experts, and by thinking far outside organisational constraints. I know that my staff and the NATO operations assessment community have benefitted tremendously from this exposure, and I hope that the operations assessment community will benefit likewise.

Vice Admiral Bruce Estes Grooms, United States Navy
Deputy Chief of Staff, Capability Development, Headquarters Supreme Allied Commander Transformation



Acknowledgements

There are a large number of people who made this volume happen. First, we are greatly indebted to the authors who provided their thoughts, ideas, and time at no cost to NATO and were happy to work through multiple revisions and reviews. It has been an honour to work with such a diverse group of experts, and the result of their efforts is a body of work that is truly at the cutting edge of this field. Second, we are very thankful to the external reviewers, in addition to the authors, who spent significant amounts of time carefully scrutinising the draft articles and contributed significant expertise. As a result of this process the work is of the highest quality. Many thanks to Sue Collins, Danielle Fenning, Mehmet Kinaci, Dave LaRivee, Simon Purton, Paul Scott, and David Shaw. Third, we want to thank the NATO Communications and Information Agency for preparing the manuscript, publishing and marketing. Finally, we want to thank all those in the operations assessment community who have directly or indirectly made this project a success and established operations assessment as a permanent fixture in military operations. This book is for you.

The Editors

Contributors

Dr. Katherine Banko is a senior defence scientist from Defence Reseach and Development Canada (DRDC) Centre for Operational Research & Analysis (OR&A) and a team member of the Canadian Joint Operations Command OR&A team. As a behavioral scientist, human dimensions researcher and social science measurement expert, she is responsible for research, analysis, and performance measurement in operational and organisational domains. She develops and incorporates social science research methods and methodologies to advance operations research into traditional operations research approaches and applies these methods to support domestic and international operations. She holds three degrees in social science disciplines (BA Psychology, MEd Educational Psychology, and PhD Psychology) from the University of Alberta, Canada.

Andrew Bell is the Operations Research Scientist for NATO's Maritime Command based in Northwood, London. He has 20 years of experience in military operations research, including maritime patrol and anti-submarine warfare tactics, air campaign planning and assessment, and submarine vulnerability studies. For the last 10 years he has worked for NATO on the planning and assessment of maritime operations, and has been closely involved in the development of the *NATO Operations Assessment Handbook*. His operational experience of operations assessment includes support to air operations in Iraq and Kosovo, two deployments to Headquarters International Security Assistance Force (ISAF) in Kabul, leading the operations assessment of NATO's counter-piracy operation, Operation OCEAN SHIELD, since it started in August 2009, and recently inheriting the responsibility for NATO's maritime counter-terrorism mission in the Mediterranean, Operation ACTIVE ENDEAVOUR.

Jim Bexfield retired in late April 2012 after serving over 10 years as a senior executive in the Office of the Secretary of Defense (OSD). Prior to joining OSD he spent 13 years at the Institute for Defense Analyses

(IDA) and over 23 years in the US Air Force. He was President of the Military Operations Research Society (MORS) from 1983-85, is a Fellow of the Society, and was the 1994 recipient of the Vance R. Wanner Memorial Award. While in OSD he led two major NATO activities: the 2009 Specialist Meeting on *Analytical Tools for Irregular Warfare* and the 2010-11 Specialist Team that produced the report titled *NATO Operations Assessment: A Case Study Based on Planning for Transition in Afghanistan*, April, 2012. He is now consulting for IDA, supporting MORS, and teaching courses in Military Modeling and Simulation in Singapore.

Emery Brusset specialises in performance assessments, with a focus on operations in fast changing and sensitive environments. A graduate of Yale University (BA) and the London School of Economics (MPhil), he routinely leads multi-disciplinary teams on the ground. After a brief career in UN humanitarian missions, over the past seventeen years he has carried out studies covering conflict prevention, human rights, rule of law, and community development. He also advises mining and oil companies in their social investment and outreach activities, with a focus on risk management and capacity development. He frequently facilitates training in assessment and planning methodologies, and has assisted NATO in methods of operations assessment. Brusset is a French national and currently resides in Belgium, outside Brussels.

Ben Connable is a Senior International Policy Analyst at RAND where he works on intelligence, irregular warfare, terrorism, and regional issues. Ben is a retired Marine Corps intelligence and foreign area officer with specialisation in the Middle East. Since coming to RAND he has written extensively on counterinsurgency and specifically on the counterinsurgency assessment process. In 2012 Ben authored *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*, an examination of U.S. assessment in Vietnam, Iraq, and Afghanistan. He also co-authored *Assessing Freedom of Movement for Counterinsurgency Campaigns*, as well as a range of project reports on assessment for a range of government sponsors. Ben holds an MA in National Security

Affairs, an MA in Strategic Intelligence, and is ABD at King's College London War Studies Department.

Dr. Cedric de Coning heads the Peace Operations and Peacebuilding Research Group at the Norwegian Institute of International Affairs (NUPI), and he is a Peacekeeping and Peacebuilding Advisor at the African Centre for the Constructive Resolution of Disputes (ACCORD). His chapter is based on his PhD dissertation in Applied Ethics with the Department of Philosophy at Stellenbosch University, South Africa. The dissertation is entitled "Complexity, Peacebuilding & Coherence: Implications of Complexity for the Peacebuilding Coherence Dilemma" and can be accessed at: http://hdl.handle.net/10019.1/71891

Johannes de Nijs is Branch Head Operational Analysis at Headquarters Supreme Allied Commander Transformation (HQ SACT) in Norfolk, U.S. Mr. de Nijs received a master's degree in Mathematics from the University of Leiden and a bachelor's degree in Political Science from the University of Amsterdam. Mr. de Nijs has been with NATO since 1982 when he joined the Supreme Headquarters Allied Power Europe (SHAPE) Technical Centre in The Hague, transferred to NATO Headquarters, Brussels, in 1988, and joined Supreme Allied Command Atlantic (SACLANT), Norfolk, in 1991. With the establishment of Allied Command Transformation (ACT) in 2003, Mr. de Nijs became Branch Head Operational Analysis. Under his responsibility, analysis was conducted for the Multinational Experiment series, in experiments for SHAPE's Comprehensive Operations Planning and Knowledge Development, and for concepts and experiments on Maritime Situational Awareness, Operations Assessment, Alternative Analysis, Hybrid Threats, Civil Military Interaction, Logistics Chain Management, Space, Ballistic Missile Defence, Cyber Security, and Counter-IED amongst others. He is the principle member for ACT in the NATO Science and Technology Systems Analysis and Studies Panel and the ACT Focal Point for Alternative Analysis.

Fabrizio Fitzgerald Farina is a consultant in the field of international development and security. After studying in the U.K. and U.S., he received his Master's degree in International Relations – Economics and Policies for Development Magna Cum Laude at the University of Bologna, Italy. Farina worked in the field of international security focusing on genocide prevention and threat-convergence issues at The Fund for Peace, in Washington D.C. He also worked as Project Manager for a local NGO in Ghana on a project to upgrade a slum area of Accra, and as international volunteer in Benin working as fundraiser and assisting senior management on projects to curb malaria and prevent HIV. After consulting for the Operational Analysis branch at NATO HQ SACT on the production of this volume, he now works as Project Coordinator for an international communication agency in Europe.

Marcus Gaul is the branch chief for Analysis and Data in the Afghan Assessment Group (AAG) at the International Security Assistance Force's (ISAF) Headquarters in Kabul, Afghanistan. He is a theoretical physicist with an extensive background in mathematical analysis. Marcus has been with ISAF since 2010 supporting decision-makers with objective analysis and regular briefings to 4 star commanders. Before joining ISAF, he was a Senior Scientist and Systems Analyst at IABG (2004-2010),where he was a project manager and conducted scientific studies in knowledge development and operations research for the German Federal Armed Forces and NATO, which resulted in his deployment to Headquarters, Kosovo Force (KFOR) in Pristina, Kosovo (2007). In his academic career, Marcus taught and worked at several physics institutes, including the Max-Planck-Institut für Physik in Munich, Germany, the Centre National de la Recherche Scientifique's (CNRS) Centre de Physique Théorique in Marseille, France, and the University of Pittsburgh, United States (1997-2001). Marcus studied Physics, Mathematics and Astronomy, and received his Diploma in Physics from the University of Heidelberg (1997). He is fluent in German, English and French.

Brittany Heal is a research assistant at Channel Research, where she has been working on the European Commission's humanitarian programmes

and on the evaluation of Norwegian support to the civilian aspects of African peace-keeping. She is a British national and a graduate of Latin America studies and History at the University of Edinburgh. She has travelled extensively in Latin America and Central Africa.

Dr. Eric Jesse serves as a Captain in the United States Air Force. He earned his doctorate in policy analysis at the Pardee RAND Graduate School in 2011. He completed his undergraduate degree at the United States Air Force Academy were he double majored in Operations Research and Economics. He was the deputy branch chief of the ISAF Headquarters Afghan Assessments Group Data and Analysis Branch in 2012/13, deployed from Headquarters United States Air Force Europe Analysis and Assessments. Some of his notable previous work includes cocaine market modeling for the U.S. Office of National Drug Control Policy, research on U.S. India policy for Ambassador Robert Blackwill, assessing vehicle fleet utilisation for the Civil Air Patrol, strategic forecasting of Iraq and Afghanistan for the US Army, and developing low observable aircraft utilisation models for the Pentagon.

Andre Kahlmeyer is a consultant specialising in the design, monitoring and evaluation of interventions in fragile, conflict and post-conflict environments. In this function, Andre has been responsible for numerous civilian fact-finding, design and evaluation missions to countries such as Afghanistan, Iraq, Yemen, Nepal, Nigeria, Democratic Republic of Congo, the Palestinian territory, Rwanda, Bosnia Herzegovina, Timor-Leste and Myanmar. He has worked as a consultant for major donors such as the U.K.'s Department for International Development (DFID) and the United States Agency for International Development (USAID), and for international NGOs such as International Alert. Before becoming a consultant, Andre was employed by Channel Research, the Parliamentary Assembly of NATO, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and the EU Common Security and Defence Policy (EU CSDP) mission in Iraq. Andre holds an MA degree in Middle Eastern/Arabic studies, History and Political Science.

Megan Grace Kennedy-Chouane is a development professional with expertise in evaluation and peace studies. She currently works as a policy analyst at the Organisation for Economic Cooperation and Development (OECD), supporting the Development Assistance Committee (DAC) Network on Development Evaluation. She carried out a major comparative analysis of development agency evaluation systems in 2010 and led the process of developing guidance on evaluating in situations of conflict and fragility from 2009-2012. Prior to working at the OECD, Ms. Kennedy-Chouane was a Thomas J. Watson fellow, carrying out independent research on strategies for engaging youth in community development projects in Africa and Latin America. She holds a Master's degree in Public Administration and International Management from the Monterey Institute of International Studies, a post-graduate diploma in Children, Youth and Development Studies from the Institute of Social Studies, and a dual Bachelor's degree in Economics and in Peace & Global Studies from Earlham College. Ms. Kennedy-Chouane is a U.S. national and currently resides in France.

Dr. Roger Mac Ginty is Professor of Peace and Conflict Studies at the Humanitarian and Conflict Response Institute, and the Department of Politics, at the University of Manchester. His main research interest is in the interaction between international intervention and local communities. His latest books are *International Peacebuilding and Local Resistance: Hybrid forms of peace* (2012) and the edited volume *Handbook of Peacebuilding* (2013).

Dr. Marcey Moss is a Senior Research Scientist at the Center for Evaluation and Education Policy (CEEP) at Indiana University. Dr. Moss has a long history of successfully working with diverse stakeholders and engaging in collaborative processes as the project director for numerous large-scale evaluations. Dr. Moss holds a Masters of Science in Rural Sociology and a Ph.D. in Sociology from South Dakota State University with an emphasis in Demography and Human Ecology.

Dr. Patricia Muller is Executive Associate Director and Director of Evaluation and Research at the Center for Evaluation and Education Policy (CEEP) at Indiana University. She has served as the Principal Investigator and Project Director for more than $10 million in funded research and evaluation at the international, national, regional levels. She is nationally recognised for her application of social science research methodology to critical research and evaluation studies, using both advanced statistical analyses and in-depth qualitative case study methodology. Her leadership of international and national work includes developing a theoretical framework for joint evaluation for the NATO. She has a Ph.D. in Higher Education, with a specialisation in inquiry/research methodology, from Indiana University; and a Masters of Science in College Student Personnel Administration from Indiana University.

LtCol (Ret'd) Bruce Pennell is a Senior Scientist working as an Operational Analyst in the NATO Communications and Information Agency based in The Hague, Netherlands. He completed his 23 year military career as a member of faculty at the UK Advanced Command and Staff Course, having seen operational experience in both Gulf Wars, in the Balkans and Northern Ireland. Since joining NATO as a civilian in 2009 he has worked extensively on strategic operations assessment for Afghanistan with SHAPE and ISAF, including multiple theatre deployments.

Dr. Anne Maree Ruddy is a Senior Research Scientist at the Center for Evaluation and Education Policy (CEEP) at Indiana University. Dr. Ruddy has extensive experience developing and implementing large-scale national and international evaluation and research projects across a variety of content areas using diverse methodologies and a comprehensive approach. Dr. Ruddy holds degrees from Edith Cowan University (Perth, Western Australia) B.Ed., The University of Western Australia (Master of Educational Administration) with an emphasis in Educational Policy and a Ph.D. from Murdoch University (Perth, Western Australia) with an emphasis in International and Comparative Educational Policy.

Dr. Jonathan Schroden is the Director of CNA's Center for Stability and Development, whose mission is to support decision-makers charged with planning, conducting, and assessing stability and development operations with objective analysis grounded in an understanding of actual operations. Schroden has been with CNA since 2003, during which time he has deployed multiple times each to Afghanistan and Iraq, travelled in the Middle East, gotten underway with numerous U.S. Navy ships, and supported disaster relief operations in the wake of Hurricane Katrina. He has served as the CNA field representative to the 2nd Marine Expeditionary Force (II MEF), Multi-National Force–West (MNF-W) in Iraq, U.S. Central Command (CENTCOM), and the International Security Assistance Force (ISAF) in Afghanistan. Schroden holds Ph.D. and M.S. degrees from Cornell University and B.S. degrees from the University of Minnesota–Duluth.

Andrew Williams is a section head in the Operational Analysis Branch in Headquarters Supreme Allied Commander Transformation. He is one of the lead developers of NATO guidance on operations assessment, including the *NATO Operations Assessment Handbook* and the formal NATO training course. He has conducted a variety of analytical studies on operations assessment, civil-military interaction, organisational collaboration, private military security companies, and concept development and experiment techniques. Prior to working at NATO, he worked as an operations analyst for the United Kingdom's Defence Science and Technology Laboratory. He holds a Masters of Physics from the University of Manchester, and is a Ph.D. candidate in public policy and administration at Old Dominion University. He is a British national and currently resides in the United States.

PART 1:

OPERATIONS ASSESSMENT, CHALLENGES IN THE LIGHT OF HISTORY

1

The Rationale, Challenges and Opportunities in Operations Assessments

Andrew Williams, James Bexfield, Andrew Bell, Bruce Pennell

Introduction

Complex crises, conflicts and disasters are characterised by dynamic and unpredictable events: political transitions, power vacuums, violent insurgencies, terrorism, crime, insecurity, ethnic tensions and conflict, government collapse, economic breakdowns, corruption, disease, and major social upheaval. Responses may include military intervention, economic aid and activities by development actors each with potentially overlapping or conflicting aims. In a few cases the criteria for success are clear and undisputed with all stakeholders agreeing on how to measure the rate of progress. More frequently, the method used to measure progress is contentious and results are disputed. While there is consensus that an assessment and feedback process is necessary, there is much less agreement on what that process should look like. However, there is general agreement that operations assessments should not replace a commander's intuition and staff experience, but rather they should provide evidence that helps inform decisions.

In NATO the process of measuring progress is known as "operations assessment." Its basic assumptions and logic are founded in the same systems thinking that underpins military planning for intervention and the development of situational awareness. Identifying the system elements of a crisis situation—actors, organisations, environment, relationships, context and intentions—and conducting evidence-based measurements of how the elements changed during an intervention (including why they changed) will inform decision makers on the impact of their intervention and help them identify potential adjustments.

Operations assessment is defined by NATO as "the function that enables the measurement of progress and results of operations in a military context, and the subsequent development of conclusions and recommendations that support

decision making" (NATO 2011, 1-1). It is broadly equivalent to "monitoring and evaluation" used by many civilian government agencies and international organisations (Williams and Morris 2009). While the focus of operations assessment is on supporting military plan development and refinement, in an intervention involving civil, military, political, diplomatic, cultural and economic factors, the assessment process should have a broader perspective.

It is critical to observe, however, that our frameworks, theories, knowledge and understanding that form the basis for intervention planning and progress assessment are provisional. Complete knowledge of a crisis environment can never be achieved. Military planners seldom know whether progress is due to their actions, some other factor, or purely a chance occurrence that may be quickly reversed. Moreover, crisis situations often feature intelligent and dynamic adversaries who exploit unexpected opportunities. The complexity of crisis environments makes the task of measuring progress particularly ambitious.

In the past decade NATO operations assessment has undergone substantial development, driven partly by the experiences of the international missions in Iraq, Afghanistan and Libya. This volume reviews recent practices, and suggests ways the process could be improved, including potential future innovations and capability developments. This introductory chapter has three aims. First, we review the rationale for conducting operations assessments. Second, we describe the various challenges of measuring progress in complex mission environments. Finally, we briefly describe the remainder of this volume in which a wide range of authors bring their experience, insight and creativity to some of these challenges.

The Rationale for Operations Assessment

The idea of measuring progress against defined objectives has a long history in the corporate, government and non-profit sectors. Indeed, objective evidence-based performance appraisal of government or corporate actions or interventions is expected in modern democratic states with market economies. While there is a strong body of evidence that justifies the practice in these sectors (Rossi, Lipsey, and Freeman 2003), it is worth examining the benefits expected from operations assessment in the military context.

Improving military planning

Military actions are guided by operational plans, which are based on higher-level political guidance, and underpinned by a number of assumptions about cause and effect. The ever-changing environment in which military operations take place means that continuous monitoring and adjustment of the plan is required, either to keep it on track, or to recognise when a situation has changed sufficiently to require plan adjustments. In short, every commander will need to answer three questions: where are we? How did we get here? And what's next? Operations assessment can help answer all three questions.

In summary, a well-designed operations assessment can improve the fidelity and quality of military plans. The act of describing planned objectives in measurable terms should give planners a deeper understanding of systems elements affected by the proposed intervention. It may also reveal interrelationships in the operational environment. Finally, it helps prevent "planning by headlines," in which a military plan contains sweeping statements about desired changes by requiring planners to describe abstract goals (e.g. "security") in specific and measurable terms.

Testing assumptions

In the process of planning, many assumptions are made about causal linkages between actions and objectives; operations assessment helps confirm the validity of these assumptions. This is particularly important in counter-insurgency or stabilisation environments, which are complex, unpredictable, and susceptible to plans quickly going off track (Stolk et al. 2011).

Improving decision making

Operations assessment provides information to a commander to support evidence-based decision making about overall strategy, operational planning, and allocation of resources to missions. The assessment process assesses whether objectives are met and whether tasks have been conducted as planned (timeline and resources). Consequently, operations assessment can support a wide range of decisions: is the force large enough? Is it in the right place doing the right things? Are new or enhanced capabilities required? Should the plan be modified? Can the strategy be accomplished?

Structuring and developing knowledge

An operations assessment uses a structured method to gather and collate evidence over the duration of the operation. The data collected can help people understand the operational environment and how military actions contributed to the success or failure of a mission. It preserves an institutional memory that can be used to learn from each other's experiences. The knowledge gained is also available for subsequent analyses that provide historical insights or test new concepts or capability developments.

Supporting the information campaign

Information operations use public diplomacy and, national and international media to influence the perceptions and actions taken by the adversary and other constituencies. Operations assessment can provide credible evidence that can be used to support the commander's information operations campaign, and help identify actions to counter adverse media or information campaign used by the enemy.

Promotes a comprehensive approach

Operations assessment is a key enabler in NATO's implementation of comprehensive approach in current and future operations. The structured process enables the military to keep track of and share information with other non-military actors operating in the theatre. The result is a better understanding of the interconnections and interdependencies between military and non-military activity, and how they should be coordinated and synchronised.

> NATO's Strategic Concept states that:
> The lessons learned from NATO operations, in particular in Afghanistan and the Western Balkans, make it clear that a comprehensive political, civilian and military approach is necessary for effective crisis management. The Alliance will engage actively with other international actors before, during and after crises to encourage collaborative analysis, planning and conduct of activities on the ground, in order to maximise coherence and effectiveness of the overall international effort. (NATO 2010, paragraph 21)

Anticipating and forecasting

Operations assessment can, in some limited cases, extrapolate past data to forecast future states and suggest actions needed to achieve them. This supports planning decisions especially in complex, political organisations where decision making takes time. Anticipatory thinking can ensure that resources and capabilities are available when needed.

The Challenges in Operations Assessment

An underlying theme of this volume is that international interventions are characterised by a complex constellation of actors, organisations, priorities and pressures, which makes it difficult to conduct operations assessments that enable senior leaders and policy makers to answer the question: how are we progressing? In order for operations assessments to be relevant and useful in this context, they should meet three equally important but conflicting aims: 1) they must be adaptive to the operational tempo and ever-changing context of military operations; 2) they must make sense to the organisational configuration, process, and leadership and 3) they must be empirically supported and, ideally, objective. Yet there are a number of challenges that underlie and exacerbate the tensions between these three aims. The goal of this volume is to help mitigate some of the following challenges.

Institutional factors

There are several institutional challenges, the first of which lies in the bureaucratic structure of military organisations. Within NATO and other military settings, operations assessment is usually conducted in a command structure with multiple reporting lines and key stakeholders distributed across many organisations. Operations assessment is primarily conceived as a tool to support a commander's decision making. Yet assessment products also drive decision making at subordinate command levels, and are sometimes used by higher level commands. An operations assessment can have multiple stakeholders at various levels, some of which have agendas that may not align with the needs of the commander. Finding a balance between measuring mission progress and meeting the various demands of other stakeholders, particularly at higher HQs, is a key challenge.

The second institutional challenge is the tension between the political imperative to demonstrate progress and the generally decades-long timescales over which change occurs in some complex operations such as counterinsurgencies. National political and budgetary agendas generally run on four or five year timescales and the political narrative in the past have advanced the notion of quick and decisive military action, thus an alignment of political and military strategy to relatively short-term timescales is inevitable (Their 2009). In results-driven organisations there will always be subtle pressures to demonstrate success. Consequently, higher level assessments of progress may be framed so as to favour a more positive view than the reality on the ground.

The third institutional challenge concerns the way "progress" is politically defined and high level objectives are left intentionally vague. Operations assessments are set in the context of a political decision making process. In NATO each member must consider their own domestic political situations in addition to the political situation within NATO as a group. In military interventions there is rarely an objective reality to be empirically discovered, with high level goals such as "change," "progress," "freedom" or "security" treated as relative concepts politically constructed by various stakeholders. As a result, evidence is often contested depending on how issues are framed (Fisher and Forester 1993; Stone 2003). This can make it difficult for operations assessment and analytical personnel who strive to approach the task from a strictly objective and scientific perspective.

Requirement for comprehensive assessment

A comprehensive assessment of progress should include the entire range of factors in the operational environment, including many factors outside the security-domain. While military forces can report on the level of violence or security due to their privileged access to data and sensors, it is hard to make causal assessments on why the level of violence changed without a deeper knowledge of changes in other domains. For example, there is a known dependency between the level of violence and the extent of international aid and development in an area (Myerson 2011; Zyck 2012), thus it is critical for military-focused operations assessment to consider all domains in order to understand key relationships.

"Over-metricising"

There is a tendency for assessment staffs to overstate the number of metrics needed, thus generating a huge data collection requirement on subordinate organisations. A major challenge is for the assessors to understand "how much is enough" in terms of metrics and to appreciate the significant amounts of work needed to collect the data. The assessors should also give the data collectors an understanding and appreciation the value of their contributions (Downes-Martin 2011).

Assessment in multi-actor environments

In an intervention there are often multiple local, national and international organisations undertaking a range of concurrent activities with different underlying objectives and goals over different time horizons (Stolk et al. 2011). Each organisation has their own definitions of success and measures of progress. Yet often the success of one actor's operation depends on the successes of other actors' operations. To be truly comprehensive, assessments should be made both with inputs from, and consideration of, all salient stakeholders in the operational theatre. NATO is currently planning to include civilian staff with relevant civil-domain expertise in crisis operations in their operations assessment process. Non-military stakeholders may "frame" the intervention objectives in different terms than the military, which can result in multiple perspectives that are difficult to reconcile.

Data collection

There are many practical difficulties in data collection. Data collectors may face security threats resulting in limited or no access to key data sources. Data collection by military-affiliated personnel may unduly influence responses by locals. Ensuring consistency in how data is collected over time can be difficult. Assessing attitudinal change in populations, especially in conflict areas, can be particularly challenging. Opinion polling or other survey methods can help monitor such attitudes, but they are difficult to conduct properly. Institutional pressures and the challenges of war may also incentivise data collectors or operators to reduce quality controls, or in the worst cases manipulate data.

Establishing reliable baselines

Many military operations are complex and turbulent and it is often difficult to establish reliable baselines or to distinguish between strategic shifts in the environment and shorter-term fluctuations (Stolk et al. 2011). Sometimes assessment designs and data collection systems are not developed in advance of an operation, resulting in evolving data collection and assessment approaches. Even if there is the capacity within a command chain to handle a large number of metrics, setting appropriate targets for those metrics is far from simplistic. Defining an acceptable condition to be reached by a metric requires an understanding of what is locally and contextually "normal." An understanding of what is "normality" may change with time or as the underlying premises and assumptions of missions change. Furthermore, the focus of missions may change over time, meaning the importance of metrics vary over time. Some commands may stop collecting data for some metrics, only to realise months or years later that the lost data would have been useful for current priorities.

Methodological challenges

There are several key methodological challenges that face operations assessment policy makers and practitioners. First, there is the difficulty in deciding when metrics are best measured using quantitative or qualitative techniques. Second, there is the challenge of data quality and reliability—whether quantitative or qualitative. Data collected can be either partially or completely inconsistent with the metric being used and establishing proper statistical protocols is often not possible.

Third, there is the challenge of determining the appropriate level of aggregation. Goals are often described as an aggregate of lower level effects to be achieved, each of which is associated with one or more metrics. The aggregation of the subordinate metrics into a higher level metric is far from trivial. The aggregation of data into a composite measure may conceal some important features of the situation. Nevertheless the challenge of aggregation remains, as it is a natural human tendency to seek simplified explanations for complex situations. This is one of the widest criticisms of recent operations assessment practice, as explained by Connable and Schroden in this volume.

Use of assessments by decision makers

Finally, there is the basic problem of how evidence is used by decision makers. Assessments are just one out of a multitude of inputs received by a commander. Social science literature offers a range of explanations on how policy makers use data and analysis to support their decisions (Stone 2003), and the varied uses of information (Weiss 1979). Furthermore, it is rare that decision situations are neatly structured and ordered such that commanders can make decisions based *only* on an assessment, even if additional context is provided with that assessment (Weiss 1980).

Bridging the Gap: An Outline of the Volume

Part I introduces the volume and provides a historical perspective. In this first chapter we have introduced operations assessment as the function that enables the measurement of progress and results of operations in a military context, and the subsequent development of conclusions and recommendations that support decision making. We then identified several ways operations assessments are used, including to help produce, test, and modify military plans, to structure, develop, and share knowledge, and to support strategic, tactical, and operational decision making. Next several challenges associated with conducting operations assessment were discussed, including institutional impediments, difficulties encountered when working in multi-actor environments, the problems encountered in consistently collecting static and temporal data, and methodological issues associated with qualitative versus quantitative metrics and metric aggregation.

The second chapter takes us back in time to the assessment processes used in the Vietnam War. It shows how the assessment methods used in this war were derivatives of those used in Korea and World War II. It goes on to demonstrate why many of the metrics used were inappropriate for the Vietnam counterinsurgency (COIN) environment and, at times, were misleading and data collection burden sometime resulted in unnecessary casualties. It concludes with some suggestions on preferred metrics and practices for current conflicts.

Part 2 contains three chapters that discuss practices and identify issues associated with operations assessments supporting current missions (primarily Afghanistan). Chapter 3 describes the assessment used at Headquarters International Security Force-Afghanistan (HQ ISAF) in 2011-2012. It explains the needs expressed by the commander, the role of the subordinate commands

and staff elements in providing a wide variety of quantitative and qualitative inputs, the insightful ways the metrics and associated rationale were displayed, and how the assessment was used in the command's decision making process.

An important part of COIN operations is understanding how our activities (military and civil) are viewed by the indigenous population. Opinion polling is the approach most often used to gain this understanding. Chapter 4 provides guidance on the proper way to design the questionnaire, identify the sample, conduct the interview to minimize respondent errors and interpret the results. It stresses the importance of collecting the right information in a consistent manner and discusses, with mitigating suggestions to counter potential social and non-response biases.

The problem of how to gain insights from data collected over time (time series analysis) is the subject of Chapter 5. The authors with recent experience in HQ ISAF show how time series analysis has been used to identify unexpected relationships that resulted in identifying opportunities exploited by the command. The chapter, one of the more technical ones in the volume, describes the data sets and the analysis approaches used, including year-over year differences, seasonal decomposition, regression analysis, and the autoregressive, integrated moving average model.

Part 3 contains four chapters full of ideas on how the military can work with other non-military organisations, both inside and outside government, to better understand today's conflict environments. Chapter 6 contains an overview of monitoring and evaluation systems used by civilian organisations to evaluate progress and compares them to the military assessment process. It concludes that both use similar terminology and logical frameworks, but differ in other important ways, including the accountability focus of civilian organisations vice the internal decision making focus of the military. In addition, civilian evaluations tend to be more "independent" as the evaluators and performers are usually from different organisations.

Chapter 7 suggests that NATO could benefit from conducting joint evaluations with civilian organisations. It begins by defining joint evaluations and then discusses several approaches, with examples, that have been used in the past. It continues by showing that joint evaluations are generally of higher quality than single organisation evaluations and the lists the benefits and challenges

associated with conducting joint evaluations. It concludes by suggesting two steps NATO could take if they decide to engage in joint evaluations.

Military operations assessments are usually planned internally involving only those that directly work for the military, and executed so as to minimize the involvement of those outside the military. Chapter 8 espouses the value of including local actors in the assessment process and suggests several ways to accomplish it. It discusses the importance of building mutual respect and trust, of planning with the locals so as to achieve local ownership and of the issues with sharing potentially classified data. It also identifies risks associated with threatening the local culture, endangering the local that work with the military, and the danger of putting military personnel in potentially unsafe areas. It concludes by describing a methodology that could be used as a template for including local actors in operations assessments.

Chapter 9, the last chapter in this part about extending assessments beyond the military, discusses the pros and cons of using private industry resources to support operations assessments. The authors suggest that much can be gained by the added expertise, the easier access to the locals, and the more timely availability of qualified personnel that industry can offer. A brief description of the private sector assessment community is followed by a discussion of the challenges of getting the resources needed and contracting with industry, and the sharing of classified information.

The final Part (Part 4) contains four chapters with new ideas and approaches for advancing the state of operations assessment. Chapter 10 defines a complex system as one that is governed by non-linear relationships with the ability to adapt, to self-organise, with individual actions being the result of the information available and interactions/feedback with others. The author suggests that using complexity theory to represent the peace building process can lead to a better understanding of conflict analysis, better plans, management and coordination of actions, and result in better metrics and assessments. It stresses the importance of attempting to foresee unintended consequences of international activities and suggests that a key challenge associated with COIN operations is how to achieve the appropriate balance between international support and local self-organisation.

Chapter 11 addresses the use of operations assessments to support the planning and execution of transition operations that occur when NATO is

disengaging from a conflict area. It was largely based on the transition planning that started in the Summer of 2010 in Afghanistan. The chapter includes sections on developing security, governance, rule of law, and economic development metric, a data sharing initiative called "DataCards" that is currently supported by the National Defense University in the U.S., and a best practices guide for conducting operations assessment along with planning and execution advice, much of which is applicable to everyone involved with assessments.

Chapter 12 explores how evidence (often gained via operations assessments) may be used to influence decisions. The author identifies 3 common modes ranging from generating evidence to support decisions already made to the evidence playing a major role in the decision. It discusses the concepts of value curves and cognitive bias and concludes with several thought-proving propositions one of which is "overcoming strongly held views rarely depends on more evidence." Finally, the chapter stresses the importance of incorporating the political domain in operations assessments.

The last chapter in the volume addresses the issue of how we measure "peace"—the absence of conflict in an area. The author argues that our current measures are inadequate, linked to specific projects, are often based on national statistics, use metrics that are misleading and have a top-down focus. Instead, he argues, the focus should be at the local level by considering what the locals perceive and any concerns/fears they have. Through this understanding NATO commanders may make very different decisions.

As the reader can observe, this volume covers a vast array of topics and issues. Each chapter will be provocative in its own way. The editors encourage the reader to question the insights, come up with their own understanding of the issues and, most importantly, generate their own ideas on how to improve NATO operations assessments.

References

Downes-Martin, S. 2011. Operations assessment in Afghanistan is broken. *Naval War College Review* 64(4):103-125.

Fisher, F., and J. Forester. 1993. *The Argumentative Turn in Policy Analysis and Planning*. Durham, NC: Duke University Press.

Myerson, R. 2011. Rethinking the fundamentals of state-building. *Prism* 2(2):91-100.

North Atlantic Treaty Organisation (NATO). 2010. Strategic Concept for the Defence and Security of the Members of the North Atlantic Treaty Organisation.

——— 2011. *NATO Operations Assessment Handbook*, Interim Version 1.0. Norfolk, VA: HQ Supreme Allied Commander Transformation.

Rossi, P. H., M. W. Lipsey, and H. E. Freeman. 2003. *Evaluation: A systematic approach*. SAGE Publications.

Stolk, C. van, T. Ling, A. Reding, and M. Bassford. 2011. *Monitoring and Evaluation in Stabilisation Interventions: Reviewing the State of the Art and Suggesting Ways Forward*. Santa Monica, CA: RAND Corporation. Available at: <http://www.rand.org/pubs/technical_reports/TR962>

Stone, D. 2002. *Policy Paradox: The Art of Political Decision-making*. New York, NY: Norton.

Thier, J. A. 2009. *The Future of Afghanistan*. Washington, DC: United States Institute of Peace.

Weiss, C. H. 1979. The many meanings of research utilisation. *Public Administration Review* 39(5):426-431.

——— 1980. Knowledge creep and decision accretion. *Science Communication* 1(3):381-404.

Williams, A. P., and J. C. Morris. 2009. Theory-based evaluation in the military: Theory on the front-line. *American Journal of Evaluation* 30(1):62-79.

Zyck, S. A. 2012. How to lose allies and finance your enemies: the economisation of conflict termination in Afghanistan. *Conflict, Security & Development* 12(3):249-271.

2

Learning from the Vietnam-Era Strategic Assessment Failure

Ben Connable

Abstract

This chapter describes and evaluates the Vietnam-era assessment process, including comments on the proper role of quantified data when performing assessments in on-going or prospective non-conventional wars. It will show how the frequent lack of clear strategic objectives in non-conventional wars—including Vietnam—resulted in both military officers and politicians seeking quantitative justification for their activities. This massive use of quantitative data for strategic assessments of the Vietnam War not only generated unrealistic data collection requirements, it also produced inaccurate and misleading findings that ultimately did not provide effective decision support. In summary, the Vietnam War provides an excellent opportunity to better understand the strengths and weaknesses of the U.S. campaign assessment processes.

Introduction

The U.S. assessment methodology implemented during the Vietnam War provides the most comprehensive and detailed example of theatre-level assessment in modern warfare through the mid-2000s, and also the best case study available with which to place the assessment efforts in Iraq and Afghanistan in useful context. The United States exerted tremendous effort and devoted what were, up to that period, unprecedented assets towards reporting and analysing battlefield actions and environmental conditions in Vietnam. Ground-up reporting fed into a complex web of semi-automated systems that, despite their relative immaturity in relation to today's networked environment, still managed to generate tens of millions of pages of data and charts. These data fed assessment reports at the Military Assistance Command Vietnam (MACV) in Saigon, and at the U.S. Office of the Secretary of Defense (OSD). Because actions and environmental feedback were recorded in detail down to the level

of the individual infantry squad and Vietnamese hamlet, and because nearly all of the reports have been declassified and are readily available online, the Vietnam War assessment process is nearly transparent to modern researchers. This transparency enables an evaluation of the assessment process used in Vietnam, much of which was an application of methods used in WWII and Korea.

In general, this evaluation reveals a grievously flawed assessment concept that spawned what could only be described, at least in retrospect, as the needless waste of both money and lives. Failure of assessment in Vietnam stemmed directly from the misapplication of WWII and Korean-era methods, and also from the lack of clear strategic objectives necessary to anchor productive strategic assessment.

Assessment reports describing the Vietnam War were generally inaccurate, irrelevant to specific military and political decisions, and often misleading.[1] Many blame Secretary of Defense Robert S. McNamara for implementing a heavily quantitative systems analysis approach to assessment that sought to produce tangible measurement while ignoring many of the critical intangibles that were truly relevant to U.S. political-military strategy. MACV and OSD analysed body counts and hamlet pacification scores, but completely misjudged the enemy's will to continue the fight and the population's desire to simply live in peace under the control or influence of either side. Secretary McNamara does deserve much of the blame assigned by historians like McMaster (1997)—McNamara has publicly acknowledged his own culpability in the Vietnam assessment failure—but the predilection to see victory through the measurement of tangible, short-term progress rather than more meaningful, often less tangible and longer-term accomplishments existed at all levels of command. Reliance on oversimplified quantitative measurement and the conflation of observable short-term gains with long-term success still exists today in some military operations, undermining the quality and usefulness of assessments. The failure of the assessment process in Vietnam is a glaring and enduring cautionary tale.

1. Observations and findings in this chapter draw extensively from previous research I conducted for RAND from 2009 through 2012 on counterinsurgency assessment, see Connable 2012.

Background

By the mid-1960s, the idea of detailed, quantitative strategic assessment of war was far from novel. Analysts had pored over reams of data from both World Wars, enumerating nearly all aspects of combat, logistics, and forces at the operational and strategic levels (e.g. Shrader 2006a; Shrader 2006b). In WWII, analysts tracked military equipment production rates, industrial capacity, strategic oil availability, destructive capacity of massed bomber forces, and other data that had relevance to allied operations and strategy. Allied victory in WWII is often described as a triumph of overwhelming production and logistics (Overy 1997), or in other cases as a clear example of the success of attrition warfare against the German armed forces and the Nazi industrial base.

Robert McNamara analysed allied bomber efficiency during WWII as part of an overall Army Air Forces analytic team that by the end of the war numbered nearly 400 men and women (Watson 2003).[2] The commanding general of 9th Bomber Command was impressed by these operations analysts, whose mathematical, statistical, and systems-based approaches formed the basis for some assessments in WWII and also for many of the strategic assessments in Vietnam, Iraq, and Afghanistan:

> These people attack problems by rational methods with a detached viewpoint which apparently makes it possible for them to discover solutions which tactical personnel have overlooked or have not had the necessary uninterrupted time to develop. (U.S. Army Air Forces 1948, 1)

Quantitative analysis reached down to the tactical level, but quantitative assessment was primarily an operational or strategic effort. Tactical battle damage assessments of specific targets were viewed through a strategic lens. Operations analysts could make a logical and quantifiable case as to how the destruction of the ball bearing plant at Schweinfurt, Germany or the oil refinery at Ploesti, Romania—both subjected to massive and costly allied air raids—could set back the entire German war effort and speed the end of the

2. McNamara was a Statistical Control Officer with the Army Air Forces. For example, Army operations analysts played a crucial role in designing the D-Day Invasion of Normandy, France, and Navy operations analysts developed and refined search and detection techniques against submarines.

war. The systems analysis approach to assessment (e.g., U.S. Army 2010) had defensible merit in the milieu of a global conventional war, a war that had a clear strategic end state: the total military defeat of the Axis powers and their unequivocal surrender.

Quantitative analysis—the process of improving operations through quantitative problem solving and modelling—became central to the U.S. military's operations support and also to assessments. Experience in the Korean War showed that the tendency to seek quantitative clarity from assessments became more prevalent in the absence of clear strategic end state conditions. As General Douglas MacArthur drove his combined U.N. forces from the Pusan Perimeter to the northern reaches of the Korean Peninsula, he struggled to keep the U.S. Army and senior political and military leaders focused on assessing progress in terms of ground gained and overall strategic victory rather than the tactical attrition of enemy forces. Once the Chinese entered the war and drove the U.N. forces south, though, the nature of the campaign changed.

Scott Sigmund Gartner and Marissa Edmund Myers make a strong case that when the U.S. shifted to a war of attrition against the Chinese and North Koreans in 1951, senior U.S. officials viewed strategic success in a different light. Absent the opportunity to outmanoeuvre an enemy land army and seize national territory, the Army and other services began assessing progress through body counts. From this point in time forward, senior generals and political leaders described end state conditions in hazy terms (e.g., to bring the enemy to the negotiating table) while focusing on the means to achieving these ill-defined ends (i.e., killing enemy soldiers) as tangible evidence of progress.

Killing became an end unto itself, and killing was then translated into what today would be called a "metric" for individual and unit accomplishment. Gartner and Myers argue that the strategic conditions during the Korean stalemate of the early 1950s changed the U.S. military from an organisation that wins wars to an organisation that accomplishes quantifiable tasks. In turn, strategic assessment became a means of measuring tactical and operational task accomplishments rather than a process that informs vital strategic decision-making (Gartner and Myers 1995).

Assessments in Vietnam

During the interregnum between the Korean and Vietnam Wars policymakers focused on developing sufficient capacity to counter a possible Warsaw Pact invasion from Eastern Europe. Strategic thought during the decade between the early 1950s and the early 1960s revolved around conventional forces. Although U.S. advisors and intelligence officers supported small irregular and covert proxy wars against the Soviets across the developing world, these "shadow wars" had little impact on capstone doctrine.

Lessons that might have been learned from the Chinese intervention and the ensuing stalemate were largely ignored. Little to no effort was made to improve campaign assessment doctrine—assessment of on-going operations—or to help policymakers and senior officers better define strategic objectives for irregular campaigns. This resulted in an assessment process in Vietnam that often led to unintended behaviour ranging from hiding bad news to data fabrication. This section describes the assessment process, including the advisor reports, commander's reports and the Hamlet Evaluation System, and provides insights regarding its effectiveness.

Lack of clear strategic objectives

Arguably, strategic end state was far more ill-defined during the Vietnam War than during the early and even the later stages of the Korean War. Douglas Kinnard (1977) published a survey of over 100 U.S. general officers who had served in Vietnam or had been involved in the war. Their primary complaint regarding their experience with the Vietnam War was the absence of clear strategic guidance from national decision-makers. U.S. strategy in Vietnam in the early to mid-1960s shifted rapidly as President Lyndon B. Johnson's administration slipped from a limited advisory mission into graduated war against the North Vietnamese. Secretary McNamara and other advisors like McGeorge Bundy tried to help Johnson frame success generically as an absence of enemy hostility and, to a lesser extent, as a self-sufficient Republic of Vietnam (RVN). McMaster and others show how national decision-makers failed to conceive of or articulate a clear and decisive strategy for the Vietnam War at any point during the war.

Demands for assessments

As in Korea, strategic ambiguity translated into a strategic assessment process that inaccurately conflated short-term and generally quantifiable gains with progress towards winning the war. In Vietnam, however, the services counted everything they could conceivably define as militarily relevant in order to demonstrate their organisational relevance and value. Officers in the field were pressured to report positive tactical data by senior officers in Saigon; analysts in Saigon and Washington, D.C. were pressured to show how these data could be equated with progress; policymakers, scrambling to find shreds of positive evidence, demanded more data from below. MACV assessment reports generally came in three forms: 1) intelligence analyses; 2) advisor reports; and 3) commanders' personal assessments. Because there was little to no doctrinal guidance in the 1960s and early 1970s to help MACV and its subordinate commands build a campaign assessment process, enemy-centric intelligence analysis often sufficed as assessment. These analyses provided some convincing—and some less convincing—insight into North Vietnamese and Viet Cong capabilities and intentions, but they were of limited use to commanders and policymakers trying to understand the ebb and flow of the overall campaign. They did little to help decision-makers understand how popular sentiment in South Vietnam might influence strategic outcomes.

Advisors' reports

Military and civilian advisors helped to address this gap in understanding with contextual narrative reporting from the field and from Saigon, but these reports were rarely (or at least not clearly) integrated into an overall campaign or strategic assessment. In 1969 MACV limited the rich, detailed, and insightful province advisor reports to no more than four pages so as to make these reports easier to read. MACV commanding generals submitted their personal assessments as long-form narrative documents through the chain of command to the Secretary of Defense and the President, but these were often so laden with boilerplate language and unanchored detail that they were of little use. General William C. Westmoreland's "Military Assessment for the Month of September," 1967, included statements that seemed tailored to lift morale and defend current planning rather than provide sober insight:

> The first week of last month was marked by enemy terrorist rocket and mortar attacks designed to intimidate the people of South Vietnam in order to disrupt the national elections. A determined people answered that challenge unmistakably. Again the enemy has been defeated in his efforts to gain a major victory. (Westmoreland 1967, under "Text of cable from General Westmoreland")

Westmoreland's assessment informs senior political leadership about important information like an estimate of enemy capability and intent, but also delivers reams of out of context and seemingly irrelevant information like the total number of South Vietnamese troops who were airlifted around one of the sub-theatre regions (33,970) and the inter-provincial Route 30 from Kien Von to Hong Ngu improved from an amber to green colour-coded security rating. Accuracy estimates and citations were generally not present in the reports. Statements like "The overall morale, combat effectiveness, and fighting spirit of the Fourth Corps units continue to be good" (Westmoreland 1967, 12) were typically not supported with evidence and it is unclear what "good" means in terms of overall campaign objectives; it could mean anything from "fight like Spartans" to simply not collapsing in the face of the enemy.

Commanders' reports

Because quantifiable tactical and operational gains substituted for strategic progress, military assessments throughout the war and at all levels of command consisted primarily of dissociated sets of aggregated data. Ultimately the assessment reports—the commanders' personal efforts to inform senior decision makers as they tried to adjust strategy and allocate vital resources—were little more than lists of these data bracketed by brief and unqualified statements of opinion or characterisation (e.g., we killed more Viet Cong this month than last month). Nearly all of these narrative reports contained dense listings of body counts, weapons seizure counts, enemy initiated attack counts, numbers of tactical operations, and even the number of bars of soap distributed to the South Vietnamese population. At a quick glance the lists of data appear impressive, and collectively they give a sense of momentum and accomplishment.

It is vaguely comforting to read that a bilateral U.S. and South Vietnamese airmobile operation killed 70 Viet Cong, and that overall the Viet Cong in MACV's fourth corps sector lost 75 more weapons in September than in August of 1967 (Westmoreland 1967). Yet these numbers are practicably meaningless absent a total and accurate count of Viet Cong and some evidence that these casualties and weapons losses were driving the insurgents towards defeat; that information either did not exist, or it was not presented in the assessments.

Hiding bad news

Reading the full volume of available MACV assessment reports, one gets the sense that the authors (most likely staff officers) intended to convey a *sense* of progress by inundating the reader with data, rather than to provide a realistic assessment and a clear acknowledgment of operational weaknesses and information gaps. McMaster (1997), Corson (1968), Lewy (1978) and others have all documented the relentless political pressure that drove some military officers in Vietnam either to exaggerate progress, to hide bad news, or to simply deliver information without clarifying analysis in order to avoid becoming embroiled in higher echelon infighting. Congressman and former Marine Paul N. McCloskey, Jr., visited the MACV headquarters for policy briefings during General William C. Westmoreland's tenure. McCloskey's visit uncovered not only the nearly unwavering penchant to emphasize favourable data, but also a purposeful effort to hide what might be considered "bad" data (e.g., distressing ARVN desertion rates) from Congress:

> There was heavy stress on numbers, i.e. body count, crew-served weapons captured, strength of VC units, and particularly favorable *trends* in those numbers in every category as compared with three months earlier, a year earlier, etc. I do not recall a single unfavorable trend reported to us, and there was a consistent and strong expression that there was a "light at the end of the tunnel," that our "nation building" program was succeeding, that the VC strength was steadily eroding, and that in due course we would be able to return to an advisory status to a strong and stable South Vietnamese government…[I saw a manual] with the title along the lines of "Standard Operating Procedure for Handling Visiting CODELs [Congressional Delegations]." The manual explicitly outlined

> the requirement that CODELs were to be provided only with facts favorable to MACV's performance and directed withholding facts that would make "MACV's mission more difficult." (Westmoreland 1984, under "Plaintiff General William C. Westmoreland's Memorandum of Law in Opposition to Defendant CBS's Motion to Dismiss and for Summary Judgment")

Body counts and data fabrication

Military units at the tactical level struggled to meet the increasingly voracious appetite for data in Saigon and Washington. Body counts and "probable" body counts became important standards of operational measurement. Seizure of territory had little to no operational or strategic importance other than when it denied the enemy an area of manoeuvre, so killing often became the only objective for battalion, regimental, and even division commanders. Killing many enemy soldiers (either Viet Cong or NVA) was good, but it was also important to have a good kill ratio.

Because in the absence of other objectives the U.S. strategy had evolved into a process of attrition, it was important for commanders to show that each U.S. life lost could be valued against many more enemy dead. In this way, the U.S. could grind down the North Vietnamese over time, thereby bringing the North to the negotiating table. Because a high body count or a good body count ratio became clearly linked with tactical and then operational success, the pressure to report high counts and ratios intensified as the military sought to show strategic progress.

Yet counting bodies was often impossible in Vietnam's imposing terrain, and commanders who lost men on body counting missions became loath to risk their lives in further such efforts. Many officers and non-commissioned officers freely abandoned the notion of accurate body count reporting. Data fabrication became commonplace, and over time the body count data became riddled with inaccuracy and egregious inflation and ceased to be a useful measure of operational progress.[3]

3. For a detailed accounting of these falsifications and inaccuracies, see Connable 2012, 106-113.

The Hamlet Evaluation System

As ground combat and air units engaged in a war of attrition against the NVA and the VC, advisors also struggled to demonstrate quantitative value to feed the demand for measurable signs of progress with the South Vietnamese population and with the RVN armed forces. Specifically, those advisors assigned to help develop militia forces or RVN government offices from the hamlet to the provincial level were tasked with showing how effectively the RVN controlled the population of South Vietnam.[4] The idea of developing the strength of the government's connection to the population has been proven valid in many COIN cases (Connable and Libicki 2010).

Measuring that connection through quantitative assessment with any degree of accuracy has proven to be nearly impossible. However, the CIA and DoD teamed up in Vietnam to develop the most detailed and comprehensive quantitative population assessment program yet devised. The Hamlet Evaluation System (HES) still stands as "the only systematic micro-level database of civil war dynamics covering such a large territory and time period" (Kalyvas and Kocher 2009, 342). Yet in terms of providing accurate, realistic assessment decision support to senior leadership, it failed.

HES was developed and improved under the Civil Operations and Revolutionary Development Support (CORDS) program, a forward-deployed office dedicated to execute the pacification effort in South Vietnam. This included improving RVN governance, generating economic growth, and improving local security capacity. It also required support from special counterintelligence operations and direct combat troops conducting military operations against the Viet Cong. CORDS needed a way to show progress in its efforts towards achieving pacification, and senior policymakers in Washington, D.C. required indicators of measurable progress across what would now be termed all "lines of operation." CIA, which supported and was in turn supported by many CORDS initiatives, helped develop HES as a comprehensive measurement system to meet these complementary demands.

HES consisted of a questionnaire distributed to each district senior advisor (DSA) in Vietnam. At the district level, an advisor might be responsible for an

4. Some HES documents describe the system as a measurement of control, while others describe it as a measurement of active support of the population for the RVN.

area that encompassed from 50 to 200 or more hamlets, and a hamlet might consist of fewer than 50 to as many as 20,000 people. Many hamlets were either in areas under direct enemy control, under enemy influence, or located in remote and inhospitable terrain. Each DSA was expected to visit each hamlet every month when possible, and then to provide a highly detailed analysis of each hamlet with input from "two Free World and two Vietnamese [sources]" (Military Assistance Command Vietnam n.d., 22). Some versions of the HES worksheet contained several columns of individual checklist questions and notation requirements, as well as boxes for narrative comments. CORDS required a worksheet be submitted for every hamlet every month, or depending on which set of MACV or CORDS guidelines one followed, changes to be submitted once a baseline was established.

Quantification of HES reports

The idea of gaining greater understanding of popular sentiment, security, and local development is sound when attempting to establish regional stability—in CORDS parlance, pacification—in counterinsurgency or similar environments. But because long-form narrative reports from the DSAs and provincial advisory teams did not provide grist for the analytic mills in Saigon and Washington, these reports were considered inadequate for assessment of progress towards achieving pacification.

These subjective observations and informed (or poorly informed) opinions of the advisors in the field were translated into numbers. Quantification was then equated with objectivity, and aggregated HES results were then treated as accurate, objective data rather than data stripped of relevant local context and formatted for multi-layered aggregation and distillation. The result was a single pacification number for all of South Vietnam's approximately 12,000 identified hamlets. This number was used to show strategic progress in increments of 1/100th percentage growth. However, while it was precise, it was not accurate.

Data fabrication in HES

HES encouraged the same kind of wilful and widespread data fabrication elicited by the body count data requirement. Even if DSAs were inclined to complete all of the onerous HES requirements, a basic assessment of a District Senior Advisor's duties would have made it obvious that it was not physically

possible to accurately and comprehensively review even a few hamlets, let alone 50 or more hamlets per month. The handbook for DSAs for Quang Nam Province, Vietnam (Military Assistance Command Vietnam 1968a), did not mention HES at all. Instead, it listed these other critical tasks, amongst many others:

- Establish and maintain a combat operations center
- Provide direct combat advisory support to RVN counterpart
- Secure the district advisory compound from infiltration and attack
- Assist RVN counterpart in calling in supporting arms against enemy forces
- Support the Phoenix Program effort to root out VC infrastructure
- Provide input and support to the RVN Pacification and Development Councils

These core responsibilities translated into hundreds of hours of difficult and often dangerous work each month. Adding in the demands of combat operations, the need to provide leadership to the district advisory team, to communicate with the province team, to establish rapport with Vietnamese counterparts, to acquire supplies, to conduct basic maintenance, to eat, to sleep, and to simply move from place to place while avoiding mines and ambushes, and it becomes clear that there would be little time to devote to HES.

Yet senior leadership levied HES on the DSAs without carefully considering its feasibility. Assuming it took only one day to move to and observe a single hamlet, to interview a minimum of four informants, and to complete the quantitative and narrative parts of the Hamlet Evaluation Worksheet (HEWS), an advisor would have to spend literally all of his time, every day, without any time off for an entire month to observe 30 hamlets. Yet even in this generalized best-case scenario, advisors were expected to rate the hamlet on everything from government influence to security to economic development.

Ample anecdotal evidence shows that many DSAs paid little attention to HES. Senior level analysts at OSD assumed that only a few of the 18 top-level indicators on the worksheet could be rated "on the basis of direct observation of a clear cut condition," that "much of the HES information can be obtained only from Vietnamese [sources]," and that "most advisors cannot visit all of their hamlets during one month" (Connable 2012, 123-124). Some filled

in the worksheets with "whatever seemed reasonable," and many filled out reports on hamlets they had never seen and about which they had little to no second-hand information.

They were loath to change a rating from better to worse for fear of higher echelon investigation or reprisal, and there were implied incentives to show improvement in hamlet ratings over time. If many DSAs decided to simply assign ratings to hamlets they had not seen, and then to have these ratings stay static or to show incremental improvement over time, then the country-level HES rating would show slow but steady and unsubstantiated improvement over time. While it is not possible to empirically prove widespread data fabrication by DSAs, there was slow and steady improvement in the strategic HES rating from 1967 through 1972. By 1972 HES showed that over 90% of the South Vietnamese countryside had been pacified. Yet advisors reported that many top-rated hamlets were quietly controlled by the VC and did not pass the sleep test: spending one night in some A-rated hamlets could be deadly (Connable 2012).

Summary of value of HES

Quantification of pacification assessment through HES had the opposite effect of the one intended. Instead of improving the completeness, accuracy and objectivity of advisor inputs to strategic assessment, HES removed or averaged out the critical yet unquantifiable details from the advisors' reporting. Forcing almost impossible tasks on men in the field all but ensured widespread data fabrication and a reduced interest in the assessment process writ large. Advisors received little to no feedback from HES that would be relevant to their local missions, so their incentive to shift energy to assessment from their primary military missions was negligible. Once aggregated, the subjective, and partially falsified HES data was of little value in decision support, and ultimately may have played a role in misleading both policymakers and the general public.

OSD reports

Despite mounting evidence of falsification and the near total inability to determine their accuracy, aggregated quantitative data like body counts and HES were fed directly into the systems analysis assessment process at MACV and OSD. Operations research and systems analyst Thomas Thayer ran the

wartime assessment branch under Secretary McNamara. Thayer voluntarily took on the assessment task at OSD—his shop was officially an intelligence analysis office. Channelling McNamara's desire for quantitative measurement and statistical analysis, Thayer (1975) applied his operations research skills to develop an ad hoc strategic assessment report. Thayer's team of analysts produced over 50 reports, which are perhaps the most detailed and comprehensive example of Vietnam-era assessments. These reports were anchored in pattern and trend analysis of quantitative data flowing from the field.

Thayer and his analysts drew correlations using what he referred to as "reasonably accurate" body count data, HES, and what are now termed blue force, or operational reports from friendly units. He acknowledged serious problems with the data and with his findings, but, nonetheless claimed to find "definite" patterns. But these patterns and trends, constituted from aggregated quantitative data of unknown completeness and accuracy, were at best generally informative and at worst misleading. Thayer's statistical correlations and data characterisations did not provide a clear understanding of progress or lack of progress in the strategic campaign, and there is little evidence to show that they effectively supported key strategic decisions.

The least relevant of the OSD report series consisted primarily of quantitative data charts or interpretations of data, while the most interesting and relevant consisted of detailed, localized, and contextual analysis of specific intelligence issues. The former were on par with the ineffective MACV assessment reports, while the latter did not provide long-range strategic insight because they never rose above the level of operational intelligence.

Capturing the Lessons of Vietnam for Current and Future Campaigns

While it is widely accepted that the U.S. was defeated in Vietnam, little direct attention has been paid to the role of assessment in the first U.S. strategic defeat. Gregory Daddis (2011) did focus specifically on assessment failures, while McMaster (1997), Goldstein (2008), McNamara and VanDeMark (1996), and Lewy (1978) touched on assessment as it related to broader critiques of U.S. strategy. Any debate over the role of assessment in Vietnam tends to lead immediately to a grander debate over the merits of attrition warfare in counterinsurgency. While these two issues are closely linked, and in retrospect

a campaign of attrition was inappropriate in Vietnam, the issue of attrition distracts from the more generalizable lessons for assessment.

The need for well-defined strategic campaign objectives

The MACV and OSD assessment processes failed for several reasons, each of which is relevant to recent experience in Iraq and Afghanistan and also the on-going development of NATO assessment concepts and methods. Absence of a clear strategic end state meant that commanders and analysts did not have well defined objectives on which to base assessments. This was amplified by the complexity of the environment and the political conditions in Vietnam and in the U.S. In an attempt to counterbalance this lack of clarity, some senior officials and military commanders sought more and more detailed quantitative data. If they could show what appeared to be objective and measurable progress—often by presenting quantitative data as both precise and unequivocally accurate—then they could fend off criticism of the war and of military combat performance.[5]

Field commanders accustomed to manoeuvre and attrition warfare from their experience in WWII and Korea readily transposed conventional assessment theory to Vietnam. But conventional assessment methods cannot be reasonably transposed to irregular warfare like COIN because success in COIN is typically defined by ambiguous and unquantifiable standards.

Reasonable reporting requirements will reflect the value of the data

Once this assessment concept was (albeit informally) established, it became necessary to levy extraordinary reporting requirements on subordinate units. Intermediate objectives like killing enemy soldiers served as ersatz strategic end state conditions. Yet this requirement was doubly illogical: it demanded tactical commanders report data that could not be reasonably or accurately collected in order to feed a strategic assessment requirement that did not show progress towards a strategic end state (Connable 2012). Many soldiers and Marines were killed on body counting missions, and over time leaders became reluctant to risk their men in order to pursue assessment data.

5. For example, hamlet security data were often presented to the public down to the tenth of a percentage point.

At both the strategic and operational level, military officers generated their own reporting requirements that, over time, became a de facto purpose of the U.S. military in Vietnam. Instead of producing a strategic end state, military officers produced thousands upon thousands of quantifiable inputs and effects. Assessment reports from MACV and OSD conflated the mass of quantitative inputs and effects data—few of which had definable value—with strategic progress or lack of progress.[6] Philip Caputo (1977), Marine platoon commander in Vietnam and the author of the landmark memoir *A Rumor of War*, wrote of the body count process:

> General Westmoreland's strategy of attrition also had an important effect on our behavior. Our mission was not to win terrain or seize positions, but simply to kill; to kill Communists and to kill as many of them as possible. Stack 'em like cordwood. Victory was a high body-count, defeat a low kill-ratio, war a matter of arithmetic. The pressure on unit commanders to produce enemy corpses was intense, and they in turn communicated it to their troops. This led to such practices as counting civilians as Viet Cong. "If it's dead and Vietnamese, it's VC," was a rule of thumb in the bush.

Data fabrication misled decision makers and the public, but it also undermined military good order and discipline by forcing thousands of tactical-level leaders to knowingly violate their integrity and to openly demonstrate disdain for senior military and political leadership. This unintended side effect of the assessment process almost certainly contributed to the general malaise in the U.S. armed forces in the mid to late-1970s.[7]

Don't underestimate the value of narrative data

If MACV had emphasized the value of the advisors' narrative reporting it could have developed an admittedly subjective but ultimately insightful and

6. To Thayer's credit, he identified the lack of value in body counts and in some other data sets during his tenure in OSD. However, he continued to defend his quantitative findings through the 1980s. See Connable 2012.

7. Many of the reports and books published in the post-war period indicated as much. See Connable 2012, Bibliography for references to these sources.

relevant assessment for senior leaders. In retrospect, advisor reports from the field were more honest, more accurate, and more operationally and strategically relevant than any of the raw quantitative information provided to higher headquarters.

A good example is the Phong Dinh Province Report submitted to CORDS headquarters for the month of February, 1968 (Military Assistance Command Vietnam 1968b). This report, written by the province advisory team and supplemented with narrative reports from the DSAs, revealed serious threats to mission success at the tactical, operational, and strategic levels of war. It also made clear recommendations as to how these threats could be mitigated in local context. While it is subjective, it is certainly no more subjective than any given HES report. And while it probably delivers far from perfect accuracy, it is probably far more accurate than body count reports or operational patrol activity reports. Because the authors were given the leeway to write only about what they knew in local context, and felt comfortable identifying things they did not know, reports like this were less likely to encourage falsification than generalized, one size fits all, quantitative reports.

Summary and Comments on Recent Military Operations

Like Vietnam, both Operation Iraqi Freedom and Operation Enduring Freedom (the U.S. mission in Afghanistan) have been relatively ill-defined campaigns with shifting strategic end state objectives. In both of these campaigns senior leaders across the various coalition nations demanded reams of quantitative data from their operational commanders which, in some cases, may have been an attempt to compensate for a lack of operational and strategic clarity and the inability to discern meaningful progress over time. At one point the assessment process in a regional command in Afghanistan demanded that subordinate units report over 400 categories of quantitative metrics, and a senior assessment officer in Kabul estimated that there were over 2,000 mandatory reportable quantitative metrics leveraged on subordinate units across the theatre in 2011 (Connable 2012; Downes-Martin 2011).

This relentless pursuit of empirical evidence of progress from the field, driven both by senior officers and policymakers, revealed their discomfort with the level of uncertainty inherent in these long running counterinsurgency

operations. Anecdotal evidence from the field shows that commanders and their staffs saw little to no value in reporting these data to higher headquarters, and they gained nothing relevant to their activities from operational or strategic assessments. Consumers at the strategic level echoed these complaints: of an estimated 300 NATO military and civilian officials involved in building or receiving assessments of the Afghanistan campaign who were interviewed for recent RAND research, none described the assessment process as useful or relevant to decision making.[8] And while levels of data fabrication are probably lower in Iraq and Afghanistan than they were in Vietnam, many officials plainly and openly admit to falsifying data in both theatres to answer demands for quantitative data that they cannot reasonably acquire.[9] In both Iraq and Afghanistan, NATO members have replicated many of the errors of over-quantification that crippled U.S. Vietnam-era campaign and strategic assessment.

This is not to say that quantification has no role in strategic assessment. Instead, the Vietnam War showed that it is necessary to use an assessment process that is appropriate to the conflict at hand. Quantitative assessments proved quite useful to support strategic decision making during WWII. It was possible to obtain quantifiable data in that context that were both accurate and relevant to the clear strategic outcome established by senior political and military leaders. Quantitative *analysis* has proven effective and has saved thousands of lives in every war of the 20th and 21st Centuries, to include in Korea, Vietnam, Iraq, and Afghanistan.[10] But a campaign and strategic assessment concept derived primarily from quantitative data and statistical, pattern, and trend analysis is not appropriate to COIN, or probably to many other types of irregular warfare. These approaches have failed to adequately support decision-making in wars with ill-defined strategies, complex environments, and shifting end states.

8. These interviews were completed in early 2012. Since that time ISAF has made significant changes to its assessment process, most of which have been positive.

9. For specific quotes, see Connable 2012. Some officials openly admitted to falsifying data in large, open, official ISAF, NATO, and other forums between 2009 and 2012.

10. For example: submarine detection analysis in WWII; logistics analysis in Korea; air sortie analysis in Vietnam; and counter-IED analysis in both Iraq and Afghanistan.

It is necessary in irregular warfare to accept the limits on knowledge and the vagaries of meaning imposed by the mission, the environment, and the type of enemy faced. In these conflicts decision makers need to rely more on the informed, subjective assessment of trusted commanders. In turn, they will have to rely on their professional judgment and on a range of other inputs to inform their decisions; they will never be provided with a clear or simple picture or path forward. Efforts to compensate for this subjectivity by seeking unachievable quantitative objectivity will be generally counterproductive. For campaign and strategic assessment, one size does not fit all.[11]

11. For a recommended alternative to the doctrinal U.S. and NATO assessment methods, see Connable 2012. Appendixes A and B contain a template for an alternative approach and an example of an alternative, bottom up, contextual assessment.

References

Caputo, P. 1977. *A Rumor of War*. Pimlico.

Connable, B. 2012. *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency.* Santa Monica, CA: RAND Corporation. Available online <www.rand.org>

Connable, B., and M. Libicki. 2010. *How Insurgencies End*. Santa Monica, CA: RAND Corporation.

Corson, W. R. 1968. *The Betrayal*. New York, NY: W. W. Norton.

Daddis, G. A. 2011. *No Sure Victory: Measuring US Army Effectiveness and Progress in the Vietnam War*. New York, NY: Oxford University Press.

Downes-Martin, S. 2011. Operations assessment in Afghanistan is broken: What is to be done? *Naval War College Review* 64(4):103-125.

Gartner, S. S., and M. E. Myers. 1995. Body counts and 'success' in the Vietnam and Korean wars. *Journal of Interdisciplinary History* 25(3):337-395.

Goldstein, G. M. 2008. *Lessons in Disaster: McGeorge Bundy and the Path to War in Vietnam*. Times Books.

Kalyvas, S. N., and M. A. Kocher. 2009. The dynamics of violence in Vietnam: An analysis of the Hamlet Evaluation System (HES). *Journal of Peace Research* 46(3):335-355.

Kinnard, D. 1977. *The War Managers*. Hanover, NH: University of New Hampshire Press.

Lewy, G. 1978. *America in Vietnam*. New York, NY: Oxford University Press.

McMaster, H. R. 1997. *Dereliction of Duty: Johnson, McNamara, the Joint Chiefs of Staff, and the Lies that Led to Vietnam*. New York, NY: Harper Perennial.

McNamara, R. S., and B. VanDeMark. 1996. *In Retrospect: The Tragedy and Lessons of Vietnam*. New York, NY: Random House.

Military Assistance Command Vietnam.1968a. *District Senior Advisor Handbook*. CORDS <www.vietnam.ttu.edu>

——— 1968b. *Phong Dinh Province 2/68*. CORDS. <www.vietnam.ttu.edu>

——— n.d. *Hamlet Evaluation System District Advisor's Handbook*. CORDS Research and Analysis Division.

Overy, R. 1997. *Why the Allies Won*. New York, NY: W.W. Norton.

Shrader, C. R. 2006a. *History of Operations Research in the United States Army, Volume 1: 1942-1962*. Washington, DC: Army Center Of Military History.

——— 2006b. *History of Operations Research in the United States Army, Volume 2: 1961-1973*. Washington, DC: Army Center Of Military History.

Thayer, T. C. 1975. *A Systems Analysis View of the Vietnam War 1965-1972.* Washington, DC: Office of the Assistant Secretary of Defense, Programs Analysis and Evaluation.

U.S. Army. 2010. *The Operations Process*. Field Manual 5-0. Department of the Army.

U.S. Army Air Forces. 1948. *Operations Analysis in World War Two*. Philadelphia, PA: Stephenson Brothers. Available online <www.mors.org>

Watson, G. M. Jr. 2003. Whiz kid: Robert S. McNamara's World War Two service. *Air Power History* (54)4.

Westmoreland, W. C. 1967. *Military Assessment for the Month of September*. Available online <http://www.mrfa.org/pdf/COMUSMACVSept67.pdf>

——— 1984. Plaintiff General William C. Westmoreland's Memorandum of Law in Opposition to Defendant CBS's Motion to Dismiss and for Summary. In *Westmoreland v. CBS*, 752 F.2d 16, 2d Cir., November 2, 1984.

PART 2:

CURRENT ISSUES IN OPERATIONS ASSESSMENT

3

Operations Assessment at ISAF: Changing Paradigms

Dr. Jonathan Schroden

Abstract

From 2009 to 2011, the assessment paradigm used internally by the International Security Assistance Force (ISAF) to assess progress in the war in Afghanistan changed significantly a number of times, vacillating between heavily quantitative attempts to "measure" progress and other approaches that sought to increase transparency and include commanders' judgments. In September 2011, ISAF got a new commander who directed the Afghan Assessment Group (AAG) to redesign the way in which ISAF was assessing progress in the war, and to be "revolutionary" in so doing. The resulting assessment paradigm was novel, non-doctrinal, and effectively addressed the unique complexities of the counterinsurgency in Afghanistan and the needs of the ISAF Commander. It had a two-tier structure consisting of both strategic and campaign assessments; the former focused on answering Strategic Questions in narrative, analytic form while the latter used standards and accompanying narrative responses to gauge accomplishment of campaign tasks. Both tiers captured the state of the war while maintaining an eye on future challenges and opportunities.

The two assessments and their associated processes were designed to stimulate discussions leading directly to decisions by senior leaders on actions they could take, direct, or request. Such actions were the observed outcome of the first five iterations of the new paradigm. While any assessment paradigm will have advantages and disadvantages, an examination of the pros and cons of the ISAF assessment paradigm in 2012 makes clear that it should be considered a "best practice" in the field of counterinsurgency assessment.

Introduction

"*Are we winning?*" It is a commonly asked question in war, and one that is not always easily answered, especially in the case of a counterinsurgency (COIN). For the war in Afghanistan, it has been asked repeatedly, and the answers to it have taken many forms—from the word of 4-star commanders and senior politicians to detailed studies performed by analysts on the ground. Throughout the course of the war, the International Security Assistance Force (ISAF) has attempted to answer this question internally using a variety of techniques, to include the use of strategic and operational (or campaign) assessment. This chapter will present a brief history of assessments conducted by ISAF in Afghanistan from 2008 to 2011 followed by a detailed description of a new assessment paradigm that the author helped create in early 2012.[1]

As this chapter will show, ISAF's assessment paradigm in 2012 was borne of the lessons learned from both failures and successes in previous assessment paradigms at ISAF and elsewhere. Like many assessment paradigms, it was crafted to better understand the current state of the war, but went beyond that in its attempts to identify future opportunities and challenges, along with the corresponding risks to mission success if these challenges were not overcome. Perhaps most significantly, this paradigm moved away from previous attempts to quantitatively "measure" progress in the war and instead used standards and structured discussions to inform and enhance strategic and operational decision-making. Given its innovations and the positive results that it engendered at ISAF, this chapter will contend that the ISAF assessment paradigm in 2012, while designed for a specific set of circumstances, should serve as a seminal example of how to assess progress and inform decision-making in a counterinsurgency.

Paradigms Past

In 2008, ISAF's attempts to gauge progress in the war stemmed mainly from the efforts of its Operational Analysis Group (OAG) and the estimates of its commander (COMISAF). Generally speaking, the command at that time

1. This timeframe is the result of the author's personal experience with assessments in Afghanistan, which began in 2008. This chapter presents an expanded description of the ISAF assessment paradigm described in Schroden et al. 2012. This chapter is the opinion of the author and does not necessarily represent the views or positions of the AAG or ISAF.

lacked a formal strategic and campaign assessment paradigm. In early 2009, ISAF recognised a need to resource and conduct assessments, so it established the AAG to serve as the theatre lead for strategic and operational assessment.[2] The AAG then endeavoured to create a formal assessment paradigm and after significant effort, developed one that utilised quantitative metrics and weighted-averaging "roll-up" techniques that were commonly found in U.S. and effects-based doctrine at that time.[3] This approach encountered difficulties akin to those noted by others with experience using similar approaches (Connable 2011; Downes-Martin 2011; Schroden 2011). For example, it was data-intensive, mechanistic, and it relied on mathematical formulations that obscured underlying causes and effects. Additionally, the specific metrics that it employed engendered significant criticism (Campbell, O'Hanlon, and Shapiro 2009; Exum et al. 2009; Schroden 2009; Agoglia, Dziedzic, and Sotirin 2010; Kilcullen 2010; O'Hanlon and Sherjan 2010). Even the head of the AAG at that time was quoted as saying, "Our metrics suck" (Soeters et al. 2010, 222).[4]

When General Stanley McChrystal took command of ISAF in mid-2009, his personal assessment reinforced these criticisms, stating:

> ISAF must develop effective assessment architectures…to measure the effects of the strategy, assess progress toward key objectives, and make necessary adjustments. ISAF must identify and refine appropriate indicators to assess progress, clarifying the difference between operational measures of effectiveness critical to practitioners on the ground and strategic measures more appropriate to national capitals. (McChrystal 2009, 2-20)

As a result, ISAF revised its assessment paradigm to one that blended qualitative information with quantitative metrics in a more holistic fashion.

2. The AAG absorbed the OAG as part of its initial structure.

3. For an example of this, see Appendix H of U.S. Army FM 5-0, *The Operations Process*.

4. Around this time, ISAF was under heavy pressure from senior decision-making entities (e.g., the U.S. Congress, U.S. National Security Council, and NATO's North Atlantic Council) to provide data on specific quantitative metrics dictated by them. This top-down guidance impacted the AAG's design of the assessment paradigm and its selection of metrics. While this pressure eventually lessened, reports such as the U.S. Congressionally-mandated "Report on Progress toward Security and Stability in Afghanistan" (also known as the 1230 Report) and NATO's Periodic Mission Review (PMR) continued to rely heavily on ISAF's assessment products for the bulk of their information throughout the timeframe of this chapter.

It continued to collect many of the quantitative metrics that formed the crux of the previous paradigm, but it placed greater emphasis on the professional military judgment of commanders in the field, such as those at the ISAF Joint Command (IJC) and NATO Training Mission—Afghanistan (NTM-A); new subordinate commands to ISAF that were stood up around that time. Due to General McChrystal's shift to a population-centric COIN strategy and the associated surge of forces into Afghanistan, the ISAF assessment paradigm shifted to assess progress along three major lines of operation (LOOs): protecting the population, building the Afghan National Security Forces (ANSF), and increasing the capacity of the Afghan government. The products of ISAF's assessment paradigm also shifted during this time, from mostly quantitative and data-intensive charts and graphs to more nuanced and narrative "judgments" that were presented in written form vice the more typical briefing format. General McChrystal also placed a heavy emphasis on transparency and openness of the assessment. He pushed hard to make ISAF's assessment products unclassified and to distribute them more broadly (e.g., to academics, think tanks, and the media). Ultimately, these efforts were overruled by senior decision makers and ISAF's primary assessment products remained classified.

This shift in the ISAF assessment paradigm was somewhat overshadowed by the creation of the IJC District Assessment Model (DAM) during roughly the same timeframe.[5] This model focused on assessing the status of security, governance, and development in the districts of Afghanistan using a set of rating definition levels (RDLs).[6] These RDLs amounted to 3-6 conditions that IJC asked its subordinate commands (the Regional Commands, or RCs) to rate along a colour scale which was then used to generate colour-coded

5. Several agencies and organisations also created assessment models during this same timeframe, largely in response to General McChrystal's comments on the ISAF assessment paradigm. At a NATO conference on assessment that took place in 2009 in Brunssum, The Netherlands, a list of operational assessments was compiled that numbered greater than twenty.

6. Examples of RDLs include the following, which were used by IJC to assess whether the overall status of a district was "active support for [the] government: environmental conditions in [the] district with respect to governance, development, and security are such that a super-majority (>70%) of the population accepts legitimacy of GIRoA [the Government of the Islamic Republic of Afghanistan] or of traditional local governance structures not actively in opposition to GIRoA; environmental conditions in [the] district with respect to governance, development, and security are such that open, active support for GIRoA is routine; and, environmental conditions in [the] district with respect to governance, development, and security are such that a large majority of individuals are satisfied with conditions in terms of their personal, community, food, health, economic, [and] political security." For more details, see "ISAF Joint Command District Assessments," available at: <http://info.publicintelligence.net/ISAFdistrictassessments.pdf>

maps of the country at the district level.[7] This model itself engendered heavy criticism from a number of civilians with experience in operations assessment (Downes-Martin 2010; Schroden 2010; Connable 2011).

In mid-2010, General David Petraeus assumed command of ISAF and the command re-wrote its operational plan (OPLAN) to have a stronger focus on transitioning security responsibility to the Afghans. The new plan contained six LOOs, each with its own operational objective and set of intermediate objectives (waypoints on the path to the operational objective), as illustrated in Figure 3.1. The corresponding assessment paradigm evolved back to a focus on quantitatively "measuring" progress towards accomplishment of the intermediate objectives in each LOO (as indicators of incremental progress towards the operational objectives). To do so, each intermediate objective was assigned a set of mostly quantitative metrics to elucidate whether the objective had been achieved, and if not, how close ISAF was to achieving it. These metrics were then used to determine where, on a "thermograph" (scale of red to green, red being bad and green being good), ISAF stood with respect to accomplishing each intermediate and operational objective. The products of this assessment

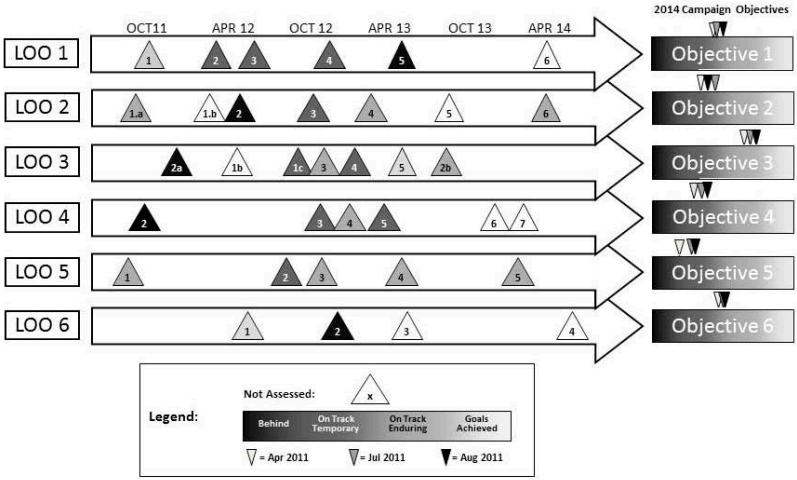


Figure 3.1: ISAF assessment paradigm circa 2010-2011 (triangle numbers, locations, and colours are notional)

7. For examples of these colour-coded maps, see ibid.

paradigm were largely slides showing these thermographs (e.g., Figure 3.1) and slides showing specific trends for a variety of quantitative indicators (e.g., level of violence)—all of which were classified.

New Commander, New Paradigm

As the above history of ISAF assessment paradigms illustrates, the nature of the assessment changed relatively often in the years 2009-2011, mostly in concert with the changes in COMISAF. Whether or not these changes in the paradigms' designs and outputs were positive or negative is a matter for debate, but it has been suggested that the staff processes associated with supporting each paradigm led to increased introspection and staff interaction, thereby adding value to the command.[8] This is supported by the observation that the emphasis the command placed on the assessment function during this timeframe was significant and increasing.[9]

It is not surprising in light of this history that when General John Allen took command of ISAF in late 2011, he directed the AAG to develop a new paradigm for assessing progress in the war. His initial direction to the AAG to be "revolutionary, vice evolutionary"[10] in so doing led to a departure from past ISAF paradigms and assessment doctrine to a paradigm that was unique, innovative, and focused on enhancing decision-making vice "measuring" progress. The remainder of this chapter will discuss, in order, the thinking behind the design of this new assessment paradigm; the paradigm itself along with examples of its products; a discussion of its pros and cons; and some concluding thoughts.

8. Email exchange with a former Director, AAG.

9. The AAG evolved over time, such that by 2012, it consisted of three main elements, led by a U.S. Colonel with an advanced degree in Operations Research/ Systems Analysis (ORSA): an operations analysis component; a survey analysis component; and a strategic and operational assessment component. The AAG itself was absorbed into the Strategic Transition Assessment Group (STAG, created in 2010), which was led by a British Brigadier. The STAG also contained a second group focused on the implementation of the *Inteqal* (transition) process. The Director, STAG reported to the Deputy Chief of Staff for Operations (DCOS OPS; led by a U.S. Major General) that also contained the plans and operations staff sections (each led by a one-star general). This eventual arrangement, in which the plans, operations, and assessments functions were led by officers of equal rank reporting to a single DCOS, was significantly different from typical U.S. staff structures where the assessment function is usually not treated as a co-equal to plans and operations.

10. Verbal communication with General John R. Allen, October 2011.

Thinking about Assessment Design

Commander's Guidance

Shortly after taking command of ISAF, and having been present for a Commander's Assessment Conference that used General Petraeus' assessment paradigm, General Allen directed the AAG to design a new paradigm for assessing the state of the war in Afghanistan and progress towards achieving strategic and campaign goals. In so doing, he directed the AAG to be "revolutionary, vice evolutionary" in its approach to the re-design. He indicated a desire for the assessment to go beyond simply presenting information to senior leaders and to actually stimulate discussion among them. He wanted the results of the assessment to be actionable; to identify challenges and opportunities and means of addressing them that were within his span of control (i.e., actions he could take, direct, or request). He wanted the assessment to be both holistic (i.e., to have his subordinate and supporting commanders involved in the assessment's inputs, outputs, and outcomes) and comprehensive (i.e., to consider all aspects of the war, not just military). Finally, he directed the AAG to reach out to, and enlist the help of, experts in the field of military assessment to ensure the best and newest ideas were being incorporated in the new paradigm.[11] General Allen's guidance served as the touchstone for the re-design of the assessment paradigm, in that the AAG repeatedly came back to his comments to check whether the products and processes being designed met his intent and satisfied his requirements.

Doctrinal and Literature Review

With COMISAF's guidance for the new paradigm in hand, the AAG examined doctrinal publications and the literature on assessment for new ideas. Being that ISAF is a command under the North Atlantic Treaty Organisation (NATO), the AAG first reviewed NATO doctrine. NATO's Operations Assessment Handbook defines operations assessment as "the function that enables the measurement of progress and results of operations in a military context, and

11. The author was one of these individuals and embedded with the AAG from November 2011 to February 2012. Ben Connable and Stephen Downes-Martin also visited the AAG for several weeks in January 2012 to conduct an after-action review of the assessment paradigm.

the subsequent development of conclusions and recommendations that support decision making" (NATO 2011, 1-1). This handbook also uses the term "strategic assessment," which it defines more broadly:

> Strategic assessment involves varying combinations of: continual measurement of effects and progress towards the achievement of military objectives; continual measurement of progress and results in non-military domains; measurement of progress and results of non-military organisations; an overall evaluation of progress towards the NATO end-state; and the subsequent development of conclusions and recommendations that support decision making. (NATO 2011, 2-5)

The AAG also considered U.S. military doctrine, which uses the term "assessment." U.S. Joint Publication (JP) 1-02 defines this as "determination of the progress toward accomplishing a task, creating an effect, or achieving an objective" (U.S. Joint Staff 2010, 21). Other U.S. doctrinal publications gave different definitions for this same term (e.g., U.S. Joint Staff 2011a *JP 3-0*, U.S. Joint Staff 2011b *JP 5-0*, U.S. Army 2010 *FM 5-0*, and U.S. Army 2006 *FM 3-24*).[12]

In addition to providing asynchronous definitions of what assessment is, these doctrinal publications differ in the purpose they assign to it. An examination of these documents reveals stated purposes ranging from informing commanders' decision making, to evaluating the performance of subordinate units, to showing causal linkages between actions and the achievement of objectives or effects (Schroden 2011). These are wide-ranging and distinct functions that may not be achievable via a single assessment paradigm. But perhaps most confounding to those tasked with assessing progress in military operations is that most of these doctrinal publications say little to nothing about how to actually conduct an assessment. It is perhaps not surprising that a number of criticisms of these documents along with proposed alternative assessment concepts have recently appeared in the literature (Connable 2011;

12. In May 2012, the U.S. Army revised FM 5-0 and published a new version as Army Doctrine Publication (ADP) 5-0. This new version uses the same definition for "assessment" as that in JP 3-0.

Downes-Martin 2011; LaRivee 2011; Schroden 2011; Upshur, Roginski, and Kilcullen 2012).

With the above doctrinal deficiencies and COMISAF's guidance to be "revolutionary" in mind, the AAG made a deliberate decision to forego using a doctrinal approach in designing the new assessment paradigm and instead design something novel. This included departing from NATO doctrine, which is still heavily rooted in the ideas and language of an Effects-Based Approach to Operations (EBAO). EBAO is a concept that has been hotly debated within, and at least partially rejected by, the U.S. military for conventional operations (the U.S. refers to these ideas using the moniker Effects-Based Operations, or EBO) and convincing arguments have been put forward against the use of EBAO/EBO in COIN operations as well (Connable 2011).

Design Constraints

The AAG identified several constraints that were important to the redesign thought process—these are important to bear in mind when determining which aspects of this paradigm can be generalized to other situations.

In addition to adhering to the Commander's guidance, one constraint that is universal to practitioners of military assessment is the requirement to "stick to the plan." Whether or not the assessment team has been involved in campaign or operations planning; the requirement to assess the plan's tasks or objectives is paramount. This also implies that the assessment design will conform to the plan's structure. For example, the ISAF OPLAN that was developed under General Allen did not state any objectives; rather it contained a set of essential tasks to be completed along with an assertion that accomplishment of those tasks would equate to mission accomplishment.[13] As will be shown below, this significantly influenced the design of the new assessment paradigm.

A second constraint was that the AAG had to adhere to the reporting and battle rhythm requirements that were already in place. ISAF had assessment reporting requirements to the NATO and U.S. chains of command (the latter via the dual-hatting of COMISAF as the Commander, U.S. Forces – Afghanistan). Thus, some aspects of the new paradigm, for example using a quarterly cycle

13. This was ISAF OPLAN 38302 Revision 6. Revisions 6.1 and 6.2 were also published during General Allen's tenure as COMISAF. While some details of the plan changed in each revision, the general structure of the plan remained largely the same.

or having a written report as a product, were not subject to change with the re-design.

Finally, the AAG had to work within the resources at its disposal to both design and execute the new paradigm. For example, the AAG was not manned nor empowered to act as a fully independent assessment entity. This led to a decision to base the new assessment processes around leveraging expertise from the entire ISAF staff and ISAF's subordinate/supporting commands, as opposed to the AAG conducting a fully independent assessment on its own.

Each of these constraints led to decisions with upsides and downsides. For example, the decision to design a staff- and command-based assessment (vice conducting an independent assessment) ensured that all of the relevant viewpoints were being exposed and discussed during the assessment process, but it also led to some views that ISAF was allowing those tasked to execute its campaign plan with assessing their own progress. These and other pros and cons of the new assessment paradigm will be discussed later in the chapter. The next section will focus on the details of the new paradigm and how it was executed in January, April, July, and October 2012, and January 2013.

ISAF's Assessment Paradigm in 2012

Assessment Framework

The sections above set the design boundaries for the new assessment paradigm. With these in mind, the AAG developed an assessment framework to help guide its thinking. To ensure the new paradigm was holistic, the AAG concluded that it must contain two components: a campaign assessment, focused on assessing progress in the execution of the ISAF OPLAN; and a strategic assessment, focused on assessing progress towards our core strategic goals in Afghanistan (described below). To ensure the new paradigm was comprehensive, the AAG concluded that it must look across four domains: security, governance, socio-economics, and regional relations.[14] Thus, an overview of the new paradigm would be that it was both a campaign and

14. While the first three of these domains had been part of previous assessment paradigms, the AAG added the regional relations domain in 2012 to ensure proper emphasis was given to this critical aspect of the campaign.

strategic assessment across four fundamental domains, designed to assess progress towards the achievement of ISAF campaign, and NATO and U.S. strategic goals in Afghanistan.

Campaign Assessment

Inputs

To assess progress in the execution of ISAF's campaign, the AAG decided that the inputs should come from ISAF's subordinate/supporting commands. The AAG believed these commands had the best visibility of conditions on the ground since they were the ones conducting operations. This approach also ensured the AAG was capturing the viewpoints of ISAF's subordinate/supporting commands in regards to how the campaign was being executed. This resulted in two fundamental inputs for the campaign assessment that are discussed next.

The first input was an assessment of the ISAF OPLAN's eight essential tasks. In regards to how best to assess these, the AAG considered a number of approaches ranging from purely quantitative (i.e., numerical metrics) to purely qualitative (i.e., subjective inputs), but ultimately settled on an approach that used standards. The use of standards will be discussed shortly, but first an overview of the ideas guiding this decision. While a strictly quantitative approach held the appeal of being able to hold subordinates accountable for their actions (since numerical metrics are arguably more objective), the AAG also knew that in practice, ideal numerical metrics may not be measureable, numbers can be easily falsified, incorrectly manipulated, or incorrectly interpreted, and an overly-narrow focus on a core set of numerical metrics tends to strip away the nuance and context that is necessary to understand the complexities of the operational environment (Connable 2011). Additionally, using a purely mathematical model for assessment makes little sense since there is no numerical model of counterinsurgency (Downes-Martin 2011).

At the other end of the spectrum, a strictly qualitative approach held the appeal of allowing those providing input to the assessment the flexibility to include whatever relevant information they had along with the nuance

Campaign Essential Task 1: Secure area A					
	Level 1	Level 2	Level 3	Level 4	Level 5
Security	Areas A is not secure	Area A is partially secured but with significant risk of reversion	Area A is partially secured but with moderate risk of reversion	Area A is partially secured but with minimal risk of reversion	Area A is fully secured with minimal risk of reversion
Governance	Key government actors are not present in area A	Some key government actors are present in area A and/ or their actions are significantly undermining security	A majority of key government actors are present in area A and/ or their actions are moderately undermining security	All key government actors are present in area A and/or their actions are minimally undermining security	All key government actors are present in area A and they are actively working to enhance security
Socio-Economic	Security conditions in/around area A are significantly hindering legitimate socio-economic activity	Security conditions in/around area A are moderately hindering legitimate socio-economic activity	Security conditions in/around area A are having minimal impact on legitimate socio-economic activity	Security conditions in/around area A are having no impact on legitimate socio-economic activity	Security conditions in/around area A are enhancing legitimate socio-economic activity
Regional Relations	Other countries are playing an overall significantly negative role with respect to security in area A	Other countries are playing an overall moderately negative role with respect to security in area A	Other countries are playing an overall minimally negative to minimally positive role with respect to security in area A	Other countries are playing an overall moderately positive role with respect to security in area A	Other countries are playing an overall significantly positive role with respect to security in area A

Table 3.1: Notional example of campaign assessment standards for a campaign essential task

and context required to fully understand their situation. But the AAG realised that a purely qualitative approach is easily manipulated in practice (since there are no benchmarks against which to compare inputs) and subject to a shifting baseline as units rotate into and out of theatre.[15]

The use of standards provided a means of balancing the pros and cons of purely quantitative or qualitative approaches, in that standards set a common framework for thinking about the campaign while also providing space for nuance and context to be captured and discussed. Thus, the AAG campaign assessment design had ISAF's subordinate/supporting commands assessing and

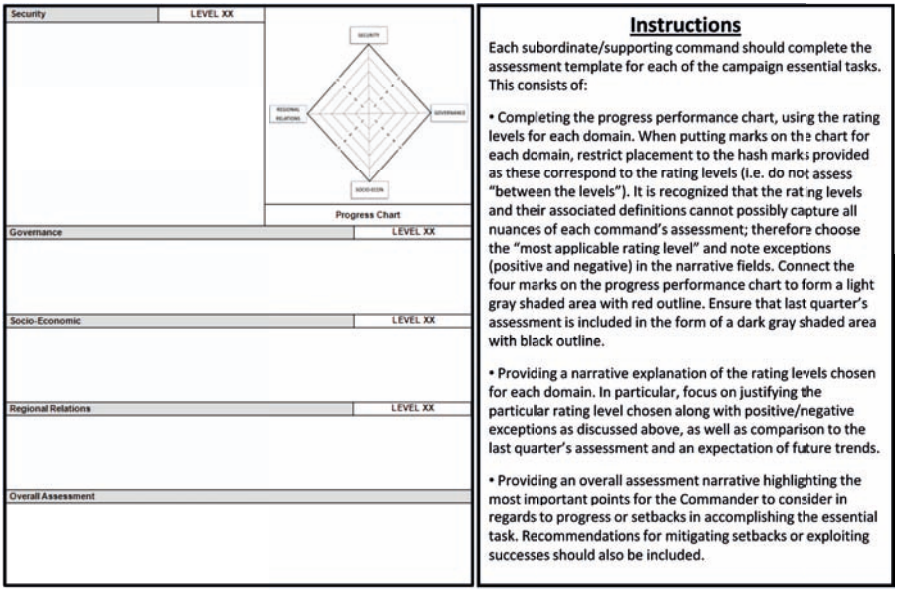


Figure 3.2: ISAF campaign assessment template

reporting their progress for each essential task using four domains and a set of five standards (levels) per domain. An example set of notional standards for the four domains for a single essential task is shown in Table 3.1. These standards are simple, high-level, declarative statements about the most important aspects

15. This is the so-called "saw tooth" effect that occurs when a new unit arrives, declares that things in its area of operations are not as good as the prior unit assessed, and then proceeds to assess things increasingly positive during the remainder of its tour, until the next replacement unit arrives and repeats the cycle.

of each domain and task.[16] They were specifically not designed to capture the rationale and nuances associated with the command's viewpoint; to do that, the AAG issued the template and instructions shown in Figure 3.2. Several items are worth noting in regards to this template.

First, the space allocated to the reporting of the standard level for each domain is quite small, with a set of much larger free-text fields making up the bulk of the template. As the corresponding instructions state, the AAG realised that its relatively simple standards could not possibly capture the whole of the operational environment; thus subordinate/supporting commands were tasked with choosing the standard that they felt was most representative of their situation. In the free-text boxes, the commands were instructed to provide narrative justification for why they chose a particular standard, along with any positive or negative exceptions to their choice, nuances of their command's thinking or situation, etc.

Second, at the bottom of the template was an "overall assessment" free-text field. Commands were instructed to treat this not as a simple "roll-up" of their assessments in each domain (note there was no rating level associated with the overall assessment), but to instead provide their thoughts on the most significant obstacles to future progress, key opportunities for ISAF to act on, or other major items of interest to COMISAF.

Third, commands were not required to rate every task across all four domains. They were told to choose the tasks and domains they wanted to rate. The AAG did not want to force commands to rate tasks or domains if the latter was uncomfortable doing so given their specific mission, nor did the AAG want to restrict commands from providing input to a particular task or domain if they felt they had equities in it.

16. It is worth noting the difference between the standards that the AAG created and the RDLs that were used as part of IJC's District Assessment Model. The latter listed three to six criteria in each line of effort (see footnote 5), which often led to confusing situations for those tasked to provide the assessment input. For example, if the person providing input felt that several criteria should be rated at level two and several others should be rated at level four, that person might "average" the concepts together and submit an overall rating of level three, despite having rated none of the criteria individually at that level. The AAG's standards were designed such that they ideally contained only one criterion per domain (in some cases two were necessary, but no more) to keep those persons providing inputs focused on the most salient aspect of the domain. Any additional nuance or concepts required to understand the situation were solicited via the campaign assessment template shown in Figure 3.2.

Finally, a radar chart (also called a spider or Kano chart) was provided (Figure 3.2, upper right of the template) as a means of visualizing the command's rating levels by domain for each task. These charts consisted of a single axis for each domain, with gradations for the five levels superimposed on them. Plotting the rating levels on these charts provided a qualitative, but standards-based, means of depicting the current status and changes that occurred in each of the domains, for each task, as shown in Figure 3.3.

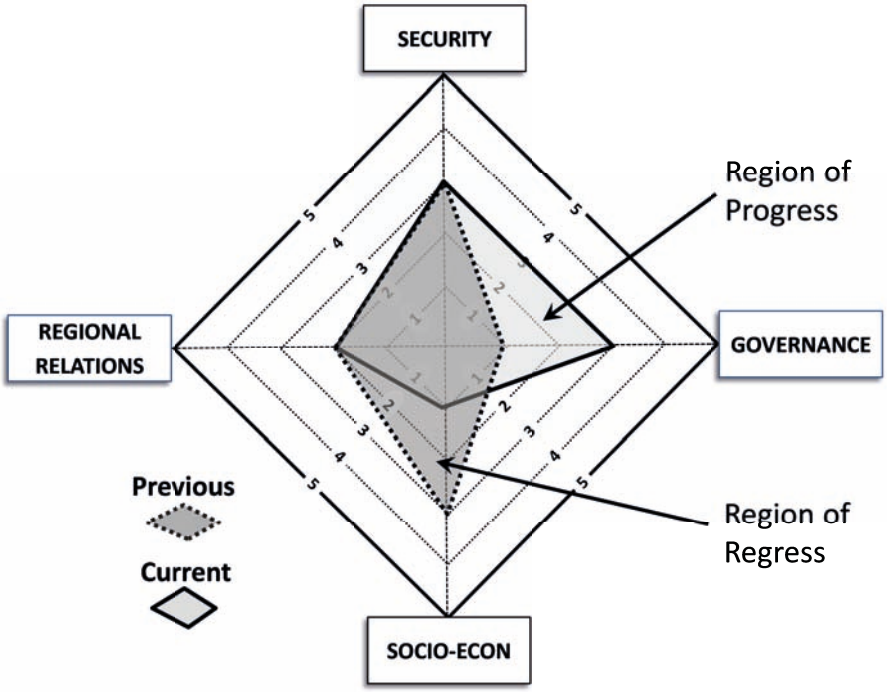


Figure 3.3: Sample progress performance (radar) chart for the campaign assessment

The quadrilateral shape that results from connecting the ratings in each domain can be shaded to provide an immediate visual depiction of the ratings; in this case, more shaded area is better and it is better to have similar ratings for each of the domains (i.e., diamonds are better than misshapen quadrilaterals). Overlaying the current assessment on previous ones quickly identifies areas of progression or regression (Figure 3.3). These charts may be foreign to many commanders, but they were easily understood after only a few showings.

The second input from ISAF's subordinate/supporting commands was less structured, but no less important. This requirement was for a personal assessment from each of ISAF's subordinate/supporting commanders addressed directly to COMISAF. As described by the ISAF Chief of Staff (COS), these personal assessments were to summarize "the heart and mind of the commanders" on their efforts to execute the ISAF OPLAN. This input was required for several reasons. First, it was understood that the first input (ratings of the tasks/domains) would likely be completed by command staffs. This second input gave commanders an opportunity to provide their unfiltered views directly to COMISAF. Second, the personal assessment also helped illuminate differences in views among the commanders as to what was working in the campaign and what was not. These differences could subsequently serve as discussion points at Commander's Assessment Conferences (more on this below). The format for this input was simple: a one- to two-page letter sent directly to COMISAF. Only if the commands did not object were they shared with the AAG.

Outputs

After receiving the commands' rating levels for the domains and tasks they chose to rate along with the commanders' personal assessments, the AAG generated two campaign assessment products. The first was a small set of slides summarizing the command inputs on the OPLAN essential tasks. The second product was a set of rank ordered issues used to help guide discussions in the Commander's Assessment Conferences.

A notional example of a summary slide for one of the OPLAN essential tasks is shown in Figure 3.4. Several aspects of this slide are noteworthy when it comes to the presentation of assessment results. Nowhere in the display of the campaign assessment, or for that matter, in the entire new paradigm, was there any simplistic colour-coding of the results (e.g., using stoplight charts or thermographs). Since colour-coding tends to mask underlying data, a deliberate decision was made to show as much of the data as possible using

other presentation techniques.[17] Figure 3.4 illustrates three different ways of doing this. The radar chart in the upper left-hand corner of the slide shows all of the responses from commands that chose to rate this particular task as multi-coloured dots (i.e., the raw data). The legend boxes to the upper left and upper right of the radar chart associate commands with coloured dots; black command names are for those that chose to rate this task, while command names in grey chose not to do so.

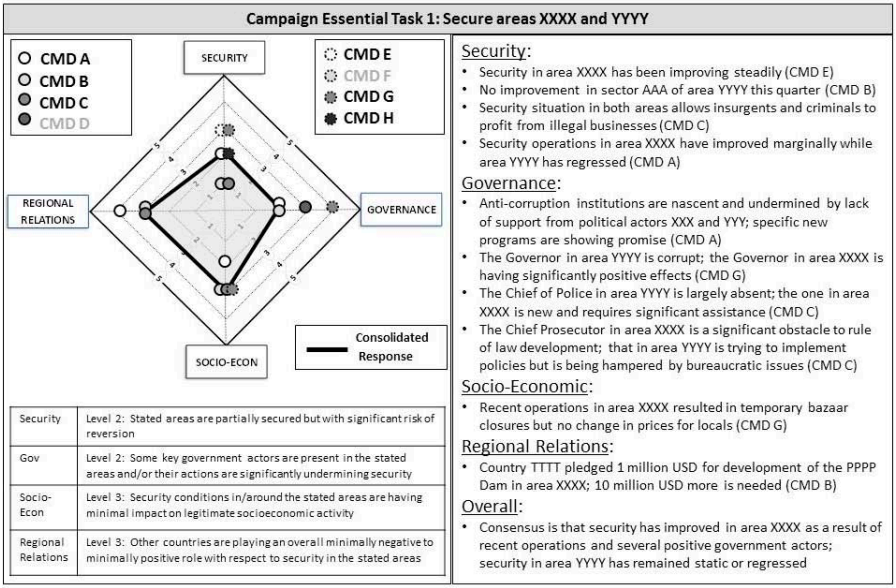


Figure 3.4: Notional campaign assessment summary slide for one campaign essential task

The quadrilateral shape with a red outline (the "consolidated response") shows AAG's rating based on reading and analysing inputs from across the commands. Note that this consolidated response was not 'calculated' (e.g., as

17. An argument has been made that the use of standards along what one reviewer called a "5-point scale" is not inherently different than using a red-amber-green stoplight chart. However, this argument misses several aspects of the new ISAF assessment. While the labels used were numerical, no mathematical computations were ever performed using these labels. More importantly, the labels were never presented in the absence of the wording of the actual standards themselves. As Figure 3.4 shows, the display mechanisms used by ISAF deliberately included the language of the standards to keep the focus on what they said, vice how they were labelled.

an average or median) but rather was 'ascertained' by analysing the totality of the commands' responses against the standards that were developed for each domain of each task. Presenting the data in this way preserved the raw inputs, made clear where various commands agreed or differed, and allowed the AAG to present its view as to where the commands stood as a whole. Having all of these data presented on a single slide stimulated significant discussion in the assessment working groups, since the slides laid bare the similarities and differences amongst the commands. It also helped "keep the commands honest" in their ratings since they knew their inputs would be openly compared to, and challenged by, those from other commands.

Below the radar chart (lower left side of Figure 3.4) are the standards corresponding to the "consolidated response." This was done to avoid the all-too-common problem in operations assessment of arguing over the presentation of data (e.g., colour-coding) vice the data themselves. Having the wording of the standards on the slide allowed the user to quickly review the definitions of the ratings (the complete lists of standards were included as backup files). This helped keep discussions focused on issues identified when comparing command ratings.

The right half of the slide (Figure 3.4) was reserved for comments culled from the completed command templates. The example shows an executive version with a small number of comments by domain. For the assessment working groups, these were stretched to two or three slides to help stimulate discussion. Putting the command comments on the slide again helped preserve and present the raw data behind the assessment. It also showed the commands that ISAF had read, analysed, and digested their inputs and it allowed COMISAF to quickly get a sense for the sentiments of his subordinate/supporting commands.

The second output of the campaign assessment was a set of issues identified via the "overall assessment" portion of the campaign assessment template (Figure 3.2). By asking the question, "what are the challenges to future progress in this task and what is the risk to the NATO core goal in Afghanistan if those challenges are not overcome?" the AAG was able to construct a rank-ordered list of issues based on risk to mission. These were combined with those identified via the strategic assessment (discussed next) and used to drive discussions at the Commander's Assessment Conference, as discussed below.

Strategic Assessment

Inputs

The second component of the new assessment paradigm was a strategic assessment addressing the overall environment in which ISAF was operating and whether or not ISAF was achieving the goals and objectives set by NATO and the U.S. for the war in Afghanistan. To inform the design of this part of the assessment, the AAG undertook a review of all of the strategic documents guiding the war in Afghanistan, both from the NATO and the U.S. chains of command.[18] The AAG constructed a set of the common themes running through these documents to identify the core objectives that ISAF was tasked to accomplish at the strategic level in Afghanistan.

In thinking through how to design an assessment to identify and address progress and setbacks towards the achievement of these core objectives, the AAG had a similar discussion as with the campaign assessment on whether to take a more quantitative or qualitative approach. As before, the AAG chose the middle ground, focusing on a process that was analytic and logical, yet flexible enough to capture and use all of the available information. To ensure the assessment product was analytic and logical, the AAG developed a set of core questions.

These Strategic Questions (SQs), when answered, would tell ISAF whether or not it was making progress toward the accomplishment of the core objectives and goals for Afghanistan. To derive the SQs, the AAG wrote an initial set of questions encompassing the common themes running through the strategic guidance documents and then iteratively combined those questions to generate an ever-smaller number. In so doing, the AAG followed the principle that the resulting SQs should be high-level, focused, enduring, and most importantly, approved by senior decision-makers as critical to ISAF success. Additionally, the AAG intentionally wrote the questions such that they would require elements from several staff sections to answer. The ensuing communication and

18. These included the U.S. White Paper on Policy toward Afghanistan and Pakistan dated March 2009 and available at <http://www.whitehouse.gov/assets/documents/Afghanistan-Pakistan_White_Paper.pdf>, the U.S. Strategic Implementation Plan for Afghanistan, several speeches delivered by the U.S. President, the U.S. Civilian-Military (Civ-Mil) Plan for Afghanistan, as well as the NATO family of plans from the Supreme Headquarters Allied Powers Europe (SHAPE) and Joint Forces Command – Brunssum (JFC-B).

coordination created a "healthy tension" within the ISAF staff that challenged assumptions, fostered new ideas, and identified critical issues that required attention.

There was admittedly more "art" than "science" in formulating the SQs, but the bottom line is that they were questions at the strategic level for which COMISAF needed to know the answers, and needed to be able to answer to his chain of command. There were seven top-line SQs, with a series of sub-questions beneath each one to help further elucidate what the SQ was asking and what information sources might be required to answer it. The AAG also identified lead and supporting Offices of Primary Responsibility (OPRs) for each top-line SQ. While the actual SQs remained classified as of this writing, they were akin to the types of questions COMISAF was asked during Congressional testimonies or in meetings with the North Atlantic Council (e.g., "Can the Afghan National Security Forces secure the country?"). Importantly, each set of sub-questions under the Strategic Questions ended with the following pair:

- "What are the primary challenges to future progress and what is the risk to the NATO core goal for Afghanistan if these are not overcome?"
- "Looking forward six months, do you anticipate continued progress, setbacks, or no change to the current situation?"

The reasons for asking these two sub-questions will become evident in the discussion of issues below.

Since ISAF is the theatre strategic headquarters, the AAG determined it would be most appropriate for the ISAF staff to answer the SQs and, based on arguments in the literature, the AAG decided that the answers should be in narrative form with supporting data (Connable 2011; Downes-Martin 2011; Schroden 2011). This decision was again made to allow the staff the flexibility to answer the questions using their subject matter expertise. That said, to minimise the submission of impertinent information and opinions, the staff was instructed not to submit assertions in the absence of supporting facts and to ensure that answers were logically-reproducible, meaning the reader should be able to use the data or information provided to arrive at roughly similar conclusions. Thus, an overview of the strategic assessment was that it

consisted of the ISAF staff answering a set of SQs in analytic, narrative form using whatever data or information were necessary, available, and relevant.

Outputs

The AAG took the staff's responses to the "challenges" question mentioned above and used them to identify a set of issues and associated risks to mission. These were combined with the issues identified via the campaign assessment to generate a comprehensive list of issues that were rank-ordered according to risk to mission.[19] These issues were then reviewed at the staff level to ensure they were described correctly and had the right level of risk associated with them. The list was then presented to the ISAF COS and subordinate/supporting commands' COSs for review, discussion, and decision on which issues would be discussed at the ISAF Commander's Assessment Conference. During the first iteration of the assessment in January 2012, eleven issues were identified as being of extremely high or high risk to the NATO core goal in Afghanistan; of these, the ISAF COS directed the top five be presented for discussion at the conference.

Assessment Process and Timeline

The process used to take the assessment inputs and turn them into outputs as described above was relatively straightforward. Sixty days in advance of the Commander's Assessment Conference, the AAG gave the ISAF staff and subordinate/supporting commands warning that the data call for the assessment would occur in two weeks' time. Forty-five days in advance, the AAG issued a fragmentary order (FRAGO) to the staff and commands to answer the Strategic Questions and provide ratings and justifications for the OPLAN's essential tasks, respectively. Thirty days in advance, the assessment inputs were due to the AAG and analysis of the inputs began.

The remaining month proceeded as follows. The first week was for ISAF internal discussions in the assessment working group (AWG; consisting of

19. To identify the level of risk, the AAG used the risk assessment framework articulated in the NATO Comprehensive Operations Planning Directive (COPD), V 1.0. The COPD identified four categories of risk as follows: extremely high (could result in failure to accomplish the mission); high (could result in failure to accomplish one or more objectives); moderate (could result in failure to meet criteria for success or exceed time, space, forces/actors limits); and low (minimal impact on mission accomplishment).

action officers) and in the campaign management working group (CMWG; Colonels and one-star General Officers). These working groups culminated in the COS Fusion Meeting, hosted by the ISAF COS and attended by the ISAF two-star directorate heads (DCOSs). These meetings focused on the ISAF staff's responses to the SQs and associated issues. The second week consisted of an AWG and CMWG attended by both ISAF staff and representatives from the subordinate/supporting commands. These working groups culminated in the COS Integration Meeting, hosted by the ISAF COS and attended by ISAF staff representatives plus the subordinate/supporting commands' COSs. Each of these meetings was focused on discussion of the commands' ratings for each OPLAN essential task and associated issues (the outcome of the COS Integration Meeting was the rank-ordered list of issues described above). The AAG took the third week of the month to draft the assessment report and prepare for the Commander's Assessment Conference. During the fourth week the report was published and the conference was held. The first week after the conference was used to write and issue a FRAGO directing the actions that were decided at the conference and the second week after the conference was used to run an after-action review (AAR) to identify areas for improvement in the next quarterly assessment.

Assessment Products

The new assessment paradigm produced a number of intangible products, not the least of which was constructive discussion, and in some instances, debate, over the critical issues that ISAF needed to address to be successful. It also stimulated critical and creative thinking about how best to address them. In terms of tangible products, in addition to the briefs, slides, scripts, and summaries that were generated for the AWGs, CMWGs, and COS Fusion and Integration meetings, there were two primary assessment products.

Quarterly Strategic Assessment Report

Upon receiving the answers to the SQs from the staff, the AAG analysed and consolidated them into a single document that would eventually form the core of the Quarterly Strategic Assessment Report (QSAR). The QSAR was the ultimate written product of the assessment paradigm and fulfilled a

reporting requirement to Joint Forces Command – Brunssum (JFC-B; courtesy copies were also sent to U.S. Central Command). It contained a preface from COMISAF (his personal assessment), a short introduction explaining the assessment process and how the report was prepared, narrative responses to each of the Strategic Questions, and a concluding section identifying the top issues and opportunities for ISAF during the quarter, requests for support from higher headquarters, and a look ahead six months at the strategic calendar and operational environment. The QSAR was approved and signed by COMISAF.

ISAF Commander's Assessment Conference

The ISAF Commander's Assessment Conference was hosted by COMISAF and attended by ISAF senior staff (Deputy and COS), the NATO Senior Civilian Representative (SCR), the ISAF two-star DCOSs, and ISAF's subordinate/supporting commanders. Given General Allen's preference for discussion vice presentation, a typical agenda for this conference consisted of an hour-long summary presentation of the results of the assessment, followed by 4-5 hours for discussion of the results, and in some cases a final hour to discuss long-range planning.

To facilitate this discussion, the issues were summarized (i.e., nature of the issue and how it was identified) and discussion points were developed. The latter were generated using the formal inputs provided by the ISAF staff and subordinate/supporting commands, discussions within the AWG/CMWG/COS meetings, and discussions within the AAG. They were designed to drive action by focusing on four areas over which COMISAF has control: changes to the OPLAN; changes to the allocation of resources; changes to organisational structure; and suggested changes to policy or requests for support.[20] The discussion of these points was moderated by the ISAF DCOS for Operations, and detailed notes were taken of the discussion to record decisions and actions directed by COMISAF. In the wake of the conference these decisions and actions were captured in a FRAGO that tasked the ISAF staff and subordinate/supporting commands with implementation.

20. Subsequent AARs identified a fifth category over which COMISAF has control: changes to strategic messaging.

Assessment of the Assessment

As the history of ISAF assessment paradigms illustrates, each one is tailored to fit a number of constraints to include the decision-making style of the commander and the resources available. The AAG's goal in designing the paradigm described above was to support General Allen's ability to understand changes in the strategic environment and the pace and direction of his campaign, and to make effective decisions accordingly, and in that it was successful.[21] That said, no assessment paradigm is perfect, so for the sake of future assessment efforts and revisions to assessment doctrine, it is worth discussing the advantages and disadvantages of the new paradigm from a more general view.

Advantages of the New Paradigm

The new assessment paradigm, by virtue of its move away from a centralized list of core (quantitative) metrics and towards a more open, narrative assessment based on a holistic examination of available data, opened the door for deeper analysis and increased the level of transparency of the assessment. This resulted in part from allowing the staff and commands to include not only the pertinent information they had at their disposal but also to include the nuance and context that needs to accompany such information. The new paradigm helped balance the past and present status of the war with maintaining an eye on the future. By consistently asking the ISAF staff and commands to address the current state of the war along with challenges to future progress and opportunities that could be exploited, COMISAF was provided with the key issues (those posing the most risk to mission or presenting the greatest possible gains) that required his attention.

Another positive aspect of the new paradigm was how it fostered cross-collaboration among the commands and ISAF staff. The working groups gave people the opportunity to defend their positions on issues, to challenge stated and unstated assumptions, and to reach workable compromises. The "healthy tension" that was created by explicitly discussing differing viewpoints directly

21. General Allen's satisfaction with the assessment paradigm in 2012 was evident by comments he made to that effect in several Commander's Assessment Conferences as well as empirically by his continued use of the paradigm throughout his tenure as COMISAF.

contributed to the quality of the assessment products, since it propelled critical thinking about what was working in the campaign and what was not.

Finally, the focus of the new paradigm on both the strategic and campaign levels helped keep the discussion at the level appropriate to ISAF, vice getting into the details of the tactical fight. The move towards a more narrative, standards-based approach kept the discussion centred on substantive and actionable items.

Disadvantages of the New Paradigm

As mentioned above, the AAG was not resourced nor empowered to conduct a completely independent assessment, and so the new paradigm is very much dependent upon inputs from the ISAF staff and subordinate/supporting commands. Some might criticize this aspect of the new paradigm as allowing those tasked to execute the OPLAN with "grading their own homework." While there is some truth to this, it is also worth recognising that there is often a difference between the quality of information and level of candour that is typically afforded by staff and commands to an independent assessment group vice to an internal group that allows the staff and commands to be intimately involved in the process. Additionally, the AAG attempted to mitigate this via the "triangulation" of results by asking the same questions of multiple staff sections and commands. This ensured that whenever possible no single staff section or command was driving the assessment's conclusions. The AAG also reserved the right to make its own assessment based on the inputs received, though it realised the importance of retaining and presenting command and staff inputs as part of the final product.

A second drawback of the new paradigm is that since it relies on inputs from a variety of sources, the analytic quality of those inputs may vary considerably. One lesson learned during the first iteration of the new process was that it is important for the assessment group to establish close relationships between its analysts and those providing inputs, to help the latter understand the analytic quality needed.

Finally, another disadvantage of the new paradigm is that the ultimate quality of the assessment products is highly dependent on the skills of a few people. It is not mechanistic, and as such requires strong analytical, critical thinking,

briefing, and writing skills in the assessment team; strong personalities in the fusion/integration roles (e.g., the COS and AWG/CMWG leads); and a strong discussion moderator at the Commander's Assessment Conference. If the people performing these tasks do not have the critical skills required, the quality and impact of the assessment products will suffer; and simply adding more people without the requisite skills will not alleviate the problem.[22]

Conclusion

In 2009, ISAF recognised a formal requirement to resource and conduct meaningful assessments of its operations in Afghanistan, and since then it has placed more emphasis on, and derived more value from, the assessment function. During the timeframe 2009-2012, the ISAF assessment paradigm underwent significant change a number of times. Depending on one's point of view, these changes may have been for the better or the worse, but in all cases they were tied to the preferences of COMISAF and the purposes ascribed to the assessment function. Some Commanders preferred an approach that sought to quantify progress and communicate it in numerical terms; others favoured a "softer" approach that focused on commander's estimates and the judgments of subject matter experts. The primary shift in thinking that occurred as a result of General Allen's guidance to the AAG was a move away from the doctrinal view of operations assessment as being designed to "measure" progress and towards a view of it as an activity to enhance the commander's ability to understand critical issues and make sound decisions on how to address them.

That shift in thinking led the AAG to design an assessment paradigm in 2012 that was innovative, certainly with respect to the published literature on assessments, but also with respect to unpublished accounts of previous efforts

22. As a point of reference, the AAG typically relied on a small team of 3-4 people to run the assessment process described in this chapter. The team was supported by the other sections of the AAG (data and survey analysis teams) and required some administrative support (to set up meetings and teleconferences).

in Afghanistan and Iraq.[23] Some of the characteristics that made it unique include: the use of SQs; the use of descriptive standards; a focus on narrative, analytic assessments leveraging both quantitative and qualitative data; the use of radar charts as a display technique; a focus on cross-collaboration, discussion, and creating "healthy tension"; and driving decisions through the use of discussion points framed by the actions a commander can take, direct, or request.

While the unique elements described above were created for a specific commander and a specific command at a specific point in time, many of them could be generalized and I believe they should be studied as part of a broader effort to revisit NATO and US assessment doctrine. The latter especially has been tested over the past decade of war in Iraq, Afghanistan, and elsewhere and has been found wanting; though this conclusion has been reached mainly by those who have tried to implement that doctrine and has not yet permeated the organisations responsible for writing it.[24] Indeed, while NATO has begun thinking about what revisions it might need to make to its assessment doctrine (going so far as to schedule a military experiment on assessment to address these issues), recent discussions with some doctrine-writing entities in the US have found them blissfully unaware of any problems with current assessment doctrine. It is my sincere hope that this chapter, and this book, will serve as a call to action for these entities and a source of innovation for the revolution in assessment doctrine that is so urgently needed.

23. Discussions with officers who served in the assessments cell at Multi-National Forces – Iraq (MNF-I; the ISAF equivalent for Operation Iraqi Freedom (OIF)) in the 2007-2009 timeframe revealed that while MNF-I had a similar paradigm in many respects to that described in this chapter, there also were some significant differences. For example, it still had as its focus an emphasis on trying to "measure" progress and it allowed "LOO owners" (i.e., individual staff elements) to control aspects of the assessment, as opposed to focusing on enhanced decision making and improved collaboration within the staff.

24. A working symposium on multinational assessments that was organised by the Military Operations Research Society (MORS) in November 2012 came strongly to the conclusion that US (and to a lesser extent, NATO) doctrine on assessment was flawed in numerous ways and was in critical need of being re-written. Unfortunately, the attendees of this symposium were largely practitioners of assessment and not those responsible for writing assessment doctrine.

References

Agoglia, J., M. Dziedzic, and B. Sotirin. 2010. *Measuring Progress in Conflict Environments: A Metrics Framework*. Washington, D.C.: U.S. Institute of Peace.

Campbell, J., M. O'Hanlon, and J. Shapiro. 2009. How to Measure the War. *Policy Review* 157:15–30.

Connable, B. 2011. *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*. Santa Monica, CA: RAND Corporation.

Downes-Martin, S. 2010. *Assessments Process for RC(SW)*. Unpublished draft.

——— 2011. Operations Assessment in Afghanistan is Broken: What Is to Be Done? *Naval War College Review* Autumn:103-125.

Exum, A., N. Fick, A. Humayun, and D. Kilcullen. 2009. *Triage: The Next Twelve Months in Afghanistan and Pakistan*. Washington, D.C.: Centre for a New American Security.

Headquarters, Department of the U.S. Army. 2006. Field Manual (FM) 3-24*, Counterinsurgency.*

——— 2010. FM 5-0*, The Operations Process.*

Kilcullen, D. 2010. Measuring Progress in Afghanistan. In *Counterinsurgency* 51-76. New York: Oxford University Press.

LaRivee, D. 2011. Best Practices Guide for Conducting Assessments in Counterinsurgencies, *Small Wars Journal* 17 August:1-34.

McChrystal, S. 2009. *COMISAF's Initial Assessment*. Unclassified redaction, available at <http://www.washingtonpost.com>

North Atlantic Treaty Organisation (NATO). 2011. *NATO Assessment Handbook*. Interim Version 1.0. Norfolk, VA: HQ Supreme Allied Command Transformation.

O'Hanlon, M., and H. Sherjan. 2010. *Toughing It Out in Afghanistan*. Washington, D.C.: Brookings Institution.

Schroden, J. 2009. Measures for Security in a Counterinsurgency. *Journal of Strategic Studies* 32:715–44.

——— 2010. Fixing Assessments in Afghanistan. Unpublished but available to U.S. government entities upon request.

——— 2011. Why Operations Assessments Fail: It's Not Just the Metrics. *Naval War College Review* Autumn: 89-102.

Schroden, J., R. Thomasson, R. Foster, M. Lukens, and R. Bell. 2012. A New Paradigm for Assessment in Counterinsurgency. *Military Operations Research Journal* (accepted for publication).

Soeters, J., P. C. van Fenema, and R. Beeres, eds. 2010. *Managing Military Organisations: Theory and Practice*. New York: Routledge.

Upshur, W., J. Roginski, and D. Kilcullen. 2012. Recognising Systems in Afghanistan: Lessons Learned and New Approaches to Operational Assessments. *Prism* 3:87-104.

U.S. Joint Staff. 2010. Joint Publication (JP) 1-02, Department of Defense Dictionary of Military and Associated Terms.

——— 2011a. JP 3-0, Joint Operations.

——— 2011b. JP 5-0, Joint Operation Planning.

4

Public Opinion Polling in Operations Assessment with Examples from Afghanistan

Dr. Katherine Banko

Abstract

In recent years, operations assessments have become much more comprehensive, are performed using a variety of data collection methods and data analysis techniques, and often incorporate perceptions of the local populace. Collecting information through the use of questionnaires is challenging, even for experts on the topic, as there are numerous opportunities for error to enter into the process. Because of this, caution must be given when interpreting the results and using the findings for operations assessment. Polling results are a snapshot of a population sample at a specific point in time. As such, results should always be understood within a social and temporal context, and preferably used in combination with other data sources that corroborate the findings. When conducted and used properly, polling can provide commanders with enhanced situational awareness and ultimately become an important component to a comprehensive assessment of a military operation. This chapter reviews basic concepts for conducting surveys with a focus on how errors can occur and how they may be avoided.

Introduction

Never before in military operations has there been as great an emphasis placed on understanding public attitudes. One reason may be the emphasis on fighting counter-insurgencies. The phrase "winning the hearts and minds" has become a popular expression among world leaders with regard to United States (U.S.) foreign policy and more importantly within counter-insurgency (COIN) operations. The U.S. Army (2006) released a counterinsurgency manual

that defined the phrase "hearts and minds."[1] "Hearts" is persuading people that their best interests are served by COIN success. "Minds" is convincing them that the allied force can protect them and that resisting it is pointless. Both effects involve persuading attitudes of the local populace. Also in 2006, the Chief Strategist of the Office of the Coordinator for Counterterrorism of the U.S. State Department released a framework for interagency cooperation in COIN operations that stressed the importance of public perceptions in the host nation of allied, local government, and insurgents in gaining influence over population groups (Kilcullen 2006). The framework helped people and organisations understand where their efforts fit into a campaign and provided a basis for measuring progress. In 2011, the North Atlantic Treaty Organisation (NATO) released doctrine emphasising that success in a COIN environment was intricately tied to understanding the attitudes of the people (NATOa 2011). Specifically, successful COIN operations hinged on influencing attitudes in the host nation. Indeed, in the early days of attitude research (early 1900s), it was generally accepted that human behaviour was guided by social attitudes and that understanding these attitudes was key to changing human behaviour (Thomas and Znaniecki 1918; Thurstone and Chave 1929).[2]

Incorporating public opinions into campaign assessments has been gaining popularity, most recently in Iraq and Afghanistan.[3] In Afghanistan, several nation-wide polls have been conducted by International Security Assistance Force (ISAF), news media, and non-government organisations. One of the longest running annual polls is conducted for The Asia Foundation (2012) and has been fielded for seven consecutive years.

Other polls have been sponsored by the American Broadcasting Corporation, British Broadcasting Corporation, Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland ("Consortium of public-law broadcasting institutions of the Federal Republic of Germany")

1. US AFM 3-24 COIN doctrine was used earlier in reference to UK COINC campaigns in the 1050s (e.g., Stubbs 2008).

2. There is a long-standing debate in the field of social psychology concerning the ability to predict attitudes. Current methods are usually found to be poor predictors of actual behaviour. In fact, it is often found that people say one thing, yet do another.

3. Public opinion research is widespread; however, it is the explicit application to campaign assessments that is a recent development.

and the Washington Post (ABC/BBC/ARD 2010), the International Republican Institute (2009), the Law and Order Trust Fund for Afghanistan (2011), Ministry of Counternarcotics of the Government of Afghanistan (2012), and Oxfam (2009).[4] Provincial polls have also been commissioned by several ISAF subordinate units (e.g., Regional Command South and Regional Command East headquarters). The Afghanistan Nationwide Quarterly Assessment Report (ANQAR) poll for ISAF Joint Command continues to be conducted at the time of this report.

In recent years, several traditional social science techniques have been used to collect perceptions of public attitudes at the village, district, and national levels to support campaign planning and assessments. Examples include focus group methodology (Kitzinger 1994) and the Tactical Conflict Assessment Planning Framework (U.S. Centre for Army Lessons Learned 2010). The most widely used, especially for understanding the host nation at the district and national level, is opinion polling.

This chapter will provide an overview of basic concepts associated with survey research. It will review errors with examples drawn mostly from Afghanistan that occur due to a) respondents and the context in which the survey is administered; b) design of the data collection instrument (questionnaire); and c) issues that arise during data collection in the field. It will conclude with specific recommendations for controlling for some of the problems and suggest how the information should be used to support campaign planning and assessments.

Basic Concepts of Survey Research

Survey research encompasses any systematic measurement procedures that involve asking questions of respondents. Questionnaires may be delivered through pencil-and-paper means, (although more recently are often administered through electronic means via the internet) that respondents complete without assistance. Questionnaires can also be administered orally through an interview with the interviewer completing the questionnaire based on answers provided by respondents. In western societies, survey research is often accomplished through telephone contact. However, in-person interviews are

4. There are many other surveys concurrently conducted in Afghanistan by non-government organisations; only those that are specifically relevant to the current conflict are mentioned.

the practice that has most commonly been used to collect information for the purpose of campaign assessments, mainly due to low literacy rates and a lack of land-based telephone lines in Afghanistan and Iraq.

It is important to understand that information collected through surveys is referred to as "self-report" data. Self-report data are data collected directly from the research participant, for example, a person being interviewed in a face-to-face situation. It is necessary to understand the distinction between self reporting of covert events, i.e., those emanating from inside the person such as an attitude toward some object and of overt events i.e., those associated with reporting of past and present behaviours. Attitudes, beliefs and values, and covert events, cannot be directly observed and are always inferred from verbal statements (Guerin 1994; Nisbett and Wilson 1977).

Reporting of both types of events, overt and covert, introduces errors into the measurement process that have implications for the interpretation of the findings. There errors are often compounded due to the in-person administration of the questionnaire. In-person interviews have unique challenges associated with them due to the effects of the interaction between the interviewer and the respondent, as well as the reduction of anonymity and confidentiality for the respondent. These challenges will be discussed within the greater context of designing and administering an opinion poll (survey).

Accessibility

A necessary starting point for conducting an opinion poll is knowing the set of elements (persons) one must draw on to extract an appropriate sample. Non-permissive environments such as military conflict zones present unique challenges for both the sampling of the population and the administration of the survey. In some places, enumerating the population may be difficult due to high rates of migration (people leave when times are most terrible and return at less volatile times), nomadic rural groups, and refugees and international displaced persons (IDPs) from neighbouring countries. IDPs are not unique to Afghanistan.

Polling conducted in Pakistan, Palestine-Jordan, Sudan, and Central America, also experience this problem (Berg 1988). In addition, there are additional inaccessible portions of the population, including people who live

in remote areas that are a) not easy to get to due to lack of roads; b) under the control of insurgent forces; and/or c) persons living in internment camps or residing in other institutions such as hospitals or prisons. This has proven to be difficult in Afghanistan. As in many conflict zones and non-permissible environments, this value is unknown. Knowing the total population in an area is important for determining a sample that is robust enough to make generalisations from it to the population.

Sampling

Sampling is the process of selecting people from a population of interest so that by studying the sample you can fairly generalize your results to the population from which the people were chosen. External validity is the degree to which the conclusions of an opinion poll would hold for other persons in other places and at other times. In order to improve external validity, researchers employ a sampling strategy and determine an appropriate sample size based on the size of the population of interest. There are numerous books on sampling design that a researcher can reference to help minimise the risk of an inappropriate or biased sample (Alwin 2007; Lohr 2010).

Opinion polling can be used to draw inferences over time via longitudinal comparisons. There are two ways of designing time series designs,[5] a panel design and a repeated cross-sectional design. In a cross-sectional design, a different sample of the target population is selected for each iteration (or wave) of data collection. In a panel design, the same participants are followed over multiple survey rounds or waves. In other words, it allows for the tracking of opinions of the same people over time and is less subject to sampling error.

One must be cautious when drawing conclusions on longitudinal trends, especially when the respondents vary with each survey. For example, a recent report produced for The Asia Foundation makes longitudinal comparisons for a six-year period using a cross-sectional design. These types of designs are used frequently in Afghanistan as accessible areas change over time (for example, districts may become too dangerous for interviewers to enter due to control by Taliban). Interpretations of trends from cross-sectional time series designs can be misleading due to poor sampling.

5. For a comprehensive treatment of longitudinal designs, see Lynn 2009.

Respondent Errors

Regardless of the type of event (overt vs. covert), respondents may be unable or unwilling to report accurately on information for many reasons. For instance, respondents in some surveys may be asked to provide information on the frequency of some event. This type of question is usually stated in the form "in the last year, how often have you been a victim of a crime?" Errors in recalling past behaviours/events generally can be the result of simple memory decay (forgetting due to the passage of time) or may be due to not being able to retrieve the required information from memory (Sudman and Bradburn 1973). In addition, the direction of the errors will vary depending on the length of the referent (Brennan et al. 1996).

For a long period of time, such as a year, respondents are susceptible to an error termed 'recall decay.' That is, they tend to under-report the number of events in the specified time period. In contrast, when very short periods of time are offered as reference periods, respondents tend to over-report the number of events referred to as 'telescoping.' One way to avoid recall decay and forward telescoping errors is by bounding the reference period by some meaningful event (Sudman, Finn, and Lannon 1984). In the Afghan context, such an event could be the poppy harvest season.

Social desirability bias

A basic aspect of our social existence is self-presentation, whereby we present who we are, or who we want others to believe we are through our words and actions (Goffman 1959). Self-reports, then, can be biased towards the audience present during the interview and are based on their perception of what is "correct" or socially acceptable (Campbell 1950). This phenomenon is called social desirability bias and has been found to occur in virtually all types of self-report measures across the social science literature (Nerderhoff 1985).

In Afghanistan, asking questions through in-person interview techniques compounds this error due to the common practice of having family members of the households present during the interviews. There are often anywhere from 2 to 15 people present during an interview. Essentially, there is no anonymity for the respondent; consequences for responses that fall outside of social norms, particularly for female respondents, are likely to be severely punitive.

For example, in a poll conducted for The Asia Foundation in the fall of 2012, in an early question more than half of the respondents reported that they would feel uncomfortable publically criticising the Government of the Islamic Republic of Afghanistan (GIROA). Following this, a series of items (a question is an item) concerning satisfaction with the performance of the local government were posed to respondents. It is not surprising that 80% of samples reported that overall they felt that the government was doing a "good" or "very good job."

There are a few ways one can avoid social desirability bias (Gordon 1987; Nederhof 1985). One such technique uses indirect questioning. Questions phrased indirectly, asking for their perceptions of the opinions of others rather than the respondent's opinion, may allow the respondent to project his or her true feelings about socially-sensitive issues (Sherwood 1981; Fisher 1993). For example, this technique was employed to discern whether people supported the Taliban. In early waves of a Kandahar provincial survey, respondents were asked directly if they supported the Taliban.

Results from this question showed that there was relatively low support for the insurgency. In contrast, when people were asked if they felt if others in their community supported the Taliban, there was a marked reversal in the responses with upward of 70% of respondents reporting they felt that there was support in their village. Shifting the questioning from direct to indirect can result in dramatically different responses. However, when using this technique to control for social desirability bias it is important to understand that one no longer has the opinion of the respondent, but what the respondent thinks might be the views of others.

Non-response bias

This bias occurs when an individual refuses to respond to one or more questions (Berg 2005). It is a type of bias that occurs if the views of those who respond to the survey (answer the question) differ from those who do not respond, resulting in a viewpoint being under-represented. This can impact the ability to generalise the results of the sample to the population. For some questions on many of the Afghanistan opinion polls, there were high incidences

of non-response.[6] One reason people may have chosen not to respond to items may have been due to a perceived fear of retribution from GIROA, insurgents, or even allied forces.

For example, across polls, high non-response rates were observed for questions on whether children, in particular female children, attended schools, whether there was support for insurgents, or whether ISAF treated people with dignity and respect. This phenomenon may be an additional source of error with important implications regarding the validity of the results.

Designing the Questionnaire

Questionnaire design has important implications for the usefulness of the results. Questionnaire design cannot be taught directly from books; every investigation presents new and different problems. Textbooks and guides, however, can help prevent some of the worst pitfalls. It is highly recommended that experts in questionnaire design be involved from the inception of a polling project. This section will give the reader a rudimentary appreciation as to how question design can affect the survey results.

Question Ordering Effects

Question order effects occur when responses are influenced by the question's placement in a survey (Bradburn and Mason 1964; Sudman and Bradburn 1974). Care must be taken when ordering the items on the questionnaire in order to avoid 'leading' questions (Lassiter, Stone, and Weigold 1987). Respondents are also often asked to express their overall satisfaction with provision of services of the government, then are asked their satisfaction with a list of items ranging from supply of basic services such as electricity to satisfaction with local courts and other developmental projects. The overly positive results on each of the items can be the result of positive response to the question of overall satisfaction. This phenomenon is referred to as acquiescence bias whereby respondents have a tendency to agree with all the items or to indicate a positive

6. It is difficult to know what the exact non-response rate was due to the contractor combining different responses into one category (i.e., combined "refused to answer," with "do not know the answer," and other responses that the contractor felt belonged together).

nuance (McClendon 1991; Messick and Jackson 1961). Care must be taken when designing an instrument to avoid these simple but common mistakes.

Question Wording and Translation

One of the major difficulties in designing a good instrument is getting the wording correct for each question. The choice of words is critical in expressing the meaning and intent of the question to the respondent and ensuring that all respondents interpret the question the same way. Even slight wording differences can confuse the respondent or lead to incorrect interpretations of the question (Shwarz 1999). Entire books have been written on the subject (Converse and Presser 1986; Fowler and Fowler 1995). This section will focus on question wording and its relationship to translation of surveys administered in multiple languages.

When conducting surveys in non-permissive environments, translation issues can arise in several ways including the initial translation of the instrument into the language of the host nation, and the coding of responses by interviewers in the field which eventually is translated again when producing the final product. Since question wording can have profound implications for the results, it is important to rigorously test items before administering a questionnaire. Since translation is part of the work contracted, researchers should, at a minimum, have questionnaires back-translated by an alternative translation source prior to fielding.

In Afghanistan, surveys have typically been translated into the two predominate languages of the country, Pashto and Dari. The translation for both stages has been done by contractors as a component of the overall contract which typically includes training of interviewers, fielding the surveys, and quality control. While there are standardised methods for translating surveys into multiple languages (Forsyth et al. 2006; McKay et al. 1996), little is known about how this has been done in Afghanistan.

Data Collection Issues

One of the most challenging aspects of conducting opinion polling in non-permissible environments occurs during the data collection phase.

Interviewer Errors

Interviewing is a complex survey operation since it involves a cognitive and social interaction between the interviewer and the respondent. There are two main schools of interviewing, conversational and standardised. With standardised interviewing, all respondents should receive the same delivery of the questionnaire (tone, emphasis on response categories, and so on) so that each interviewer's influence on the response is minimised. Conversational interviewing, also known as flexible interviewing, allows interviewers to ask respondents if they did not understand a question and provide unscripted feedback to clarify the meaning of questions. A key distinction between standardised and conversational interviewing is that standardisation requires the interpretation of questions to be accomplished entirely by respondents (Fontana and Frey 2005). The literature discussing different aspects of the interview process is extensive and beyond the scope of this paper. Our focus will be on the standardised interviewing process as this is the approach used in Afghanistan.

Interviewers may contribute to total survey error in a number of ways. Interviewers vary in their ability to gain cooperation and counter non-response bias. Typically, respondents that are interviewed by the same person tend to provide similar responses. The result is a clustering effect by individual interviewer (Kish 1962). Reasons for the clustering effect include a) interviewers rewording questions, b) interviewers using different probing techniques, and c) interviewers using different strategies for accepting a "I don't know" answer, to name a few. These errors are a result of standardised administration procedures not working correctly.

The importance of acquiring competent research firms to field a survey in a conflict zone is very important. Different surveying companies will have different training procedures which will ultimately affect the quality of the data and how much trust one can invest in the results. The company's training program for standardised interviewing techniques should be an important consideration when selecting a contractor. In addition, in a conflict environment one may want to give preference to a contractor that strives to employ local nationals who have affiliations to particular districts and villages (i.e., interviewers who live in a district or have relatives or friends in the area). This

practice overcomes two issues. It increases the likelihood that respondents will open the door for the interviewers and then continue to cooperate with the interview. More importantly, it may help ensure the safety of the interviewers.

In Afghanistan the dangerous environment limits the effectiveness of the interviewers. There is often no accessibility to households due to few or no paved roads. Modes of travel are primitive; cars are rare and gasoline is expensive. Sometimes work is done by means of foot, camel or donkey and carts for longer distances. Roads are dangerous to use due to the placement of improvised explosive devices and often require the payment of a "user fee." Data are often collected only during midday hours when it is the 'least dangerous' to be out. Sometimes contractors are unable to fulfil the full requirements of a contract due to the extremely dangerous conditions in some villages. Method reports have stated that interviewers frequently have to hide from insurgents and hire locals for protection and to get access to the population. In the end, the sophisticated sampling frame and methodology have to be discarded and replaced with ad hoc procedures collecting information from passers-by who would talk to interviewers. Being seen in public with a briefcase full of papers (i.e., completed survey forms) can be life threatening. As a result of these conditions, interviewer turnover is very high and can impact standardized interviewing. As a result of the challenging conditions, strong threats to validity can occur requiring careful analyses.

Interpreting Polling Results

This section contains three cautions concerning the use of polling data.

Questionable Inferences

One can make a broad distinction between two types of survey: the purely descriptive, enumerative, census types of survey (the type typically conducted for campaign assessments), and the analytical, relational type of survey, designed to show relationships between variables (Beatty 1995). Public-opinion polls are usually of the descriptive variety but are often used to make predictions by comparing results of surveys at different times and producing a trend. For example, in comparing the results of public opinion surveys conducted with Afghans, we see a decrease over time of reporting crimes by the victim.

In the interpretation of the results it was suggested that "the decline in reporting crime or violence to an authority might be due to a decrease of citizen confidence in government conflict and dispute resolution mechanisms/ institutions" (The Asia Foundation 2012, 41). However, the polling results showing a decrease in crime reporting is not sufficient to suggest that these two variables are related and therefore, the conclusion may be inappropriate. Polling data cannot show causal relationships.

Presenting Incomplete Evidence

One of the most serious misapplications of polling data occurs when users fail to look at the entirety of a poll and instead 'cherry pick' results that support their view of a situation or suppress evidence that does not confirm a particular position. One reason this occurs is due to pressure on analysts and operational and tactical headquarters staff tasked with completing spreadsheets that list hundreds of metrics. The results are turned over to higher strategic headquarters that, in turn, use some of the items for other reporting purposes that further distort the 'true' measure; single items selected from a poll usually cannot provide correct and therefore useful information.

Importance of the Social Context

Finally, it should be noted that polling data provides only a snapshot of the population sample at a particular point in time. Events that precede data collection will undoubtedly affect how people respond to questions. For example, a poll conducted in Kandahar province in September 2009 showed marked changes in respondents perceptions of safety; many people reported that they felt much less safe compared to a poll conducted just three months earlier. Just weeks earlier, an explosion had devastated blocks of houses killing 43 and injuring at least 65. Polling results must always be understood within a social and temporal context.

Applying Polling Results for Campaign Assessment

Producing assessments for operational and higher-level leadership on the missions in Afghanistan has become an increasingly important and challenging task. Many articles have been written discussing measures of performance

(MOPs) and measures of effectiveness (MOEs) (NATO 2011a). Numerous complicated frameworks and dashboards have been developed to determine what to measure and how to quantify abstract concepts such as safety and security (U.S. Institute of Peace; U.S. Army Corps of Engineers; and U.S. Army Peace Keeping and Stability Operations Institute 2008). Still, there is little agreement on what should be measured, how it should be measured, and how best to report it.

In recent years, several special task groups have been devoted to this specific problem, i.e., NATO System Analysis and Studies-091 (NATO 2011b) and NATO Human Factors and Medicine-185 (NATO 2012). Most recently, a special meeting of international measurement and assessment experts was held in November 2012 and hosted by the U.S. Military Operations Research Society. While some of the major data collection and data storage issues have been addressed as a result of these efforts, there still remain many gaps and unknowns with respect to what to count, how frequently to count it, who should count it, and why it should be counted at all. Because of the challenges, there is a need for identifying good practices for opinion polling conducted in conflict environments and with populations that may have different cultural practices.

Reporting Results

In all reports, products and other assessments, it is important to not only give results of the surveys but to also describe the methodology used and the problems encountered. Too often methodologies provided are 'sanitised' containing only basic information about the administration process. The report has information on the dates the surveys were conducted, the sampling plan employed including margins of error and replacement of sampling points with explanations, number and sex of interviewers, respondent selection and substitution, and non-response and completion information. Usually only brief information is provided on problems encountered during the fieldwork. In Afghanistan the most commonly cited problems are the inability to locate a village, the village is under the control of insurgents, or there is kinetic activity occurring. More attention should be given to the dangers of the job of data collection and how this may cause deviations from the original methodology.

Conclusion

Conducting surveys is a complex activity with many potential ways for errors to occur and ways to misuse results. Instrument design, translation, administration, and respondent error are just a few of the areas the researcher must consider when designing the survey. The compounding of errors can draw uncertainty to the validity of the results.

When used for operations assessment purposes, polling results should always be reported within a social and temporal context. Recent salient events can impact not only what, but how a respondent chooses (or declines) to answer. Single items from a poll should not be used to populate extensive frameworks, dashboards and spreadsheets of MOEs.

Given all the potential sources for error that can occur when conducting polling, extreme caution is advised when using results to support operational planning and decisions affecting campaigns. Careful attention should be paid to reports on the methods used to conduct the poll to ensure that errors were minimised and to ensure that results are properly used. Finally, caution should be given when comparing the results of multiple polls over time.

In summary, opinion polling can provide the human perspective within complex operational environments, one that has been lacking in traditional assessments. When used in combination with other data sources that corroborate the findings, the evidence provided by polling can give a commander better situational awareness and contribute to overall mission assessment.

References

ABC/BBC/ARD. 2010. *Afghanistan: Where Things Stand*. <http://abcnews.go.com/images/PollingUnit/1099a1Afghanistan-WhereThingsStand.pdf>

Alwin, D. 2007. *Margins of Error: A Study of Reliability in Survey Measurement.* Hoboken, NJ: Wiley & Sons.

Beatty, P. 1995. Understanding the standardized/non-standardized interviewing controversy. *Journal of Official Statistics* 11:147-160.

Berg, N. 2005. Non-response bias. In Kempf-Leonard, K. ed. *Encyclopedia of Social Measurement* 2:865-873. London: Academic Press.

Berg, S. 1988. *Snowball sampling. In Kotz, S.,* and N. L. Johnson*, eds. Encyclopaedia of Statistical Sciences* 8:528-532. New York: Wiley.

Bradburn N. H., and W. M. Mason. 1964. The effects of question order on responses. *Journal of Marketing Research* 1:57-61.

Brennan, M., J. Chan, D. Hini, and D. Esslemont. 1996. Improving the accuracy of recall data: A test of two procedures. *Marketing Bulletin* 7:20-29.

Campbell, D. T. 1950. The indirect assessment of social attitudes. Psychological Bulletin 47:15-38.

Converse, J., and S. Presser. 1986. Survey Questions: *Handcrafting the Standardized Questionnaire*. Beverly Hills, CA: Sage.

Fisher, R. J. 1993. *Social desirability bias and the validity of indirect questioning.* Journal of Consumer Research 20:280-328.

Forsyth, B. H., M. S. Kudela, D. Lawrence, K. Levin, and G. B. Willis. 2006. *Methods for Translating Survey Questionnaires*. Paper presented to American Association for Public Opinion Research, Montreal, Canada.

Fowler, Floyd Jr., and Floyd Fowler. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage.

Goffman, E. 1959. *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday Anchor Books.

Gordon, R. A. 1987. Social desirability bias: A demonstration and technique for its reduction. *Teaching of Psychology* 14:40-42.

Guerin, B. 1994. Analysing Social Behaviour: *Behaviour Analysis and the Social Sciences*. Reno, NV: Context Press.

Headquarters, Department of the Army. 2006. Field Manual 3-24, *Counterinsurgency*. MCWP-3-33.5. <http://www.scribd.com/doc/9137276/US-Army-Field-Manual-FM-324-Counterinsurgency>

International Republican Institute. 2009. *Afghanistan Public Opinion Survey: July 16-26, 2009*. <http://www.iri.org/sites/default/files/2009-News-Release-Files/2009%20August%2014%20Survey%20of%20Afghan%20Public%20Opinion%20July%2016-26,%202009.pdf>

Kilcullen, D. J. 2006. *Three Pillars of Counterinsurgency.* Remarks delivered at the U.S. Government Counterinsurgency Conference, Washington D.C., 28 September 2006. Available online <http://www.au.af.mil/au/awc/awcgate/uscoin/3pillars_of_counterinsurgency.pdf>

Kish, L. 1962. Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association* 57:92-115.

Kitzinger, J. 1994. The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health and Illness* 16:103-121.

Lassiter, G. D., J. I. Stone, and M. F. Weigold. 1987. Effect of leading questions on the self-monitoring-memory correlation. *Personality & Social Psychology Bulletin* 13:537-545.

Law and Order Trust Fund of Afghanistan. 2011. *Police Perception Survey 2011: The Afghan Perspective*. <http://www.undp.org.af/Publications/KeyDocuments/2011/PPS-Eng%20Version-2011%20Final%20Lowest%20Res.pdf>

Lohr, S. 2010. *Sampling Design and Analysis*, 2nd Edition. Boston, MA: Brooks/Cole.

Lynn, P., ed. 2009. *Methodology of Longitudinal Surveys.* Hoboken, NJ: John Wiley and Sons.

McClendon, M. J. 1991. Acquiescence and recency response-order effects in interview surveys. *Sociological Methods & Research* 20:60-103.

McKay R. B., M. J. Breslow, R. L. Sangster, S. M. Gabbard, R. W. Reynolds, J. M. Nakamoto, and J. Tarnai. 1996. Translating survey questionnaires: Lessons learned. *New Directions for evaluation* 70:93-104.

Melia, T. O. 2002. *Afghan Perspectives on Democracy: A Report on Focus Groups in the Kabul Area on the Eve of the Emergency Loya Jirga*. Prepared for the

National Democratic Institute for International Affairs. <http://www.accessdemocracy.org/files/1411_af_report_052802.pdf>

Messick, S., and D. N. Jackson. 1961. Acquiescence and the factorial interpretation of the MMPI. *Psychological Bulletin* 58:299-304.

Ministry of Counternarcotics of the Government of Afghanistan. 2012. Afghanistan Opium Survey 2012: *Opium Risk Assessment for all Regions*. <http://www.unodc.org/documents/crop-monitoring/Afghanistan/ORAS_report_2012.pdf>

Nederhof, A. J. 1985. Methods of coping with social desirability bias: A review. *European Journal of Social Psychology* 15:263-280.

Nisbett, R. E., and T. D. Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. Psychological Review 84:231-259.

North Atlantic Treaty Organisation (NATO). 2011a. *NATO Operations Assessment Handbook*, Interim version 1.0. Norfolk, VA: Supreme Allied Commander Transformation.

——— 2011b. SAS-091, NATO Operations Assessment: A Case Study Based on Planning for Transition in Afghanistan.

——— 2012. HFM 185. Processes for Assessing Outcomes of Multinational Missions.

Oxfam. 2009. *The Cost of War: Afghan Experiences of Conflict, 1978-2009.* <http://news.bbc.co.uk/2/shared/bsp/hi/pdfs/18_11_09_oxfam_afghan.pdf>

Sherwood, G. G. 1981. Self-serving biases in person perception: A re-examination of projection as a mechanism of defence. *Psychological Bulletin* 90:445-459.

Shwarz, N. 1999. Self-reports: How the questions shape the answers. *American Psychologist* 54:93-105.

Stubbs, R. 2008. From search and destroy to hearts and minds: The evolution of British strategy in Malaya (1948-60). In Martson, D. and Malkasian, C., eds. *Counterinsurgency in Modern Warfare.* Osprey Publishing.

Sudman, S, and N. M. Bradburn.1973. Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association* 68:805-815.

———1974. Response effects in surveys. Chicago: Aldine.

Sudman, S., A. Finn, and L. Lannon. 1984. The use of bounded recall procedures in single interviews. *Public Opinion Quarterly* 48:520-524.

The Asia Foundation. 2012. *Afghanistan in 2012: A Survey of the Afghan People.* <http://asiafoundation.org/publications/pdf/1163>

Thomas, W. I., and F. Znaniecki. 1918. *The Polish Peasant in Europe and America (Vol. 1)*. Boston: Badger.

Thurstone, L. L., and E. J. Chave. 1929. The measurement of social attitudes. *Journal of Abnormal and Social psychology* 26:249-269.

United States Centre for Army Lessons Learned. 2010. Handbook 10-41, Chapter 2. Assessment Support to Planning. <http://usacac.army.mil/cac2/call/docs/10-41/ch_2.asp>

United States Institute of Peace; United States Army Corps of Engineers; and United States Army Peace Keeping and Stability Operations Institute. 2008. *Measuring Progress in Conflict Environments (MPICE): A Metrics Framework for Assessing Conflict Transformation and Stabilisation.* August 2008, Version 1.0.

5

Time Series Analysis in a Theatre Strategic Headquarter

Marcus Gaul, Eric Jesse

Abstract

Time series analysis in its many facets and shapes, from data exploration to spatial analysis, and from qualitative to quantitative considerations offers a set of highly valuable tools for analysts in support of senior leader's decision making. Time series analysis can not only generate crucial insights by combining time and spatial domains, but it can help to identify unexpected relationships and opportunities that can be exploited by a command, and enable the assessment of operational effectiveness and developments in insurgent tactics. In this article we present the fundamentals of time series analysis and discuss some of its applications and uses stemming from our experiences of working in the Afghan Assessment Group (AAG) at the International Security Assistance Force's (ISAF) Headquarters in Kabul, Afghanistan.

Introduction

The Afghan Assessment Group (AAG) is the International Security Assistance Force's (ISAF) Headquarter's principal analysis and assessment organisation. The AAG provides regular analysis and assessments of the campaign to the ISAF Commander, the Government of the Islamic Republic of Afghanistan (GIRoA), and the governments of troop contributing nations, the press, and the public. In assessing the campaign, AAG uses a variety of analytic methodologies, one of which is time series analysis, the focus of this discussion. Time series analysis, in particular, leverages the information advantage ISAF has achieved through its advanced theatre-wide data collection processes.

Our chapter begins with a general review of time series analysis and associated terminology. Then we will cover selected common analytical issues and challenges, to include data collection and consistency, data aggregation, variable

selection, seasonality, and non-constant variance. Aware of these problems which frame analytic approaches, we next discuss methods AAG employs to analyse insurgent activity in Afghanistan, covering year-over-year differencing, seasonal decomposition, and regression analysis. This carries into a discussion on the identification of statistically significant changes in insurgent behavior. Finally, we discuss the incorporation of geospatial considerations with temporal data. By adding spatial data we can visualize, assess and explain trends, which might otherwise go unobserved by analysts and decision makers. We conclude with a discussion of the challenges quantitative analysis encounters during the ISAF campaign, and the facilitating measures taken to remove them.

Time Series Analysis

A time series is a collection of observations obtained through repeated measurements over time. Time series are most often considered subsets of stochastic models or stochastic processes. Stochastic processes concern sequences of events governed by probabilistic laws. Stochastic models infer characteristics of a population using temporal or spatial data. Time series analysis utilises the natural temporal order of the data to extract information about the underlying stochastic process.

The extent of assumptions about the underlying data in a statistical analysis defines two classes of statistics, parametric and non-parametric. The most common models utilise parametric statistics. Parametric statistics assume that the data has some underlying structure, most typically about the underlying distribution (often the normal distribution) and variance (often constant over time). Some widely known parametric models are ordinary least squares regression, generalized least squares, autoregressive and moving average models. Non-parametric (or distribution free) models do not assume an underlying distribution for the data; instead, they try to determine the underlying system structure within the data. As a result, larger datasets are required for non-parametric models to make similar conclusions as analogous parametric test. Non-parametric methods often rely on ranking or ordering of data for analysis. The Kolmogorov-Smirnov test is one such well-known non-parametric method used to test whether a data set follows an assumed probability distribution, or whether two data sets come from the same generating process.

Our discussion of time series analysis focuses on parametric statistics, but we need to be aware that non-parametric analysis techniques exist when parametric assumptions are shown to be invalid.

Statistical analysis is further subdivided into linear and *non-linear* analysis. Linear analysis assumes the dependent variable (response variables) can be explained by a linear combination of independent variables (explanatory variables). Linear analysis assumes an additive relationship between variables, which enables a closed form solution to estimation. While linear models are intuitively easier to interpret and understand, most real world systems do not follow strictly additive processes. In non-linear analysis, the dependent variable is explained by a non-linear combination of independent variables. Examples of non-linear functions include Lorenz curves, exponential functions, logarithmic functions, and Gaussian functions. Exponential and logarithmic functions can be transformed into linear forms enabling a closed form solution, but in general, non-linear functions use numerical techniques to estimate parameters.

Time series analysis can be used to answer two types of inquiries: descriptive and predictive. Descriptive inquires seek to ascertain the existence of hypothesized relationships between the dependent and independent variables. For example, how do enemy-initiated attacks, the opportunities for leave, seasonality, or major Islamic holidays affect attrition, i.e. the (planned and unplanned) loss of personnel in the Afghan National Army. Descriptive analyses look for significance and direction of relationships between the dependent and independent variables. Predictive analysis projects potential future system states based on the historical relationship between a dependent variable and a set of independent variables. An example is predicting the level of enemy activity in 2013 using the historical relationship between the enemy activity level and independent variables such as the number of coalition forces.[1]

Data Collection and Aggregation

ISAF collects vast amounts of data with most of the data being provided by military units using standard reporting procedures. These units are not focused primarily on data collection, so data quality, completeness, and accuracy can vary. Reported and collected data also tends to evolve with mission objectives,

1. See Gons et al. 2012 for an example of such a scenario.

potentially adding further complexity to analysis. The collected data are refined with subsequent reports categorized and transmitted to an analytical database that also serves as a historical archive. A centralized data collation function provides an additional layer of quality assurance by enhancing data consistency, eliminating duplicate reports, and identifying and reconciling database anomalies.

ISAF tracks additional data sources beyond military operational reporting. These reports include open source, United Nations, and NGO reporting. Military operational reporting is usually entered into the database within 24 hours of occurrence and data entry. This lag can mislead or bias near term analysis. Understanding the data is the foremost task of the analysts, since quantity, quality, and reliability of the data is a primary factor shaping the analysis.

Data aggregation is the next major consideration for the analyst. Data should not be aggregated beyond the relevant level of inquiry. For example, looking at theatre-wide improvised explosive device (IED) events would be an inappropriate means of assessing a new mine rolling system in Kandahar Province.

The aggregation of security-related data in Afghanistan is at least a three-dimensional issue, with data having spatial, temporal, and categorical aspects. Spatial disaggregation presents a particular challenge due to the number of ways Afghanistan can be divided. ISAF seeks to map data with an accuracy of a few meters to a specific military grid reference system (MGRS) coordinate, but this level of accuracy is not always achieved, especially for reports from the Afghan National Security Force (ANSF). Almost all data is identified at the district level. The approximately 400 districts are in 34 provinces, which in turn are part of six geographically distinct regions in Afghanistan: East, South, Southwest, West, North, and the Capital—the ISAF Regional Commands (RCs). Numerous further groupings are possible, for example defined by climate, terrain, and population demographics.

Aggregation can influence analytic results. Simpson's paradox is the phenomenon where a trend that appears in data subsets disappears or reverses when these subsets are combined. As a hypothetical example, we might find an improving security situation in 28 of Afghanistan's 34 provinces when analysed separately, but when looking across the entire theatre, the trend could be reversed and security would appear to be worsening. This trend could be

driven by dominating negative conditions in just 2 provinces, for example Helmand and Kandahar.

Disaggregation can potentially obscure important trends due to the effect of specific events, operations, or simple random noise. Disaggregation can also limit the power of statistical analysis by reducing the number of observations. There is no "correct" level of aggregation to employ across all analysis. Aggregate models are very important to understanding general theatre trends, while disaggregated models enable the understandings that are important at operational and tactical levels. Disaggregate models help with the analysis of local effects of actions such as clear/hold operations, cache finds, or weapon seizures. The appropriate level of aggregation is a judgment call by the operational analysts based on their understanding of the question and the tractability of the data.

Adjustments for Seasonality

Insurgent activity in Afghanistan follows a strong cyclical process. This type of recurring pattern in data is known as 'seasonality.'[2] Nationwide insurgent activity in Afghanistan is closely tied to the rhythm of daily life, with the progression of the seasons strongly influencing engagement levels. Planting and harvesting crops demand the attention of individuals who might otherwise be fighting coalition or Government of the Islamic Republic of Afghanistan (GIRoA) forces, while the winter cold and snows close mountain passes limits movement that results in diminishing contact between insurgents and coalition forces.

As measures for insurgent activity, we consider, for example, *security incidents*[3] or *enemy-initiated attacks*[4] (EIAs). EIAs provide a perspective on the active, executed components of insurgent engagements against ISAF and the ANSF, while security incidents also include elements that failed to execute such

2. Many time series show cyclic variations that are known as seasonal fluctuations or seasonality. Insurgent activity in large parts of Afghanistan is correlated to the seasonal temperature fluctuations over the course of a year.

3. Security incidents comprise enemy-initiated attacks (enemy-initiated direct fire, indirect fire, surface-to-air fire and executed improvised explosive device (IED) attacks), as well as potential IED attacks such as IEDs and mines that were found and cleared.

4. Enemy-initiated attacks is a term that is used for the total of enemy-initiated direct fire, indirect fire, and surface-to-air fire attacks, as well as executed improvised explosive device (IED) attacks, namely IED explosions and mine strikes.

as explosive devices that were found and cleared. For the purpose of this article, we consider enemy-initiated attacks as an approximation for insurgent activity.

Figure 5.1 shows the periodic pattern of insurgent activity over the period from 1 January 2007 through 31 December 2012. In July of 2011, EIAs averaged over 700 per week. Five months later, in December 2011, EIAs averaged just over 350 per week, a drop of approximately 50 percent. Interpreting this decrease as a sign of an improving security situation could be misleading for several reasons. First of all, this comparison fails to account for seasonality, i.e. fluctuations in the insurgent activity due to the seasonal temperature changes over the course of a year. We can see this by considering EIAs several months later. In July of 2012, EIAs again averaged over 700 per week.

An assessment of the security situation should include not only insurgent

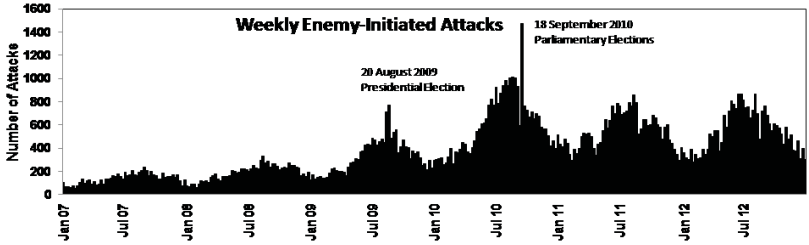


Figure 5.1: Weekly enemy-initiated attacks in Afghanistan
(1 January 2007- 31 December 2012)

activity, but also quantitative and qualitative metrics of population perceptions of development and governance, as well as interdependencies among the variables. A comprehensive analysis of this type extends beyond the scope of this chapter.

When analyzing the insurgency, it is useful to differentiate between *systemic changes* and those changes that are attributable to the cyclical pattern of the conflict. ISAF assessments regularly utilise three approaches to adjust for seasonality year-over-year differences, classical seasonal decomposition, and regression analysis.

Year-over-year (YoY) differences

YoY differences allow comparisons between seasonally analogous periods, e.g. comparing EIAs in January 2013 with January 2012. The comparison window is important, as periods that are too small risk confusing natural variance in the data with significant trends. HQ ISAF AAG found that a one week period was the shortest possible period to identify meaningful changes in insurgent activity from data noise. On the other hand, if YoY analysis spans too much time, operationally relevant factors may be obscured, giving the commander misleading information.

When seasonal trends evolve slowly, it is sometimes possible to compare the last period with that immediately preceding to search for changes, e.g. for data with an annual seasonal trend we can compare the last week with the preceding week. This is an analogous approach to more traditional YoY differencing in that the two periods should have very similar seasonal influences. When they differ considerably, it could be indications of an emerging trend.

Classical seasonal decomposition

Classical seasonal decomposition is a second approach commonly employed to seasonally adjust the data. This method removes seasonal fluctuations from the data allowing the identification of long-term trends. The technique factors time series data (X_t) into two parts: a seasonal component (S_t) and a trend component (T_t). While both additive and multiplicative decompositions are available, multiplicative decompositions perform best when modeling insurgent activity in Afghanistan ($X_t = T_t * S_t$). Since we know *a priori* that the cycle affecting the time series of enemy-initiated attack data (X_t) is annual, we can calculate the two components of the seasonal decomposition using the following steps:

1. Compute a moving average for the series X_t, with the moving average window being of equal length to 1 season (i.e. 12 months, 52 weeks, 365 days). This step removes intra-seasonal variability from the data resulting in a smoothed time series (SM_t).
2. Compute the ratio of the observed (X_t) and smoothed (SM_t) series. This ratio is the raw seasonal component.

3. Next the average seasonal component (S_t) is computed by averaging the raw seasonal components for each point in the time series (i.e. each month, each week, and each day). Various forms of averaging the raw seasonal components to determine the average seasonal component are possible. We recommend normalized medial averaging. This approach excludes outliers from the average, preventing particularly abnormal observations from influencing the seasonal adjustment in the same period for other cycles.
4. The original time series (X_t) can now be seasonally adjusted by dividing it by the average seasonal component (S_t), producing the trend component (T_t).

A seasonal decomposition of weekly EIAs in Afghanistan is depicted in Figure 5.2 showing the usefulness of this approach.

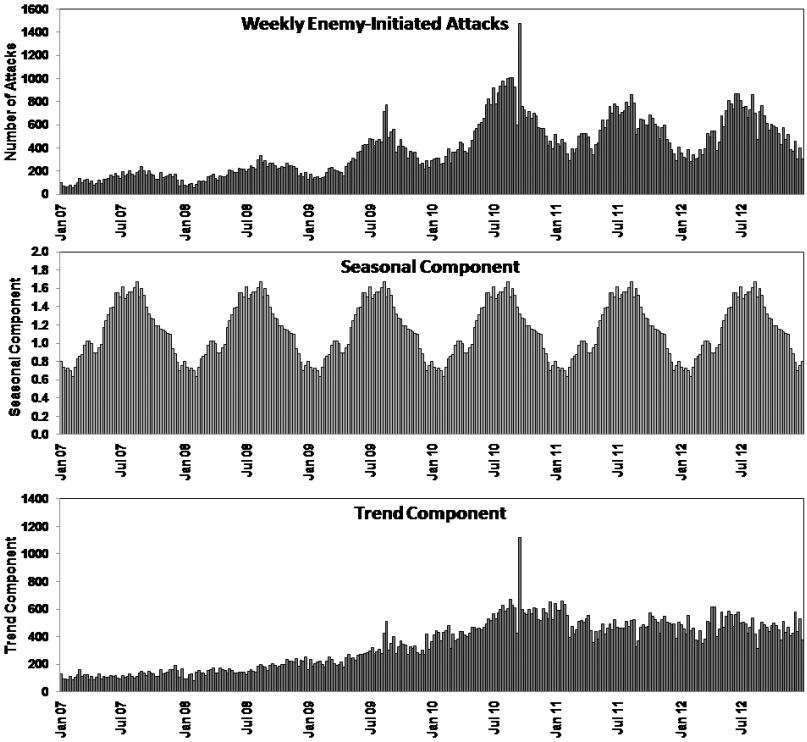


Figure 5.2: Seasonal decomposition of weekly enemy-initiated attacks in Afghanistan (1 January 2007- 31 December 2012)

Regression analysis

Regression analysis is the third approach to seasonally adjust data and identify trends. Regression analysis is used to estimate relationships between a dependent variable and one or more independent variables. A major strength of regression analysis is that it assesses the relationships between multiple independent variables to the dependent variable. Variable selection is the process of identifying control variables in regression analysis. In Afghanistan, factors known to influence insurgent activity include coalition and ANSF force levels and significant events such as elections and Islamic holy days. The number of insurgent attacks on ISAF and ANSF patrols is related to the location and frequency of these patrols. In the case of elections and Islamic holy days, insurgent activity has two opposing characteristics. Prior to and during elections, insurgents attempting to intimidate the population accelerate attacks. Conversely, some Islamic holy days see lower levels of activity due to religious customs.

When selecting control variables for factors that influence insurgent activity, it is important to distinguish between regular periodicity and irregular periodicity, i.e. periods or events that recur at the same time each year and those periods or events that occur on different calendar dates. Events that occur at the same time annually are subsumed in the seasonal adjustment. For example, insurgent activity decreases during the poppy harvest season each year, but since the harvest occurs at approximately the same time each year it is captured by the seasonal adjustment. Islamic holy days, however, follow a lunar calendar in which a year is always 12 lunar months, making the months not linked to the seasons. Key dates shift each year 10 to 12 days. As a result, the effects of the annual Islamic observances of Eid al-Fitr (the "Festival of Fast-Breaking" marking the end of Ramadan) and Eid al-Adha (the "Festival of Sacrifice") will not be accounted for in a seasonal adjustment. Regression models, being the most powerful of the seasonal adjustment techniques will be explored in greater detail below.

Linear Regression Models

There are several types of regression models. In this section we discuss how simple linear regression models can be applied to estimating trends in

enemy-initiated attacks. Other approaches will be addressed in later sections. Determining the best model structure, as measured by statistically accurate parameter estimates, can be challenging. Ordinary least squares (OLS) regression assumes the independent variables are additive (linear).

In Afghanistan, as discussed earlier, some of the major factors related to insurgent activity, in particular seasonality, have a multiplicative relationship. We use a log transformation converting an approximately multiplicative process into a linear model.[5] Techniques are available for treating seasonal adjustments simultaneously with other explanatory variables. If treated separately, seasonal adjustments will include some of the effects of the explanatory variables. Treating the adjustment simultaneously with other explanatory factors ensures that the seasonal adjustment itself does not include effects that are more correctly attributed to the other explanatory factors in the model. In a regression model, the addition of "indicator" variables for each observation period in the seasonal cycle (which is one year for the Afghan insurgency) is equivalent to a classical seasonal decomposition. An indicator variable is assigned a value of one when a condition is present and zero otherwise. For example, the seasonal decomposition in a regression model using weekly data includes 51 additional indicator variables, one for each week. The 52nd week is captured in the intercept, as it is perfectly identified by the 51 other weekly indicator variables and therefore cannot be directly estimated.

A simple (multiple) linear regression model of weekly insurgent activity expressed by seasonality, coalition and Afghan Nation Army (ANA) force levels, and the holy days of Eid al-Fitr, and Eid al-Adha, can be written as follows (with each variable considered a vector of the data indexed by week):

5. Coalition force levels strongly influence insurgent activity levels, but coalition force levels have varied significantly from less than 40,000 troops in 2006 to over 150,000 at the peak of the surge in 2011. As a result, the variance in insurgent activity has changed significantly over time. Uneven variance over time, heteroskedasticity, causes ordinary least squares estimation to be inefficient and overstates statistical confidence in measured estimates. A log transformation of the dependent variable makes a multiplicative model additive and in many cases significantly reduces or entirely removes the problem of heteroskedasticity. There are numerous methods to stabilise variance. A square root transformation is often also appropriate for count processes. Other variance stabilisations include the Anscombe transform (see Anscombe 1948), Freeman-Tukey transform (see Freeman and Tukey 1950), and Box-Cox transform (see Box and Cox 1964) amongst others.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 \ldots + \beta_{57} x_{57} + \varepsilon$$

where

Y = log (insurgent activity),

$x1$ = log (coalition forces),

$x2$ = log (ANA forces),

$x3$ = Φ(Eid al-Fitr), where Φ is a scalar function assigning the value of one to a week in which the days of Eid al-Fitr fall, and the value of zero otherwise

$x4$ = Φ (Eid al-"Adha), where is a scalar function assigning the value of one to a week in which the days of Eid al-Adha fall, and the value of zero otherwise,

$x5$ = Φ(elections), where is a scalar function assigning the value of one to a week in which an election falls, and the value zero otherwise,

$x5$ = Φ(week) is the indicator (or dummy) variables for each observation period, i.e. week, in the year, and

ε is the residual error.

The parameters β are computed using OLS, i.e. by minimizing the sum of squared deviations between the dependent variable (observations) and the estimate using the independent variables. Each parameter measures the observed historical relationship the independent variable has with the dependent variable.

The observed and predicted results of this model for the Afghan theatre of operation are shown in Figure 5.3. The results of the OLS for nationwide and four Regional Command (RC) models are presented in Table 5.1.

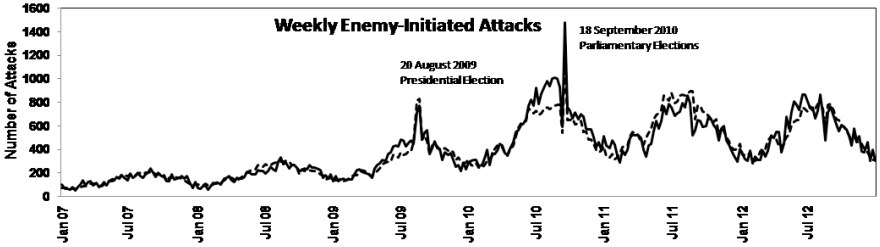


Figure 5.3: Observed versus predicted enemy-initiated attacks (EIAs)

Coefficient	Theatre	RC South	RC Southwest	RC East	RC North
Intercept	-9.76 (0.26) ***	-9.31 (0.46) ***	-15.59 (0.54) ***	-8.75 (0.29) ***	-15.59 (0.54) ***
log (Coalition)	1.11 (0.07) ***	0.76 (0.13) ***	1.98 (0.15) ***	0.48 (0.08) ***	1.98 (0.15) ***
log(ANA)	0.21 (0.06) ***	0.38 (0.11) ***	-0.22 (0.12) *	0.63 (0.07) ***	-0.22 (0.12) *
Eid al-Fitr	-0.26 (0.07) ***	-0.30 (0.12) **	-0.33 (0.14) **	-0.21 (0.08) ***	-0.33 (0.14) **
Eid al-Adha	-0.10 (0.07)	-0.18 (0.12)	-0.04 (0.14)	-0.04 (0.08)	-0.04 (0.14)
Election Week	0.54 (0.11) ***	0.50 (0.19) ***	0.36 (0.22)	0.63 (0.12) ***	0.36 (0.22)
R^2	0.9525	0.8596	0.8736	0.9423	0.8736

Table Format: Coefficient (Standard Error) Significance
Significance: *** at 0.01, ** at 0.05, * at 0.1

Table 5.1: Results of the OLS model by region in Afghanistan

We see in Table 5.1 that the parameters point in the hypothesized directions. Insurgent activity rises with increases in coalition and ANA force levels, Islamic holidays lead to decreases in insurgent activity, and elections result in increased insurgent activity. The only variable that does not point in its hypothesized direction is ANA troop level in RC North. This may have occurred because the ANA troop levels are modeled at the theatre level, and this may be an inadequate measure of ANA forces in RC North. It could also have occurred by chance with the measured effect being only weakly significant at a confidence level of 0.1, i.e. with a 90% chance of being a significant relationship.

Looking at the nationwide model, assuming all else to be equal, the interpretation of the model results is as follows:

- A 10% decrease in the number of coalition forces is correlated with a 11% decrease in insurgent activity
- A 10% increase in ANA forces is correlated with a 2% increase in insurgent activity
- A week encompassing Eid al-Fitr has a 26% decrease in insurgent activity
- An election weeks results in a 54% increase in insurgent activity

Finally, it should be noted that each of the 5 regression results in Table 5.1 have an associated R^2. This is sometimes called the coefficient of determinations

and it is a measure of the proportionate reduction of total variation associated with the use of the independent variables. It can take on values between zero and one, with higher numbers like those seen in the table implying that the regression model is a good fit to the data.

The levels of significance displayed in Table 5.1 are calculated assuming that the residuals are normally distributed, with constant variance, and are not correlated with each other. In general, it is necessary to test these implicit assumptions of parametric models and verify *model validity*. Violation of these assumptions can bias results and mislead conclusions.

Autocorrelation is one common issue in time series data. It occurs when the current observation is correlated with an observation at another time period, e.g. the immediately preceding one, violating the independence assumption of OLS.

Non-constant variance is a second potential issue. As coalition force levels have varied significantly, from less than 40,000 troops in 2006 to over 150,000 at the peak of the surge in 2011, the variance in reported insurgent activity has changed significantly over time. Uneven variance over time, heteroskedasticity, may result in overstated statistical confidence in parameter estimates.

We can test the OLS assumptions visually and with more formal tests. Figure 5.4 shows selected residual analysis for the nationwide model in Table 5.1. The top plot shows studentized residuals. Studentized residuals are normalized by their variance and thus should align with a standard normal curve with most observations falling between -2 and +2. The bottom plot compares a histogram of the residuals to a normal distribution. We can see that the residuals do not appear to violate the normality assumption.

The residuals show signs of autocorrelation, with insurgent activity in one period appearing strongly related to insurgent activity in adjacent periods, as visible in Figure 5.4. Having homoskedastic residuals (residuals with constant variance over time) that are reasonably balanced (centered about zero over time), our estimates are unbiased but our confidence in these estimates is overstated. In order to correct the effects of autocorrelation (or heteroskedasticity if balanced) in the error terms from OLS regression, the autocorrelation-consistent Newey-West variance-covariance matrix estimator of the residuals is often

used. Alternatively, we can utilise a generalized linear model and explicitly address the autocorrelation.

Autoregressive, Integrated Moving Average (ARIMA) Models

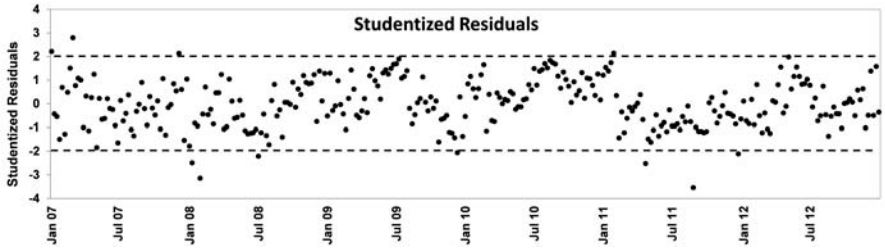


Figure 5.4: Selected residual analyses

Ordinary least squares is one of many forms of regression analysis (Hamilton 1995). As the event data is actually count data, Poisson or negative binomial models under certain conditions can yield much more accurate predictions. If data fidelity improved and limitations in computing power were removed, point process models like a Cox process result in even higher fidelity characterisations of the data over space and time (Cox 1955; Cox and Isham 1980; Cook and Lawless 2007; Allen, Borgan, and Gjessing 2008).

A class of models, which perform well in forecasting non-stationary time series data, like insurgent activity in Afghanistan, are autoregressive, integrated moving average (ARIMA) models (Box, Jenkins, and Reinsel 2008).[6] ARIMA models consist of three parts, an autoregressive (AR) part, an integrated (I), and a moving average part (MA). The model is usually referred to as an ARIMA (*p, d, q*) model where the parameters *p, d*, and q are non-negative integers. The parameter *p* refers to the order of the autoregressive part, i.e. the number of autoregressive terms, *d* refers to the order of the involved differencing (note that a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series), and *q* is the order of the moving average part of the model. An ARIMA model specification can be a

6. A stationary process is a stochastic process whose joint probability distribution does not change over time. This means parameters like the mean and variances are constant over time, i.e. time invariant. The data here exhibits cyclostationary properties–it varies annually. In order to perform time series analysis we transform the data to become stationary (Priestley 1988).

subset of the autoregressive, integrated, and moving average components. None of the parts of the ARIMA is required be included in its final specification.

An *autoregressive* model seeks to define the current observation as a function of preceding observations. For example, an autoregressive model of order one is given by $y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$, while the general autoregressive model of order p is given by

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-1} + \varepsilon_t$$

.

The *integrated* part of the model utilises successive first differences. For example, for a difference of order 2, the relationship analysed is with the variable defined by the difference between the current and previous periods minus the difference between the previous period and the period preceding, mathematically expressed as $z_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$ The most common use of such an integrated term is to account for a quadratic trend in the data.

A *moving average model* defines the current observation as the mean μ of the series, and current and previous error terms . For example, a moving average model of order one is given by $y_t = \mu + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1}$, and a moving average model of order *q* is given by $y_t = \quad + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$. Because the lagged error terms are not directly observable, iterative non-linear fitting procedures need to be used in place of linear least squares to estimate model parameters.

A seasonal ARIMA is an adaptation of the standard ARIMA that incorporates *a priori* information about the recurring seasonal pattern. In a seasonal ARIMA model with a recurrence period of time S, observation y_t is predicted by data values, differences, and errors with time lags that are multiples of S (the span of the complete seasonal cycle). For example, a seasonal autoregressive model of order two is given by $y_t = \beta_0 + \beta_1 y_{t-S} + \beta_1 y_{t-2S} + \varepsilon_t$.

An ARIMA model as briefly discussed above has several benefits. It allows us to model seasonality using fewer terms than in the (multiple) linear regression model described earlier, increasing estimation power. It also allows simultaneous estimation of non-seasonal factors like force levels, or Islamic holy days Eid al-Adha and Eid al-Fitr. An ARIMA model will also identify, capture, and reflect unspecified underlying processes that might be eliminated as noise in other types of estimation models, thus making it a very powerful tool for forecasting and prediction.

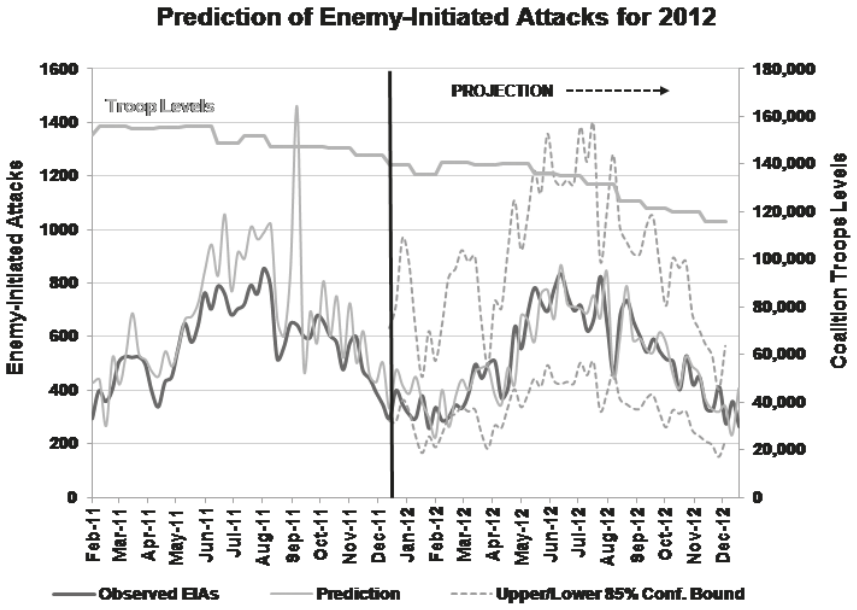


Figure 5.5: Seasonal ARIMA prediction model for insurgent activity

At the theatre strategic level we have used seasonal ARIMA models to forecast insurgent activity taking several factors into account, primarily the projected troop level changes. Figure 5.5 shows the model predictions of insurgent activity in 2012 based on historical data from 2007 to 2011. The troop level changes in 2012 that are factored in the prediction are shown, as well as the confidence bounds on the predicted EIA levels. The seasonal ARIMA has a non-seasonal part of (1, 0, 0) and a seasonal specification of (0, 2, 1); it also includes control variables for troop levels, Eid al-Fitr, and Eid al-Adha; however, no examination of the structural, systemic causes for insurgent activity, or the changes and dynamics are considered.

Note that the model's overestimation of insurgent activity for 2011, as seen in Figure 5.5, is caused by not controlling for voter registration periods or the 18 September 2010 Parliamentary Election.

Identifying Operationally Relevant Changes

The identification of operationally relevant changes is one goal of analysis at a strategic headquarters. The ability to rapidly evaluate and assess changes, identify opportunities first, and provide the option to react faster, can provide a significant advantage over the opposition. ISAF developed a methodology that utilises regression models to quickly identify measurable changes in insurgent activity and determine the statistical significance of this change (Jesse 2013).

The procedure begins by defining a model to explain insurgent activity, and estimating it against a benchmark period of data. We next compute the residuals for this benchmark period, which are defined as the difference between model predictions and observed insurgent activity. The residuals, or fitting errors, to regression models often follow a distribution, which is in many cases a normal distribution with mean zero. We can estimate the parameters of this residual distribution and design a statistical test to search for un-modeled changes in insurgent activity. With the model predictions over the benchmark period, we can forecast insurgent activity levels for new time periods and compare these with observed insurgent activity. From the sequence of errors obtained by this procedure, and the known information about the residuals' distribution, we can find the statistical likelihood for a significant change (Chandler and Scott 2011).

One use of the above approach at ISAF is to monitor daily insurgent activity. We use a two stage least-squares model[7] of daily enemy-initiated attacks baselined on the period from 1 January 2010 to 31 October 2012 to monitor changes in insurgent activity.

The advantage of a two-stage model is its separate accounting for additive and multiplicative relations. The first stage models enemy-initiated attacks using ISAF and ANSF troop levels, which explains much of the growth in insurgent activity. For the second stage, the observed enemy-initiated attack values are normalized by the predictions from the first stage. This procedure results in the dependent variable being normalized around one (seasonal fluctuations persist), and it should remove any heteroskedasticity. Fitting the

7. A two-stage model can allow implicit correction of heteroskedastic variance in a dependent variable. In the first modeling stage we control for additive independent variables and in the second stage multiplicative independent variables. This approach can outperform dependent variable transformations, as it incorporates a priori knowledge about modeled relationships. The approach yields unbiased results as long as the independent variables in the first stage are uncorrelated with the independent variables in the second stage.

second stage on the normalized dependent variable, independent variable coefficients estimate the relative change from the mean caused by the variables. In the example, fitting the second stage model, we get estimates of the effects of the Islamic holidays of Ramadan, Eid al-Adha, and Eid al-Fitr seasonality approximated by month.

In the next step, we compute the residuals of the two stage model. The residuals of this model should closely approximate a normal distribution with a mean of zero and constant standard error. We then employ a *t*-test to calculate the likelihood that the observed insurgent activity deviates significantly from activity explained by modeled factors. Or stated differently, it tests the null hypothesis that the mean of a sample population, the residuals, is equal to a specified value μ_0.

The *t*-statistic is found by taking the mean of the sample population $\overline{x}$, minus the test value , and dividing this by the sample standard deviation $s / \sqrt{n}$, where n is the sample size. For most regression models, to include those presented here, the test value μ_0 is zero, with the mean of the errors averaging to zero by construction. The *t*-statistic is:

$$t = \frac{\overline{x} - \mu_0}{s / \sqrt{n}}$$

The *t*-statistic comes from a Student's *t*-distribution, which maps directly to probability values allowing the determination of statistical significance. Rearranging the *t*-test equation, thresholds can be defined at the standard levels for statistical significance 0.10, 0.05, and 0.01, which corresponds to the identification of a change from the baseline with 90%, 95%, and 99% certainty.[8] The determination of the thresholds is accomplished by first finding the *t*-distribution value for the desired confidence level p, and degrees of freedom in the sample size $n-1$. The t value, scaled by the sample standard deviation $s / \sqrt{n}$ provides upper and lower bounds for the sample mean minus the test value:

$$t_{n-1,p}\left(\frac{s}{\sqrt{n}}\right) < \overline{x} - \mu_0 < t_{n-1,p}\left(\frac{s}{\sqrt{n}}\right)$$

8. Analyst selection of an appropriate significance level will be dictated by context. It must balance type 1 error, an incorrect rejection of the null hypothesis and type II error, falsely rejecting the null for the alternative hypothesis when in fact the null hypothesis is true (Pratt, Raiffa, and Schlaifer 2001).

The exact likelihood of having observed a change can be found by solving explicitly for *t*; all the required sample characteristics are known (sample mean $\overline{x}$, number of events in the sample *n*, test value μ_0, and sample variance $s / \sqrt{n}$):

$$\overline{x} = \mu_0 \pm t_{n-1,p}\left(\frac{s}{\sqrt{n}}\right)$$

$$t_{n-1,p} = \frac{\left(\overline{x} - \mu_0\right)}{\left(\frac{s}{\sqrt{n}}\right)}$$

Utilising the detailed approach, Figure 5.6 shows one possible visualization of significant unexpected deviations in insurgent activity that occurred in 2012. Significant increases are denoted by shades of red, with darker reds indicating increased statistical confidence. Significant decreases are denoted by shades of

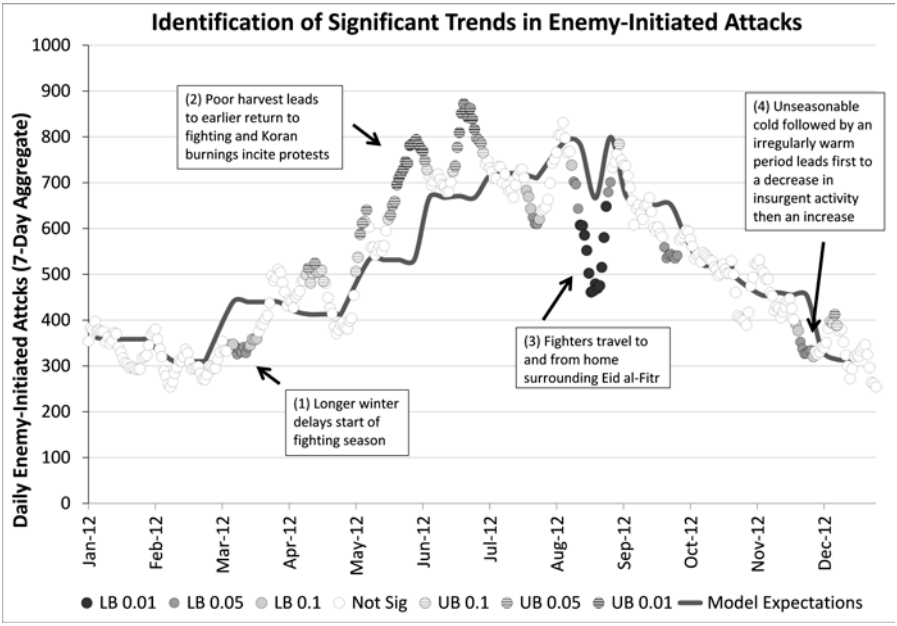


Figure 5.6: Results of a two-stage least squares model to identify significant unexpected trends in nationwide enemy-initiated attacks in Afghanistan in 2012

green, with darker shades indicating increased statistical confidence. UB/LB means upper and lower confidence bounds, gray circles represent non-significant incidents. While the underlying analysis leverages daily residuals, Figure 5.6 plots moving seven-day averages of enemy-initiated attacks, smoothing the trend line for visual analysis.

The strength of this process is that in recurring analysis it allows us to readily identify deviations from expected insurgent activity as they occur. With the identification of significant changes, possible explanations for these shifts are as follows:

1. An atypically severe winter might have caused a delay in the start of the fighting season.
2. A poorer harvest likely returned insurgent fighters to the fight earlier. At the same time, the burning of Korans at Bagram airbase reaches international and Afghan media outlets inciting protests.
3. EIAs decrease significantly surrounding Eid al-Fitr as insurgents travel to and from home for the observance.
4. Unseasonable cold temperatures likely led to lower than usual expected insurgent activities, but these were almost immediately followed by a period of unseasonable warm temperatures which probably led to an increase in activity levels.

Methods for Visualizing Time Series Data

Visualization is a very useful tool for showing changes in patterns or trends that might otherwise go unnoticed. Standard ways to display changes over time comprise among others line charts, bar charts, or scatter plots such as shown in Figure 5.6 with a supplementary colour dimension to convey additional information, in this case statistical significance. Time series data can be categorized by considering their context or frame of reference–usually distinguished into abstract and spatial. An abstract frame of reference simply refers to data that has been collected in a non-spatial context, i.e. without connection to a spatial layout. A spatial frame of reference, on the other hand, implies the existence of a spatial layout for the data set–it allows viewing space as some kind of indexing system for events (Anselin 1992; Cressie 1993; Cressie and Wilke 2011).

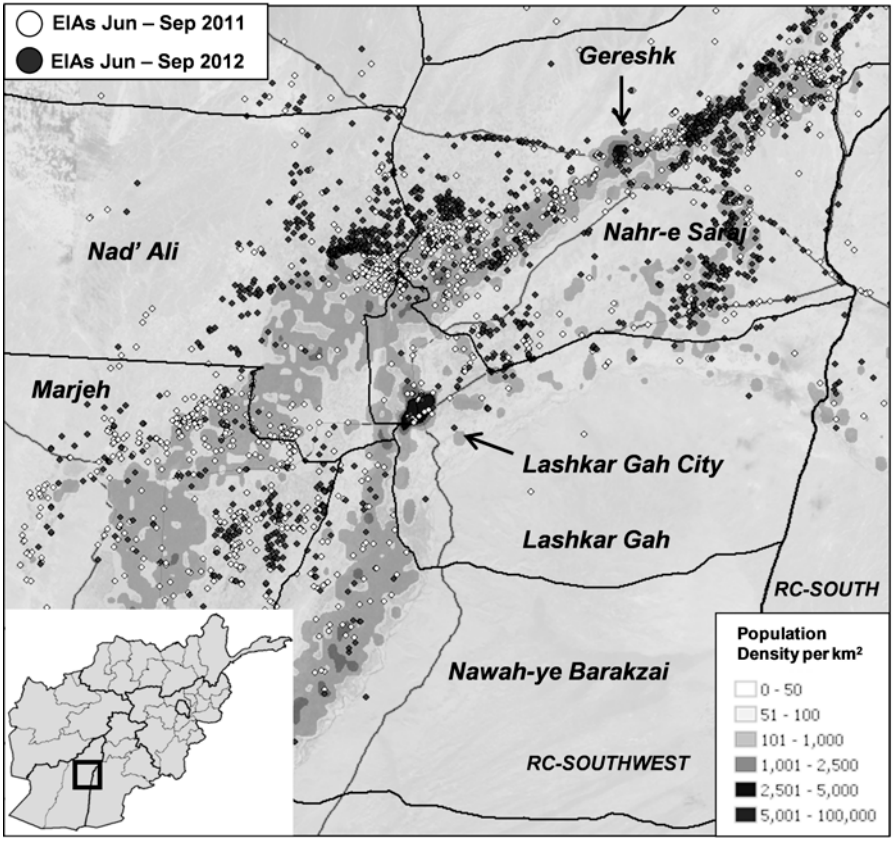


Figure 5.7a: A static, geospatial scatter plot visualizing individual enemy-initiated attacks in the period from June to September 2012 and the corresponding period in 2011

In the Afghanistan Theatre of Operations spatial information is collected– its geographic location for each event. Although it is important to know the insurgent activity changes over time, it is equally important to understand where insurgent actions occur, their spatial structures, spatial interactions, and the population that is impacted. The power for spatial data analysis lies in its geo-relational database structure, i.e. in the combination of value information and locational information. A geographical information system (GIS) is the natural tool to aid in the display and analysis of spatial data. The advanced display capabilities contained in a GIS can be useful for the visualization of

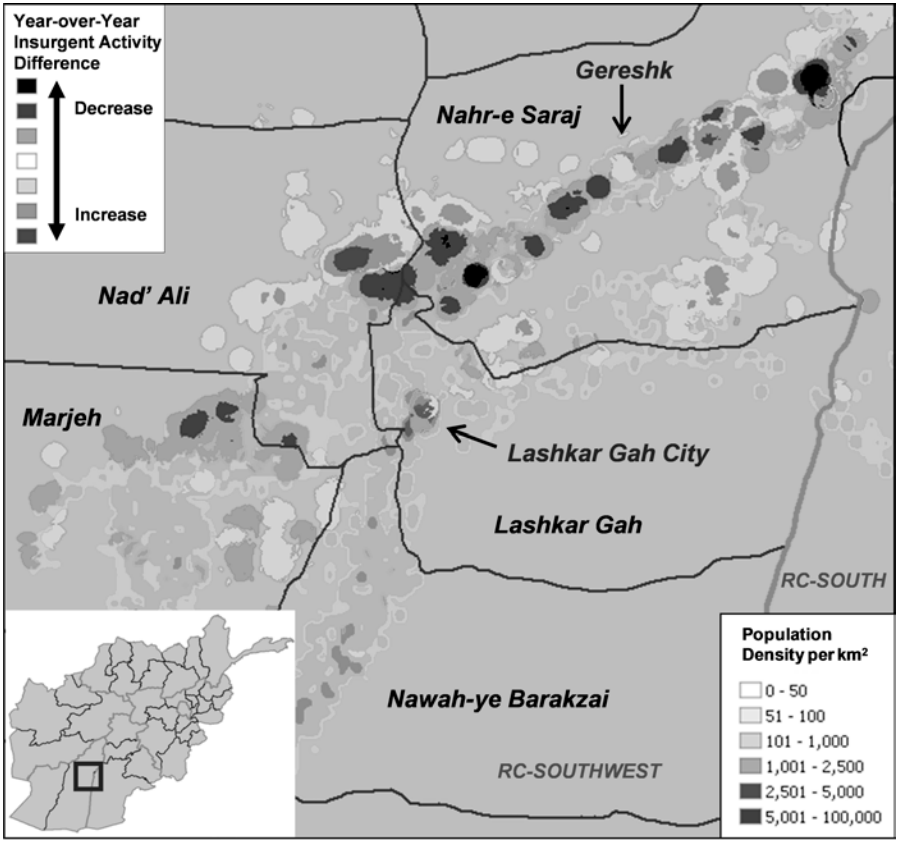


Figure 5.7b: A static, geospatial density plot visualizing changes in enemy-initiated attacks for the period from June to September 2012 compared to the corresponding period in 2011

processed, geo-referenced data, or the results of statistical analyses. These spatial visualizations can have a temporal dimension. In Figures 5.7a and 5.7b we show examples for two static geospatial representations.[9]

Figure 5.7a displays insurgent activity in Afghanistan's central Helmand Province relative to the population that is visualized as a background layer. The location of each individual insurgent attack is represented by a coloured

9. Static representations visualize the data in still images (i.e. the representations do no changes automatically over time), while dynamic representations convey the time dependency of the data (i.e. representations over sequential time periods).

point, where the colour represents one of the two time periods under consideration–June to September 2012 and, for comparability, the same period in 2011. This representation shows that insurgent activity moved from higher population densities in the central part (during the 2011 time period), to areas with lower population densities in 2012. Combined team operations and a high operational tempo appear to have had a considerable impact on the insurgent's campaign by reducing attacks in key population centers in Helmand Province, namely in Marjeh District, as well as Lashkar Gah, Nad'Ali, and western Nahr-e Saraj Districts.

Figure 5.7b shows another static representation of the enemy activity for, central Helmand Province. Figure 5.7a shows the difference between insurgent attach densities for 2011 and 2012 time periods for this same region. This representation reveals a relative *change* in insurgent activity from one year to another. Representations like those in Figures 5.7a and 5.7b can greatly improve the situational and contextual understanding of both the analyst and more importantly the leadership and decision makers.

More robust spatial time series analysis can help to quantify these observed trends (Gaul, McAlinden, and Alvarez 2013). We achieve this by measuring the proximity of the population to insurgent activity, or more precisely by considering the average number of Afghans that live within 1 km of each enemy-initiated attack. A geographical information system is utilised to conduct the laborious spatial analysis.

Figure 5.8 shows the result, i.e. the average number of Afghans in proximity of an attack by month for all of Afghanistan since January 2010. We evidently see a continuing downward trend in this metric since 2010 when the average number of Afghans living within 1 km of an enemy-initiated attack was approximately 2000, while in November 2012, around 1300 Afghans were objectively affected on average. Each EIA affects (on average) fewer Afghans, or in other words, the insurgent attacks migrate to less densely populated areas where they are close to fewer and fewer Afghans. The visualized downwards trend is statistically significant.

Modern warfare is creating vast amounts of data. Leveraging this information across time and space can crucially advance operational and strategic goals. It is critical for analysts to be equipped with the right tools, skills and

opportunities to process and manipulate spatiotemporal data generated by a (military) campaign, thus facilitating expedient analysis of enemy actions and blue force operations.

Challenges

During our time at ISAF Headquarters we have encountered several challenges that have impeded our ability to perform effective analysis. They include the cost and time needed to establish an operational database (i.e., data collection, access to the data and data sharing; data quality control and quality assessment; the merging of data sets) and access to analytic software. Beyond these structural challenges, there often exists a gap between decision makers and analysts. The analysts need to understand the issues facing the top level leadership in order to best focus their analysis.

In addition to the large upfront costs of establishing a database structure useable across a command, there is the time and manpower needed to maintain it (e.g. collate operational reports into the database, reconciling differences

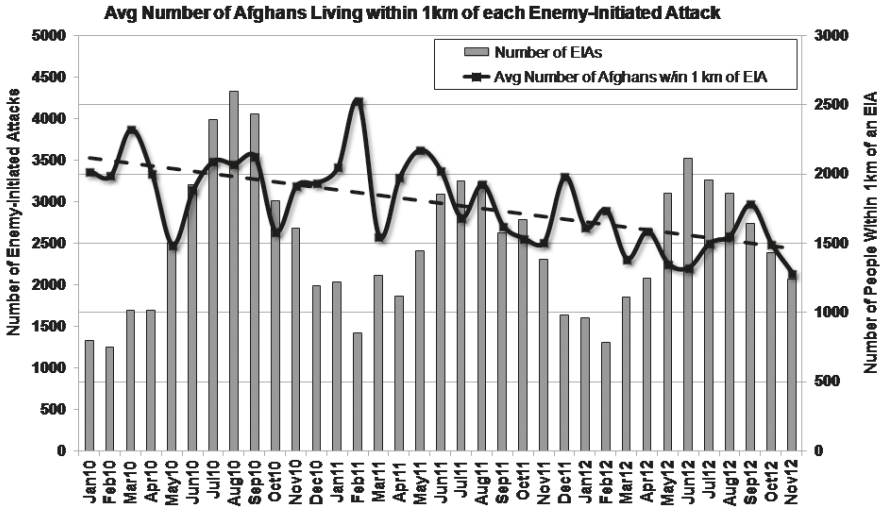


Figure 5.8: Average number of Afghans living within 1 kilometer of each enemy-initiated attack (EIA)

between inputs, resoling lost or incomplete data, etc.). This requires planning and a dedicated funding pool.

Across a theatre of operations different organisations often maintain databases or data repositories for various reasons, such as the collection of a specific type of information that is not collected elsewhere, or because of a lack of knowledge of existing databases. This can cause the analyst additional challenges, the primary ones being: 1) lower standards concerning quality control or data consistency and data integrity compared to the authoritative "command databases," and 2) difficulties in assessing and accessing the data. The ability to access data is crucial to analysis and assessment–additional non-authoritative databases/repositories although potentially well-intended can have counterproductive effects.

Understanding data quality and the constraints it places on subsequent analysis is an important prerequisite to formulating an assessment and making recommendations. The ability to control data quality by cleaning, processing and organising data is a necessary skill for analysts. In some cases analysis that could be done should be avoided because underlying data inconsistencies would lead to conclusions too unreliable to guide decision making.

The data necessary for an analysis may not exist in one repository. The ability to join large datasets is what enables many analyses. Combining datasets and looking across previously different lines of effort allows for the identification of formerly undetected interdependencies and interrelationships that can bring important campaign insights.

Access to analytic software is the final major constraint for the analyst. The tools to manage and manipulate large quantities of spatiotemporal data can be expensive. Funding such tools requires improved awareness of their operational benefits. Similarly, system protocols and security related constraints often hinder access to required software such as MS Excel's VBA (Visual Basic for Applications) or tools publically available for statistical analysis such as R data analysis software.

Operational analysis can provide commanders critical information about the campaign. In this line, analysts need to link directly into the pulse of the campaign and understand current commander priorities, while leadership should seek to remove hurdles hindering analysis. A good analyst can often

help transform commander's priorities and guide valuable research that informs future decisions.

Summary and Conclusion

Leveraging the vast amounts of data collected across time and space in modern military campaigns can crucially advance operational and strategic goals. It is critical for analysts to be equipped with the right tools, skills and opportunities to process and manipulate this data. Time series analysis is one of these highly valuable tools and can generate crucial insights for a command, enabling the assessment of operational effectiveness and developments in insurgent tactics. By combining time and spatial domains it is possible to identify further relationships and opportunities that can be exploited. In this chapter we have presented the fundamentals of time series analysis and discussed some of its uses stemming from our experiences of working in the Afghan Assessment Group (AAG) at ISAF Headquarters in Kabul, Afghanistan. We concluded by presenting some of the assessment challenges encountered during the ISAF campaign, lessons which may help future analysts and commanders better structure their campaigns for success.

References

Allen, O. O., O. Borgan, and H. K. Gjessing. 2008. *Survival and Event History Analysis: A Process Point of View*. Statistics for Biology and Health, Springer Science and Business Media. New York, NY: LLC.

Anscombe, F. J. 1948. The transformation of Poisson, Binomial, and Negative Bionmial data. *Biometrika* 35:246-254.

Anselin, L. 1992. *Spatial data analysis with GIS: an introduction to application in the social sciences*. Santa Barbara, CA: National Center for Geographic Information and Analysis University of California, Santa Barbara.

Box, G. E. P., and D. R. Cox. 1964. An analysis of transformations. *Journal of Royal Statistical Society*. Series B, 26:211-252.

Box G. E. P., G. M. Jenkins, and G. C. Reinsel. 2008. *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons, Inc. 4th ed.

Chandler, R. E., and E. M. Scott. 2011. *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. Chichester, UK: John Wiley & Sons Ltd.

Cook, R. J., and J. F. Lawless. 2010. *The Statistical Analysis of Recurrent Events*. Statistics for Biology and Health, Springer Science and Business Media. 2nd ed., 2002. Corrected Eighth Printing. New York, NY: LLC.

Cox, D. R. 1955. Some statistical methods connected with series of events (with discussion). *Journal of the Royal Statistical Society*. Series B, 17:289-300.

Cox, D. R., and V. Isham. 1980. *Point Processes*. Boca Raton, FL: Chapman and Hall/CRC.

Cressie, N. A. C. 1993. *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics. Revised Ed. New York, NY: John Wiley and Sons Inc.

Cressie, N. A. C., and C. K. Wilke. 2011. *Statistics for Spatio-Temporal Data*. Wiley Series in Probability and Mathematical Statistics. Hoboken, NJ: John Wiley and Sons Inc.

Freeman, M. F., and J. W. Tukey. 1950. Transformations related to the angular and the square root. *The Annals of Mathematical Statistics* 21:607-611.

Gaul, M., R. McAlinden, and P. Alvarez. 2013. *Population-Centric Counterinsurgency: The Effects of Violence on Population Clusters.* (Accepted for publication).

Gons, E., J. Schroden, R. McAlinden, M. Gaul, and B. Van Poppel. 2012. Challenges of measuring progress in Afghanistan using violence trends: The effects of aggregation, military operations, seasonality, weather, and other causal factors. *Defense & Security Analysis* 28(2):100-113.

Hamilton, J. D. 1995. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Jesse, E. 2013. *Identifying Significant Trends in Operational Data.* (Accepted for publication).

Pratt, J. W., H. Raiffa, and R. Schlaifer. 2001. *Introduction to Statistical Decision Theory*. Third Printing. Cambridge, MA: MIT Press.

Priestley, M. B. 1988. *Non-linear and Non-stationary Time Series Analysis*. London: Academic Press.

PART 3:

EXTENDING BEYOND THE MILITARY DOMAIN

6

Civilian and Military Evaluation and Assessment: Synergies and Differences

Fabrizio Fitzgerald Farina, Andrew Williams, Megan Kennedy-Chouane

Abstract

This paper explores the synergies and differences between civilian monitoring and evaluation systems and military operations assessments in five areas: terminology, the use of logical framework approaches, organisational structures, organisational independence, and communications. Both communities used similar terminology and used logical frameworks as the central instrument of the management cycle. Larger differences were observed in the other three areas. The organisation doing the planning (e.g., World Bank) is typically not the organisation executing the project (e.g., borrowing country). In the military, planning and execution occur in the same organisational structure. As a consequence, the two communities have different reasons for conducting evaluations and assessments: civilian systems monitor accountability for the donor; military systems measure progress and inform decisions. In addition, the two communities differ regarding the independence of the evaluator/assessor with civilian evaluations usually viewed as more independent as the performers are not members of the planning (donor) or executing organisations. Finally, internal communication in the military is well developed through a formally-established reporting system, whereas external communication is highly selective. In contrast, international organisations and development agencies show high degree of external exchange of information, but rather informal channels of internal communication. We end by identifying some areas for initial cooperation between the civilian and military evaluation and assessment communities.

Introduction

There is a growing international consensus that the effectiveness of international peace-building interventions should be verified using rigorous evaluation, both as a means to ensure that public funds are spent wisely and as a way of learning more about the impacts of different types of interventions in order to improve performance. In recent years it has become widely accepted that this process of evaluation should be in partnership with all organisations involved and local actors (Picciotto, 2002, 2003, OECD 2012). With the advent of more coordinated approaches to international intervention in conflict the need for better civil-military coordination has become evident. Evaluation in the civilian sphere has likewise moved towards more collaborative approaches (OECD 2010).

Coordinated approaches to international interventions and their evaluation are thought to be more effective, provide opportunities for cross-organisational learning, reduce transaction costs including the burden on local partners, and increase policy coherence. Development and humanitarian actors are increasingly monitoring and evaluating peacebuilding and statebuilding activities, collecting data and producing valuable evidence about the effectiveness of different strategies. This chapter suggests that NATO, and military forces in general, would be well-placed to increase cooperation and collaboration with international development actors in assessing intervention success. While significant differences exist between these actors in terms of institutional set-up and culture, there are also many similarities—importantly in terms of data and methods. As improving results in missions where NATO and development actors are working together requires better mechanisms of cooperation, we must start with a deeper understanding of each other's similarities and differences.

NATO has increasingly emphasised the importance of working cooperatively with other actors, driven by recent operational experiences and international consensus (NATO 2010). Given that success of an overall intervention is interdependent between the various actors present, NATO recognises that evaluation of progress is also interdependent. Recent concept development and experimentation work (NATO 2009), national (US JFCOM 2010) and NATO experiences (NATO 2011b) in operations have confirmed that working interdependently with other actors, requires interdependence in the act of

measuring progress. Likewise civilian development organisations, in particular the members of the Organisation for Economic Cooperation and Development's (OECD) Development Assistance Committee (DAC)[1] recognise that no actor single-handily determines the outcome of an international intervention; assessing progress therefore requires taking into account the broader context and actions of others.

The 2005 Paris Declaration on Aid Effectiveness exhorts development partners to work together and to improve alignment with national priorities and systems, including in monitoring and evaluation (OECD 2005). Thus by improving cooperation, NATO would contribute to a growing international standard of practice in aid, development and peace-building interventions. These international standards are derived from the political necessity to improve and better demonstrate effectiveness—or lack thereof—of interventions.

In order to initiate discussion about encouraging cooperation and collaboration between civilian and military organisations in terms of measuring the progress of international interventions, this chapter presents a comparison between monitoring and evaluation (M&E) in civilian development organisations and the military equivalent—operations assessment—within the NATO system. We illustrate that while the underlying logic behind military assessment and civilian evaluation is similar, the intended uses are quite different. We make some suggestions for how synergies can be exploited, and how differences can be bridged, providing examples of potential avenues for collaboration.

The chapter begins with an overview of monitoring and evaluation in international (civilian) development cooperation and NATO operations assessment. Then, we present the terminology utilised in M&E and operations assessment. Second, we compare and contrast the use of logical frameworks in the civilian and military systems. Third, we analyse the two different structures of M&E and operations assessment and compare the importance of the principles of accountability. Fourth, we analyse independence as the central tenet of

1. The OECD DAC works to make development co-operation more effective by collecting data on development finance and issuing policy guidance, monitoring both the quantity and quality of official development assistance (ODA). For more information, see website www.oecd.org/dac. The DAC includes a number of specialised subsidiary bodies, including the DAC Network on Development Evaluation which brings together the central evaluation units of DAC member countries as well as seven multilateral organisation's independent evaluation units.

evaluation processes. Fifth, we investigate the communicative channels and different products whereby findings are disseminated within and outside the organisation. Finally, we summarise the key findings and suggest some key conclusions and recommendations for NATO.

Monitoring, Evaluation and Operations Assessment: An Overview

Monitoring, results-based management and evaluation play distinct but complimentary roles in international development co-operation today. While monitoring is primarily a routine management function, evaluation takes a step back to assess the value and worth of activities—to question not just whether pre-defined objectives were achieved, but also whether these are the right objectives in the first place, whether they were achieved efficiently and whether the benefits will be sustained overtime (OECD DAC 2002). Evaluation should provide credible evidence about effectiveness and results. In development circles, a broad consensus has formed around the use of five main criteria for evaluating development activities: relevance, efficiency, effectiveness, impact and sustainability. The criteria are widely used by both bilateral and multilateral donors and other development actors (implementing agencies, non-governmental organisations (NGOs), humanitarian organisations, etc.) supporting poverty reduction, economic growth and peacebuilding activities in developing countries (OECD 2010).

There is a long history of results-based management and evaluation in international development.[2] This is in part a reflection of the fact that, unlike other areas of public policy, development activities take place outside the scope of normal accountability mechanisms. The need to create stronger feedback loops has thus resulted in intense pressure to evaluate development activities. The Paris Declaration on Aid Effectiveness has also placed a high priority on results based management of development co-operation and mutual

2. There is currently a very lively debate in development circles about the value of results based approaches to aid and there has recently been some backlash to this approach, driven primarily by concerns that a donor-driven results agenda undermines country ownership and that a narrow focus on measuring quantitative results will detract from investing in complex, political change processes required for long –term development. See Vahamaki, Schmidt, and Molander 2011. For the purpose of comparison this article focuses on the established evaluation (and to some extent monitoring and results based management) systems of the World Bank and the members of the OECD DAC.

accountability between development partners, increasing focus in these areas (OECD 2005). In recent years, there has been a move to use more rigorous evaluation techniques to assess not just inputs (funds, technical assistance, etc.) and outputs (such as teacher training) but higher order impacts (such as the effects of teacher training on student learning outcomes and life-time earnings). In the context of development co-operation, monitoring and evaluation are also driven by what is perceived as a "learning gap"—a real need to better understand how poverty reduction and economic growth "work" and how international partners can best support development processes.

Both approaches share the common aim of improving performance by providing reliable information about the results of interventions. While the civilian side often references the dual function of learning and accountability (and sometimes the tradeoff between the two), military assessment is primarily focus on learning. This is in large part due to the different institutional settings, described further below.

In development agencies, monitoring and results information are used internally and often only readily available to staff and management within the organisation. In recent years there has been a move towards improving transparency around aid data at the programme level and improving reporting of activities and results, also discussed in more detail below. Evaluation has a long standing track record of transparency, as this is considered a critical element of credibility. The OECD DAC has established a clear set of norms and standards for development evaluation, against which all major bilateral donors are reviewed periodically in peer reviews. Two core principles of evaluation are credibility and transparency. Nearly all OECD DAC members have a legal mandate to make all evaluation findings publicly available. Most reports are published online by the agency themselves and also by the OECD DAC on its Evaluation Resource Centre website.[3] National policy mandates for evaluation help ensure that the evaluation department has final say on the conclusions and recommendations, to avoid pressure from management or other potential bias. This contrasts with the military approach, which tends to be much more internally-focused and where sharing of information is blocked by various procedural and security barriers.

3. See www.oecd.org/derec

Evaluation/assessment functions reflect the broader institutional context (be it humanitarian, development or military) in which they operate. In the civilian development context monitoring and results based management are used at the program level (in-country) and also feed into results reporting and decision making across the institution. Evaluation is also usually split into at least two functions, with internal or self-evaluation taking place within programs (either carried out by staff or by external consultants overseen by program staff) and a central, independent evaluation unit set up outside of the management structure of the institution. The role of these evaluation units (represented by the membership of the OECD DAC Network on Development Evaluation) is usually to provide objective, critical perspectives on the effectiveness and results of the institution's development activities, often by providing more strategic-level evaluations (OECD 2010).

While the command structure in NATO and military command structures essentially functions in one direction, development evaluation on the civilian side is more of a multi-directional activity, providing feedback and (when it works well) giving critical reviews and even questioning assumptions and "speaking truth to power." Of course development organisations also suffer from a positive-bias, where results reporting tends to confirm pre-existing ideas and evaluation findings are used selectively for political ends. Still the aim (and accepted good practice) of a credible evaluation system is to provide objective evidence on results—whether or not the reality of these results is palatable for staff, management or policy makers. In contrast, assessments in NATO primarily focus on the whether or not stated objectives have been achieved. As discussed further below, good practice in civilian evaluation would also involve questioning the objectives themselves and taking a critical look at strategies being pursued.

In the context of peacebuilding and statebuilding interventions, further differences are created by the different types of activities carried out by civilian and military actors and the overall "philosophy" of intervention. While civilian development actors view statebuilding and peacebuilding as primarily endogenous practices that can be supported (or undermined) by international partners (OECD 2011b), military interventions are viewed as a self-contained and not necessarily requiring external support—although current policy calls for comprehensive approaches and consideration of local perspectives. Military

activities are likely more readily comparable to humanitarian interventions in terms of pursuing relatively well-defined short-term objectives in an emergency context. In contrast, the time scale and breadth of development interventions are much more ambitious, making their evaluation also more complex.

The relative chaos of the international aid system (compared to a fairly streamlined command structure of military actors) is also significant when trying to understand and compare the two. For aid, public funds go through multilateral organisations (like UNICEF), NGOs (like Save the Children), development agencies (USAID) or directly to implementing partners in the country (like a local NGO in Afghanistan) or the partner government (Ministry of Health). Each of these different development operations is governed by particular rules and the responsibilities for evaluation and results management and reporting.

Despite these differences, the tools for monitoring, evaluation and operations assessment are quite similar. Staff turnover, poor knowledge management, low institutional memory and weak data are all common challenges in both civilian and military domains. The fast-paced rate of departure is indeed a threat to continuity of the overall evaluation effort that development agencies and military apparatuses need to address carefully. An additional point of commonality between the two systems can be seen in the context and political-economy of international intervention in violent conflict. Both military and development actors are under pressure to demonstrate positive results and to reassure anxious politicians and tax payers that they are not wasting (scarce) public funds on futile peacebuilding adventures overseas. Such political dimensions make the technical aspects of monitoring and evaluation (selecting appropriate indicators to monitor security sector reform, for example, or attributing positive changes in the security context to a particular intervention) immensely more complex.

Civilian Monitoring and Evaluation and Military Operations Assessment Terminology

The first step in order to grasp the synergies and differences between civilian and military is to compare the terminology of the two systems. Table 6.1 presents a set of definitions of the most basic lexicon utilised by civilian and military organisations for evaluation and assessment purposes. Other more specific definitions will be presented throughout the chapter. Although civilian international organisations (IOs), NGOs, and development agencies use slightly different terminologies, the glossary developed by the OECD is widely accepted as an international standard. Therefore, the civilian definitions are derived from OECD Development Assistance Committee's (DAC) *Glossary of Key Terms in Evaluation and Results Based Management* (OECD DAC 2002). On the military side, we looked at the definitions formulated in the *NATO Operations Assessment Handbook* (NATO 2011a).

Table 6.1 illustrates each concept's meaning and also clarifies the relationship between different concepts in the civilian and military domains. Undoubtedly, civilian and military concepts differ by virtue of the two very different domains—development and aid on the one hand, and security on the other—in which they are applied. By taking a closer look, however, we observe overlap in the definition of the terms.

Logical Frameworks in Civilian and Military Systems

A logical framework, or 'logframe,' is the core analytical tool and reference document used throughout the entire project management cycle to design, plan, implement, monitor, and evaluate projects (Williams and Morris 2009). Logframes are a standard technique in the civilian context, falling under the general rubric of results-based management (RBM). In the past decade, NATO and many of its member nations have also adopted RBM by shifting from the traditional military planning based on the input-target nexus to the outcome-centred Effects-Based Approach to Operations (EBAO) (NATO 2011a).

Table 6.2 describes the core features of logframes and compares the civilian and military approaches. As far as the former is concerned, The World Bank's *The LogFrame Handbook* (The World Bank 2005) is benchmark for civilian development organisations. With respect to the latter, we considered the information contained in the *NATO Operations Assessment Handbook* (NATO

Table 6.1

Comparison of civilian and military terminology

Civilian		Military	
Monitoring & Evaluation	Monitoring – A continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds. Evaluation - systematic and objective assessment of an on-going or completed project, program, or policy, and its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact, and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision making process of both recipients and donors.	*Operations Assessment*	Activity that enables the measurement of progress and results of operations in a military context, and the subsequent development of conclusions and recommendations that support decision-making.
Indicators	Quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor.	*Measure of Effectiveness (MOE)* *Measure of Performance (MOP)*	Metric used to measure current state system. Metric used to determine the accomplishment of actions.
Goal	Higher order objective to which a development intervention is intended to contribute.	*End State*	NAC statement of conditions that defines an acceptable concluding situation for NATO's involvement.
Objectives	Intended physical, financial, institutional, social, environmental, or other results to which a project or programme is expected to contribute.	*Objectives* *Decisive Condition*	Clearly defined and attainable goal to be achieved. Combination of circumstances, effects, or a specific key event, critical factor, or function that when achieved allows commander to gain a marked advantage on opponent or contribute materially to achieving an operational objective.
Results	Impact: Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended. Outcomes: the likely or achieved short-term and medium-term; effects or changes of an intervention's outputs. Outputs: the products, capital goods, and services that result from a development intervention.	*Effects*	Change in the behavioural or physical state of a system (or system elements), that results from one or more actions, or other causes.
Activity	Actions taken or work performed through which inputs, such as funds, technical assistance and other types of resources are mobilized to produce specific outputs.	*Action*	The process of engaging any instrument at each level in the engagement space in order to create (a) specific effect(s) in support of an objective
Inputs	Financial, human, and material resources used for the development intervention.	*Resources*	Available resources.

Adapted from OECD DAC 2010 and NATO 2011

2011a). Recalling the definitions provided in Table 6.1 and despite the fact that the concepts have different names, the features of the two systems look similar. For the sake of simplicity, we omit some crucial technical aspects of RBM processes (i.e. risk management, data collection techniques).[4] Nonetheless, it is worth analysing the similar logical underpinnings of the two systems.

Results-based management (RBM)

First of all, in both systems RBM puts forward a causal logic: inputs (or resources), activities (or actions), results (or effects), objectives (and decisive conditions) must be logically linked to one another and aligned to the achievement of the overall goal (or end state). Second, both systems utilise quantitative and qualitative measures—QQT (quantity, quality, and time) indicators in the case of civilian organisations[5] and MOEs and MOPs in the military—in order to evaluate the progress of the intervention. Third, the use of targets to judge performance is common to both systems.

Development agencies use milestones and triggers to estimate how much success is required to move to the next phase. The rate of 'success' can be estimated in different ways: where the approach includes quantitative indicators targets will be represented by a numerical threshold to measure *how much* change is required to achieve a particular objective. On the other hand, when qualitative indicators are selected targets will be more informational and determine *what kind* of change is considered necessary to accomplish an objective (The World Bank 2005). Similarly, the military uses a combination of one or more targets, such as acceptable condition, rate of change, thresholds of success (ToS), and thresholds of failure (ToF), in order to set a level or a tipping point at which a desirable situation has been achieved.[6]

4. The discussion of the very technical aspects goes beyond the scope of our study. For more detailed information, please consult The World Bank 2005 and NATO 2011a.

5. Civilian organisations that utilise these QQT Indicators (also referred to as OVIs, objectively verifiable indicators, or SMART, specific, measurable, attainable, relevant, and timebound) include: DFID, SIDA, DANIDA, SDC, The World Bank, AFD, NORAD, KFW/GIZ, OXFAM, AfDB.

6. Acceptable Condition is defined as "a target level for the metric at which a desirable situation has been achieved;" Rate of Change is referred to as "the change measured in a metric over time during an operation;" Threshold of Success is "a tipping point at which a positive level of achievement becomes unstoppable and most likely irreversible;" Threshold of Failure is "a tipping point at which an unrecoverable situation is reached" (NATO 2011a, 3-7, 3-8).

Table 6.2

Civilian vs. Military Logframes

Results-Based Management		Civilian	Military
Set Goal		Define Country Assistance Strategy (CAS) Goal	Define End State
Formulate Objectives		Define Development Objectives (D.O.), Outcomes, Activities	Define Objectives, Decisive Conditions, Effects, and Actions
Select Indicators		Select QQT Indicators (measurable in quantity, quality, and time) or proxy measures for Goal, Objectives, Impact, Outcomes, and Outputs	Define Measures of Effectiveness (MOE) for End State, Objectives, Decisive Conditions, and Effects. Define Measures of Performance (MOP) for each Action
Set explicit targets for each Indicator to judge performance		Set Triggers and Milestones	Establish Targets, Acceptable Condition (AC), Rates of Change (RoC), Thresholds of Success and Failure (ToS and ToF) for each MOE and MOP
Establish Monitoring and Evaluation design (civilian) - Operations Assessment system (military)	*Review, analyse, and report actual results vis-à-vis the targets*	Monitor day-to-day Activities and implementation; collect testimonies from customers and target groups via surveys and interviews; encourage team to look for evidence of impact of Activities, Objectives towards the achievement of CAS Goal	Conduct assessment of Effects, Decisive Condition and Objective status; conduct assessment of Action status to determine progress towards End State
	Integrate evaluations to provide complementary performance information for decision making	Evaluate linkages between Results, Objectives, Activities and Inputs and present findings at Mid-year and Annual Performance Improvement Planning Workshops; Year End Review Workshops; Annual Strategic Planning Retreat	Conduct assessment of Effects versus Action status; Decisive Condition versus Effect, and Objective versus Decisive Condition status to find insight as to why results were or were not achieved; use results of the assessment to inform the Commander and staff; identify required changes to both plan and assessment

Adapted from The World Bank 2005 and NATO 2011

Accountability

In contrast to the similarities described above, there is a key difference between civilian and military approaches in terms of the use of logframes—the aspect of accountability. OECD defines accountability as:

[The] obligation to demonstrate that work has been conducted in compliance with agreed rules and standards or to report fairly and accurately on performance results vis-à-vis mandated roles and/or plans. This may require a careful, even legally defensible, demonstration that the work is consistent with the contract terms. Note: Accountability in development may refer to the obligations of partners to act according to clearly defined responsibilities, roles and performance expectations, often with respect to the prudent use of resources. For evaluators, it connotes the responsibility to provide accurate, fair and credible monitoring reports and performance assessments. (OECD DAC 2002, 15)

Amongst other organisations engaged in development assistance The World Bank, for instance, provides funds, in terms of grants or loans, to borrowing countries that are responsible for the implementation of the program or project (The World Bank 2011a).[7] In this light, the question of accountability—intended as the obligation of the partner, namely the borrowing country, to spend the money appropriately—becomes a key aspect from the donor's viewpoint. M&E systems, as part of logframes, are the template against which parties are held to account. Logframes are primarily used as an internal management tool and are meant to guide programme design and management, but also feed into to broader results management and reporting, including in the context of partnership agreements between the donor and the implementing actors (whether a borrowing country or a local NGO, for instance).

Logframes in the military context are the underlying fabric of a military operational plan, although they are often not identified as such. Military organisations as direct agents of a State are responsible only for efficient and legal

7. The OECD defines official development assistance as cash, commodities or services to countries and territories on the Development Assistance Committee List of Official Development Assistance Recipients and to multilateral development institutions which are: 1) provided by official agencies, including state and local governments, or by their executive agencies; and 2) each transaction of which: a) is administered with the promotion of the economic development and welfare of developing countries as its main objective; and b) is concessional in character and conveys a grant element of at least 25 per cent (calculated at a rate of discount of 10 per cent) (OECD DAC 2008).

operations in-theatre and assessments of progress are focused internally within the military structure. While military organisations are politically accountable to government leadership, the aspect of accountability is not a focus point in military operations assessment. The aspect of accountability will be further clarified in the next section where the different structures of civilian M&E and military assessment processes will be analysed.

Civilian Evaluation and Military Operations Assessment Structures

In this section we draw a comparison between civilian M&E and military Operation Assessment structures. In order to illustrate the point, we compare the World Bank structure in the civilian domain with the operations assessment process within NATO,[8] recognising that there is a large variation in organisational structure within the civilian domain (see OECD 2010 for a full discussion). Overall, the two structures present some similarities. First, M&E and operations assessment consist of multilevel structures. The World Bank describes the levels as follows:

- Program level – Evaluation: reviewing early evidence or leading indicators of Development Objectives and CAS Goal;
- Management level – Performance Monitoring: reviewing achievement of Outputs and progress towards the PDO [Program Development Objective] outcome;
- Operational level – Progress Monitoring: reviewing actual activity implementation status vs. time plans and disbursement schedule and milestones. (The World Bank 2005, 49-50)

Bilateral donors are also increasingly carrying out assessments at the country level and at the institutional level, to guide policy making. Thematic or sector evaluations, looking at operations in a particular area (women's empowerment or health sector support, for example) are used to provide more generalizable lessons and inform strategic planning (OECD 2010).
In NATO, the three levels are:

8. We assume that NATO is a good example of military in general. US, UK, Canada, Netherlands, Norway, and Germany use, indeed, doctrines very similar to NATO.

- Strategic level – Strategic Assessment: function that involves varying combinations of: continual measurement of strategic progress and results in non-military domains; an overall evaluation of progress towards NATO End State;
- Operational level: continuous monitoring and evaluation of all Effects and Objectives specified in the operational level military plan; assessment of desired and undesired Effects across all PMESII [political, military, economic, infrastructure, and information] domains, where they significantly impact the operational level military plan, or where they are explicitly stated in the military plan;
- Tactical level – Progress Monitoring: measuring the achievement of planned actions, tasks or activities using MOP for each particular component; in some cases, measuring the achievement of Decisive Conditions and creation of operational Effects using MOE. (NATO 2011a)

Second, both management cycles involve top-down planning and a bottom-up execution. This means that goal (end state), objectives (and decisive condition), and indicators (MOPs and MOEs) are formulated at the highest level, whereas the actual measurement of results and reporting are produced from the lowest level.

Let us turn now to analyse the key differences in the two structures. A crucial aspect is the ownership of the top-down planning and bottom-up execution phases. The whole process of operations assessment from top-down planning to bottom-up execution at all three levels occurs *within* the military structure. Whereas M&E in civilian organisations is split up with the bottom-up execution phase taking place *outside* the structure of the organisation. While top-down planning is a unique responsibility of the development agency, program implementation and M&E are generally a responsibility of the borrowing country as showed in Figure 6.1. Program level impact evaluation is carried out by the development agency itself (in the case of The World Bank, by the Independent Evaluation Group). Thus it is challenging to clearly demarcate functions during bottom-up execution between the donor and the borrower. There is, indeed, collaboration in the monitoring phase between the donor agency and the borrowing country (especially at Management level) although the responsibility formally remains in the hands of the borrower.

As stated in the *Good Practice Note for Development Policy Lending*:

> While implementation is the responsibility of the borrower, the Bank can play the following significant roles: (a) supporting implementation by the borrower, including through complementing a development policy operation by a separate parallel technical assistance operation; (b) focusing the Bank's own supervision on results, and (c) ensuring timeliness of supervision. […] In light of the importance of borrower ownership of the program, [The World Bank] reiterates borrower responsibilities for implementation of the actions supported by the development policy operation, monitoring of progress during implementation, and evaluation of results on completion. The role of the Bank in the process of supporting implementation by the borrower can be summarized as follows: assess borrower implementation capacity […]; assess the M&E systems to be used by the borrower […]; include capacity enhancement measures […]. (The World Bank 2011a, 15-16)

The separation of bottom-up execution from top-down planning in civilian structures leads to the second main difference with the military—the principle of accountability. We have already briefly mentioned this difference

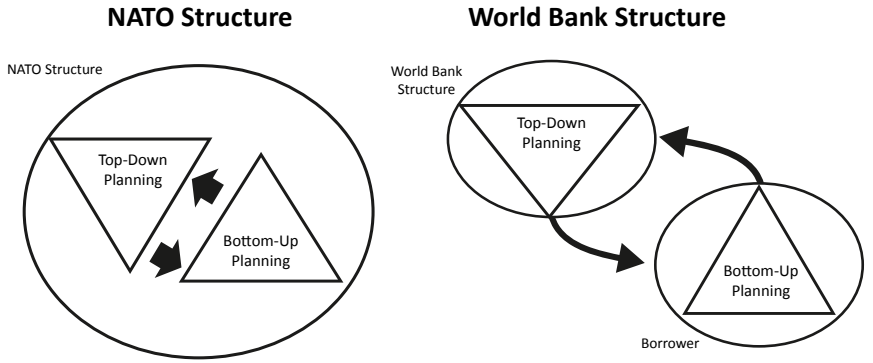


Figure 6.1: NATO structure vs World Bank structure

in the previous section; however, in light of the analysis of the two different structures it deserves further emphasis. The importance of accountability within civilian organisations is intertwined with the structure of M&E, and in particular it is correlated to the fact that the implementation is responsibility of the borrower. Donor agencies must be certain that the funds are being spent appropriately by the borrower thereby making accountability a crucial point in the evaluation process. In contrast, accountability is not considered in the same manner in military operations assessment, as the military has a single formal chain of reporting.

Independence

In this section we present an analysis of the principle of independence in relation to the process of evaluation in civilian and military systems, because it represents the cornerstone of sound evaluations and assessments. OECD defines independence in the evaluation process as follows:

> An evaluation carried out by entities and persons free of the control of those responsible for the design and implementation of the development intervention. Note: The credibility of an evaluation depends in part on how independently it has been carried out. Independence implies freedom from political influence and organisational pressure. It is characterised by full access to information and by full autonomy in carrying out investigations and reporting findings. (OECD DAC 2010, 24)

The assumption is that the higher degree of independence evaluation units enjoy, the more reliable and rigorous the assessment. On the one hand, we present a general picture of development agencies, and then we turn to look more in depth at The World Bank's Independent Evaluation Group (IEG). On the other hand, we will investigate independence within NATO. The analysis hinges around three features: a) organisational independence; b) behavioural independence; and c) protection from external influence. OECD (2010) provides an overview of independence across bilateral and multilateral agencies (using basically the same framework).

Organisational independence

In most development organisations and agencies the evaluation unit reports directly to a governing body. Many development agencies, indeed, adopt the 'separate unit model,' which implies that the evaluation unit is managerially separated from the staff responsible for the program activities (Foresti et at 2007). In other words, evaluators—either staff members or staff and consultants hired outside the agency that cannot be hired at the end of the mandate—are not subordinate to management or operational bodies of the organisation. This is, for instance, the case of IEG. According to its mandate, indeed, IEG's Director General (DGE) "is directly responsible to the Executive Director/ Board of Directors of IBRD/IDA, IFC, and MIGA [the three main bodies that constitute The World Bank Group]" (IEG 2013, under "*The Director-General*").

In NATO, most operations assessment is conducted by specialised assessment units or by personnel attached to plans or intelligence organisations. While objective analysis is part of their organisational mandate, they cannot be considered as independent. Recent experience has shown that assessment systems are highly dependent upon the commander (see Schroden 2013 in this volume), and often political imperatives strongly affect the detailed assessment design, even down to the indicator level. Furthermore, military assessment tends to emphasise progress measurement against already defined objectives, rather than critical evaluation of whether planned objectives are right.

Behavioural independence

AFD and ODI define behavioural independence as "the capacity and space for the evaluation function to critique the agency's work and to make recommendations for improvement" (Foresti et al. 2007, 18). In the case of The World Bank, IEG carries out self-evaluation with the purpose of shaping lessons from the implementation of programs; produce recommendations to improve the policies applied; and address the shortcomings. IEG fulfils these functions under no management-imposed restrictions that could curb transparency and limit the disclosure of the findings (The World Bank 2011b). It is also important to point out, however, that the IEG's reports are created almost exclusively by staff members and may be somewhat isolated from external critiques. A slightly different interpretation of behavioural independence corresponds to the idea of a sort of "evaluation of evaluation." Many development organisations—with

the exception of IEG—hire external consultants to carry out an examination of the evaluation carried out by the agency's staff. In this case, however, external consultants could be under pressure in order to assure the contract thereby producing an evaluation that is not completely unbiased.

In NATO, there is no formal requirement for meta-evaluations or internally-focused self-assessments. The Joint Analysis and Lessons Learned Centre (JALLC) conducts lessons learned, but these concern organisational effectiveness issues rather than operational progress or evaluation. From 2009 onwards, SHAPE has contracted, via the NATO Communications and Information Agency (NCIA), an independent strategic level assessment capability for the ISAF mission. Yet this was meant to inform the strategic commander on progress and other requested issues, rather than be an independent evaluation function in the same sense in the civilian context.

Protection from external influence

Protection from external influence means that the evaluation unit performs the evaluation without any interference; the findings cannot be changed by an external authority; and the members of the evaluation group do not perceive any pressure in relation to their appointment, in terms of performance appraisal, renewal, and compensation. It also refers to the capacity of evaluation units to set their own work agenda, establish the necessary manpower, and define the budget (The World Bank 2011b). Civilian agencies show a mixed picture: in some agencies, evaluation units submit the Board a budget proposal based on the planned initiative for the coming one, two, or three years; in other cases, the budget is authorized independently from the management, and established directly by the Board (Foresti et al. 2007, 19). The IEG produces a three-year work plan and related budget and submits it for the approval of the Board. It is important to note that the IEG informally consults management and operational staffs during the preparation phase of the work plan and budget in the attempt of mitigating potential tensions between the different departments (The World Bank 2001b).

In NATO, as the assessment function is entirely within the organisation it is not meaningful to consider the question of external influence in the same context as the civilian case. Military assessment staff may perform many other functions or activities falling under responsibility of the military commander.

Furthermore, certain measures of progress may be kept classified due to operational secrecy concerns, or false figures may be published for deterrence and deception operations. Taking into account all institutional factors, in strict comparison with the civilian evaluation context, the military system of operations assessment cannot be considered truly objective nor free from influence.

Communication and Distribution

The last theme we analyse consists of the distribution phase of evaluation findings. Communication is two-sided: external communication has the objective of involving the widest possible audience in the distribution of evaluation results; internal communication represents mechanisms and channels whereby lessons learned are shared; there is an exchange of views through the provision feedback; and the overall capitalisation of the experience occurs. Moreover, depending on the type of communication, different products are developed to properly fit the communicative purpose. In the civilian context, the systematic dissemination of evaluation findings is a fundamental pillar in the transparency and credibility of the evaluation process. It is considered particularly important for ensuring that findings are used and responded to appropriately and to avoid bias or the suppression of "negative" findings.

With respect to external communication, there is a growing emphasis in the development world to make as much information as possible available online. Only a few development agencies, however, have defined clear disclosure procedures. The most common products for external communication are evaluation reports and annual reports. Other products, although not systematically utilised, are e-newsletters, short synthesis papers, policy briefs, and seminars. Seminars are also a communicative tool utilised for internal communication in order to share lessons learned and exchange feedback on the initiatives undertaken. IEG, for example, has a comprehensive range of products accessible online and explicit dissemination procedures. IEG has recently started to engage with the media with the objective to publicise press releases and press conferences. Many other development agencies are following suite, and the OECD DAC maintains a database of evaluation reports from its members' development agencies.[9]

9. See, for example, http://www.oecd.org/derec/

Internal communication tools or the way in which evaluation studies are used varies widely across development agencies. Mechanisms to follow up and monitor the agency or organisation's response to recommendation are in most cases informal (Foresti et al. 2007). With respect to internal communication mechanisms, IEG's recommendations are endorsed by the Board through the Committee on Development Effectiveness (CODE). CODE, in turn, reports every year on the progress made.[10]

In NATO and any military organisation in general, external communication of assessment results is extremely limited, mainly resulting from the firmly entrenched institutional and bureaucratic culture of secrecy. As subordinates several layers down the chain of command, assessment units have no authority to determine which information should be released. External exchange is usually of two types. First, ad-hoc agreements that arise in the context of a particular mission, such as the quarterly report that ISAF provides to the UN. Second, militaries release regular information to the public, but these can be interpreted as public information for strategic communications, in addition to fulfilling the need for openness and transparency.

In terms of internal communication, the military hierarchy provides assessments with a high rate of utilisation. Certainly, within NATO, operations assessment is part of the formal operations doctrine and is institutionally mandated, although the particular format and design may vary significantly with commander and missions' context. The extent to which assessment information affects decision making, however, is uncertain. Assessment reports typically are set against complex and dynamic mission situations, and there is rarely a "decision situation" where assessment *causes* a clear and decisive choice. Furthermore, in common with the civilian domain in both the development sector and government, evaluations and assessments typically are used in a variety of different ways, other than purely instrumental (Weiss 1979).

10. CODE acknowledges the recommendations provided by IEG and informs the Board through informal meetings. An example on "Improving Effectiveness and Outcomes for the Poor in Health, Nutrition, and Population" can be retrieved at: http://siteresources.worldbank.org/EXTWBASSHEANUTPOP/Resources/code.pdf

Conclusion

This chapter has conducted a comparison between key elements of the civilian and military assessment and evaluation systems by underlining the main synergies and differences. This final section re-emphasises the main findings and Table 6.3 summarises the main similarities and differences discussed.

Table 6.3
Summary of Synergies and Differences

Similarities	Differences
Terminology	Structure
Logframe	Accountability
Evaluation Approaches	Independence
Staff Turnover	Communication

Terminology

First, despite the fact that the M&E and operations assessment concepts are termed differently, quite similar definitions emerge in each concept. Evidently, a complete overlap cannot exist and some crucial differences remain. These are attributable to the different nature of civilian and military interventions on the ground that influences the scope and the aim of evaluations, plus the institutional heritage that generated the terminology.

Logical frameworks

Second, we looked at logframes as the principal tool utilised to arrange the management cycle. Logframes were developed and first utilised by development agencies but later adopted also in the military as a result of the shift from traditional planning to EBAO. Although the military applies somewhat different instruments throughout the management cycle, the use of logframe can be considered another similarity between the two systems, although the military do not often equate a military plan with a logframe. Evaluation is an integral part to both civilian program management cycle and military OPLAN. The overall objective of carrying out evaluations is to assess progress. To do so, development organisations and military apparatuses employ a mix of quantitative and qualitative methods.

Structure

Third, military and civilian organisations' structures appear to be multilevel with a management cycle that consists of two phases: a top-down planning and a bottom-up execution, however, the key difference lies in the fact that in the military the two phases take place *within* the military structure, whereas in development agencies the bottom-up execution occurs *outside* the civilian organisation's structure. This implies that the responsibility to carry out M&E at management and operational levels is in the hand of the borrower country-although development agencies often step in and provide technical assistance. A direct consequence of this point is the need to consider accountability. Since development agencies lend funds to borrowing countries, which are responsible for the implementation of the programs, they have to ensure that the resources are spent appropriately. A primary purpose for evaluation is as a mechanism of accountability. In contrast, military evaluation—operations assessment—is mainly concerned with measuring progress for internal decision making, rather than demonstrating accountability.

Independence

Fourth, we scrutinized the concept of independence by looking at: a) organisational independence, b) behavioural independence, and c) protection from external influence. While this is a complex issue, we can make a broad generalisation that civilian evaluation systems, in the development field, are generally more independent on all three accounts that the military system, which is highly integrated within the planning and management branches of the military organisation. In the context of civilian evaluation, independence makes sense given that evaluations are conducted to ensure accountability for donor funds committed. In the military context, though accountability is not a primary purpose for operations assessment and organisational independence is lower, the credibility and validity of assessments must nonetheless be insured.

Communication of findings

Fifth, we analysed the internal and external communication channels of assessment and evaluation products. At the program level, there is a high degree of external communication in the civilian evaluation system, especially in major international organisations such as the World Bank, and within government

development agencies. Internal communication of evaluation findings is very variable, but there are questions about exactly how evaluations drive decision making and what impact they have. The military system is the converse: external communication is almost non-existent, but internal communication is very high and assessment reports have a high utilisation rate, even if the particular type of utilisation is not well-specified.

Building on Synergies—Ideas for the Future

The purpose of this chapter was to start a discussion about the differences—and potential synergies—of evaluation and assessment activities conducted by military and civilian organisations intervening in crisis situations and conflict affected areas. In this conclusion, we highlight some potential areas where the military operations assessment and civilian evaluation communities could pragmatically engage with each other. While there are obvious political and institutional challenges in the idea of cooperation between civilian and military worlds, we offer these suggestions in the spirit of opening an initial debate.

Some efforts have already been made and can serve as useful points of departure. For instance, the OECD DAC's guidance on *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility* (OECD 2012) was primarily developed with a development audience in mind. However, during the testing of the guidance, it garnered strong interest from various military actors. For example, the Swedish military used the guidance to evaluate its Provincial Reconstruction Team interventions in Afghanistan. While the evaluation report is unpublished and classified, feedback from the evaluation team showed that the guidance itself was in fact very useful in military assessment context and, importantly, pointed to the value of the DAC criteria and key development evaluation concepts like testing theories of change and using conflict analysis as an analytical framework (OECD DAC and NORAD 2011).

Other civilian evaluation resources, such as the OECD DAC's *Quality Standards for Development Evaluation* (OECD 2011a) or recent real-time evaluation in the humanitarian community (Cosgrave, Ramalingam, and Beck 2009) could usefully be adapted to NATO work. Development evaluation experience with participatory evaluation approaches and assessing longer term change processes also has potential applications. Likewise, the methodology developed in ISAF for conducing large scale evaluations across multiple lines

of operations may well be of interest to civilian agencies working in conflict settings.

Another area for potential collaboration is in sharing data. Projects such as the UK Department for International Development (DFID) financed Coffey International project to produce a solid database of basic household and development statistics in Helmand Province in Afghanistan (Ahmar and Kolbe 2011) can provide a common basis for monitoring and evaluation work. In most contexts where international development and military actors are intervening to support peacebuilding and statebuilding, statistics are sorely lacking; without good data, evaluation and assessment is impossible. Data collection and management (as discussed elsewhere in this volume) is thus a logical starting point for collaboration. Furthermore, statistics capacity development and data collection efforts are often funded by the same NATO/OECD governments. In theory, this should mean they can be shared with relatively few barriers of security; however, in practice, this may not always be the case.[11]

Finally, there is scope for collaboration in testing common theories of change–that is probing the underlying logic of why certain activities are thought to achieve certain objectives (Vogel 2012). The German government partnered with a group of academic researchers (Böhnke, Koehler, and Zürcher 2010) to rigorously test the widely held assumption that financing aid projects (such as water wells, food aid, roads or other small scale infrastructure) can help improve local attitudes towards international actors and the central government in Afghanistan. Evaluation work in South Sudan (Bennett and Farran 2010) explored whether funding social services (health, education, etc.) would actually create a peace dividend and reduce violent conflict. Such evaluative efforts help build up the knowledge base about what types of strategies, programmes and activities are most effective at achieving shared peacebuilding and statebuilding objectives–regardless of whether these objectives are being pursued in the context of development co-operation or through NATO efforts. In any case, learning from such exercises should be more broadly shared to inform strategic planning and future interventions.

11. A workshop held between NATO operations assessment policy and operators, and civilian evaluation experts from development agencies and consultancies, in Paris, May 2012, highlighted—from recent mission experiences—the significant bureaucratic challenges in resolving this problem.

Both military and civilian development actors must accept that they are operating in fundamentally complex, political settings. No single technical solution, no perfect indicator set or log-frame, will solve all of the myriad learning and accountability challenges faced in the context of international interventions in violent conflict and state fragility. On this front civilian and military actors will find much common ground.

References

Ahmar, S. and C. Kolbe. 2011. *Innovative Approach to Evaluating Interventions in Fragile and Conflict-affected States: The Case of Helmand*. Working Paper. Sydney: Coffey International Limited.

Bennett, J. and A.G. Farran. 2010. *Aiding Peace: A Multi-donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan 2005 – 2010*. ITAD Ltd.

Böhnke, J., J. Koehler, and C. Zürcher. 2010. *Assessing the Impact of Development Co-operation in North East Afghanistan 2005-2009*: *Final Report*. Evaluation Reports 049. Bonn: BMZ.

Cosgrave, J., B. Ramalingam, and T. Beck. 2009. *Real time evaluations of Humanitarian Action: An ALNAP Guide*. Pilot Version. ALNAP.

Foresti, M., C. Archer, T. O'Neil, and R. Longhurst. 2007. *A Comparative Study of Evaluation Policies and Practices in Development Agencies*. Agence Française de Développement (AFD) in collaboration with Overseas Development Institute (ODI). Paris: AFD. Available online: <http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/4343.pdf>

Independent Evaluation Group (IEG). 2013. *Mandate of the Director-General, Evaluation*. "About IEG." Available online: <http://ieg.worldbankgroup.org/content/dam/ieg/aboutIEG/dge_mandate_tor.pdf>

North Atlantic Treaty Organisation (NATO). 2009. *Multinational Experiment 5 Analysis Report*. Norfolk, VA: HQ Supreme Allied Command Transformation.

——— 2010. *NATO's Strategic Concept for the Defence and Security of the Members of the North Atlantic Treaty Organisation.* Adopted by Heads of State and Government at the NATO Summit in Lisbon 19-20 November 2010. Available online: <www.nato.int/lisbon2010/strategic-concept-2010-eng.pdf>

——— 2011a. *NATO Operations Assessment Handbook*. Interim Version 1.0. Norfolk, VA: HQ Supreme Allied Command Transformation.

——— 2011b. *Operations Assessment: A Case Study Based On Planning for Transition in Afghanistan*. RTO-TR-SAS-091. Paris: NATO Research and Technology Organisation.

Organisation for Economic Co-operation and Development (OECD). 2005. *The Paris Declaration on Aid Effectiveness*. Paris: OECD. Available online: < http://www.oecd.org/dac/effectiveness/34428351.pdf>

——— 2010. *Evaluation in Development Agencies.* Better Aid. Paris: OECD

——— 2011a. *Evaluating Development Co-operation: Summary of Key Norms and Standards*. 2nd Edition, OECD DAC Network on Development Evaluation. Paris: OECD. Available online: <www.oecd.org/dataoecd/12/56/41612905.pdf>

——— 2011b. *Supporting Statebuilding in Situations of Conflict and fragility – Policy Guidance.* DAC Guidelines and Reference Series. Paris: OECD.

——— 2012. *Guidance on Evaluating Peacebuilding Activities in Settings of Conflict and Fragility - Improving Learning for Results*. Paris: OECD. Available online: <http://bit.ly/T4qZrS>

OECD Development Assistance Committee (DAC). 2002. *Glossary of Key Terms in Evaluation and Results Based Management*. Reprint 2010. Paris: OECD. Available online: <http://www.oecd.org/development/peer-reviews/2754804.pdf>

——— 2008. *Is it ODA?* Factsheet November 2008. Available online: <http://www.oecd.org/dac/stats/34086975.pdf>

OECD DAC and Norwegian Agency for Development and Cooperation (NORAD). 2011. Chair's summary - Evaluating conflict prevention and peacebuilding activities: Lessons from the application phase. Workshop. Oslo. Available online: <http://www.oecd.org/development/evaluationofdevelopment-programmes/dcdndep/48092571.pdf>

Picciotto, R. 2002. *Development Cooperation and Performance Evaluation: The Monterrey Challenge*. Washington, DC: Operations Evaluation Department, World Bank.

——— 2003. International trends and development evaluation: The need for ideas. *American Journal of Evaluation 24*(2):227-234.

Schroden, J. 2013. Operations assessment at ISAF: Changing paradigms. In Williams, A., F. F. Farina, J. Bexfield, and J. de Nijs. 2013. *Innovations in Operations Assessement: Recent Developments in Measuring Results in Conflict Environment*. Norfolk, VA: NATO HQ Supreme Allied Commander Transformation.

United States, Joint Forces Command (US JFCOM). 2010. *Multinational Experiment 6 Analysis Report*. Suffolk, VA.

Vahamaki, J., M. Schmidt, and J. Molander. 2011. *Review: Results Based Management in Development Cooperation*. Riksbankens Jubileumsfond.

Vogel, I. 2012. *Review of the Use of 'Theory of Change' in International Development*. Report for the UK Department of International Development (DFID). Available online: <http://r4d.dfid.gov.uk/pdf/outputs/mis_spc/DFID_ToC_Review_VogelV7.pdf>

Weiss, C. H. 1979. The many meanings of research utilisation. *Public Administration Review* 39(5):426-431.

Williams, A. P., and J. C. Morris. 2009. The development of theory-driven evaluation in the military theory on the front line. *American Journal of Evaluation* 30(1):62-79.

The World Bank. 2005. *The LogFrame Handbook: A Logical Framework Approach to Project Cycle Management*. Washington, DC: The World Bank Group. Available online: <http://gametlibrary.worldbank.org/FILES/440_Logical%20Framework%20Handbook%20-%20World%20Bank.pdf>

——— 2011a. *Good Practice Note for Development Policy Lending: Results in Development Policy Lending*. Washington, DC: The World Bank Group. Available online: <http://siteresources.worldbank.org/PROJECTS/Resources/40940-1244732625424/GPNonResultsJanuary2011.pdf>

——— 2011b. *Self-Evaluation of the Independent Evaluation Group*. Washington, DC: The World Bank Group. Available online: <http://ieg.worldbankgroup.org/content/dam/ieg/aboutIEG/ieg_self_eval.pdf>

7

A Review of Joint Evaluation: Opportunities for NATO?[1]

Patricia Muller, Andrew Williams, Anne-Maree Ruddy, Marcey Moss

Abstract

This chapter is a companion piece to Chapter 6 and expands further on the potential for cooperation between civilian and military actors in the domain of operations assessment and evaluation, specifically in the practice of joint evaluation. While evaluations are typically arranged between a requesting customer and an evaluating body, the internationalization of evaluation and increasing drive for cost-efficiencies has pushed the emergence of joint evaluations involving more complex arrangements between multiple customers and evaluating bodies. The chapter provides the foundation needed to advance understanding of the current and future roles of joint evaluation and the potential for military participation. The chapter defines joint evaluation, provides context and background related to joint evaluations, provides an overview of key findings from existing research and literature, and makes recommendations for broadening the field by considering how evaluation and operations assessment are, or should be changing given the landscape of increasing collaborative governance and inter-organisational collaboration. Some practical suggestions for NATO operators are suggested, to initiate the practice of joint evaluations in the future.

Introduction

In the civilian domain, evaluations have traditionally been arranged as principal-agent relationships between a management, leadership, funding or accountability body, and the evaluating body. With the acceleration of the internationalization of evaluation and increasing drive for cost-efficiencies,

1. A Note from the Editors: This work is derived from a collaborative project between Indiana University and HQ SACT, conducted between April and November 2012. Indiana University was partially funded by HQ SACT (ACT PO Ref: 124000571.) An earlier version of this chapter was presented at the 10th European Evaluation Society Biennial Conference, 3–5 October, 2012, in Helsinki, Finland.

however, recent practice has seen the emergence of "joint evaluations" involving more complex arrangements between multiple principals and agents. While the overall number of joint evaluations in the field remains "rather limited" (Feinstein and Ingram 2003; Hageboeck 2009), the trend towards joint evaluations is evident from the steadily increasing numbers of joint evaluations of humanitarian actions and international development cooperation and assistance programs (Beck and Buchanan-Smith 2008; Brier 2005; OECD 2010), as well as the growing intent to use joint evaluations (OECD 2005b; Telford 2009) and the on-going strong international interest in creating an all-inclusive network that can promote joint evaluations (OECD 2012).

Currently, more than half of all member agencies of the Development Assistance Committee (DAC) of the OECD have evaluation policies that explicitly give direction with regards to joint evaluations, and approximately one quarter of DAC member evaluations each year are now joint evaluations (OECD 2010); and at least half of all evaluations by SIDA (Sweden), Danida (Denmark), BMZ (Germany), and Finland are joint evaluations (OECD 2010). In fact, U.S. Agency for International Development (USAID) asserts that one of the greatest shifts in evaluation in the last decade has been a greater focus on joint evaluation; a shift that they state is at least partially attributable to the 2005 Paris Declaration on Aid Effectiveness (Hageboeck 2009).

The purpose of this chapter is to provide the foundation needed to advance understanding of the current and future roles of joint evaluation and the potential for military participation. We define joint evaluation, provide context and background related to joint evaluations, provide an overview of key findings from existing research and literature, and make recommendations for broadening the field by considering how evaluation and operations assessment are, or should be changing given the landscape of increasing collaborative governance and inter-organisational collaboration. This chapter should be read as a companion to Chapter 6 in this volume, and is intended for expert level policy makers and evaluators in the development and military fields.

Defining Joint Evaluation

The OECD *Evaluation Glossary* defines joint evaluation as "an evaluation to which different donors and/or partners participate" (OECD 2002, 26); and for the purposes of a meta-evaluation of evaluations of humanitarian action, Beck

and Buchanan-Smith defined joint evaluations as "evaluations carried out by two or more agencies, evaluating the work of two or more agencies" (Beck and Buchanan-Smith 2008, 85). Beyond these very broad and generic definitions of joint evaluation, however, there is little consensus in the field about what constitutes a joint evaluation. Current working definitions used both by entities implementing joint evaluations as well as those studying joint evaluation vary greatly in terms of the types of partnerships included, the organisational level of interaction, the focus of the evaluation, the spectrum or scope of interaction or "jointness," and the nature and mechanisms of exchange.

Some definitions are more limited in scope given their focus on a specific "type" of joint evaluation. For example, Beck provides the following working definition specifically for joint *impact* evaluation: "joint judgment of the merit of the intended and unintended and negative and positive end results of interventions; and attribution of results to particular interventions; and/or program and generalizable lesson learning" (Beck 2009, 8). This particular definition, perhaps because of its more narrow focus on impact evaluation, includes "joint judgment" of the overall merit and worth of a program as a core component of the definition.

Yet other definitions include a wide range of alternative ways of joint working, ranging from joint data collection and/or exchanging assessments with external actors with each partner conducting its own analyses and separate report, to more truly collaborative endeavors in which each partner is mutually and equally responsible for the evaluation design, implementation and development of joint recommendations. For example, the United Nations Development Programme (UNDP) defines joint evaluations as collaborative efforts between UNDP and government and/or other national implementing partners where UNDP and partners are mutually and equally responsible for the evaluation exercise (UNDP 2009) whereas the DAC defines joint evaluation in a manner that recognises the various degrees of "jointness" depending on the extent to which the individual partners co-operate in the evaluation process (DAC 2005).

Given this diversity in focus and scope of joint evaluations, several typologies have been proposed that address differences in the degree and mode of "jointness." For example, the DAC differentiates between three different types of joint working: classic joint evaluation or multi-partner evaluation in

which participation is open to all stakeholder agencies and all partners contribute actively and on equal terms; qualified joint evaluation or multi-partner evaluation in which participation is open only to those who qualify through membership of a certain grouping or through active participation in the program being evaluated; and hybrid joint evaluation or multi-partner evaluation which includes a wide range of alternative ways of joint working (Brier 2005). This DAC typology primarily differentiates joint evaluations based on who is participating in the joint work.

Extending this typology to a consideration of humanitarian relief efforts, and based on a meta-evaluation of joint aid efforts in this area, Beck (2009) also developed a typology that focused on two categories: purpose and scope (i.e., institutional, sectoral or thematic, multisectoral related to a particular humanitarian crisis and usually bound geographically, and global); and how actors work together (i.e., partnership, like-minded agencies, hybrid multipartner, and system-wide). However, the current typologies for joint evaluation appear to be limited in their usefulness for describing the diversity of scope and focus in the current and potential future landscape of joint evaluations.

For the purposes of this chapter, the following loose definition of joint evaluation is adopted that includes a broad spectrum of "jointness" in order to ensure that our review and assessment of the current landscape is as comprehensive as possible: evaluations, in whole or in part, being carried out by more than one organisation, and/or evaluating the programs, initiatives or interventions being implemented by more than one organisation. The adoption of this definition for the purposes of this chapter does not suggest that such a broad definition should (or should not) be used in the field of evaluation more generally, but is simply used here as a framework for determining the scope of our review.

The review is also inclusive of joint evaluations conducted using other terminology, as long as the construct employed is similar to the general concept of joint evaluation as defined for the purposes of this chapter. Therefore, research and literature that uses the following terminology are also generally included: multi-agency evaluation, multi-partner evaluation, inter-agency evaluation, and sector-wide evaluation. However, methodology that is typically called participatory evaluation and/or collaborative evaluation (e.g., Cousins, Donohue, and Bloom 1996; Pollitt 1999; Borton, Buchanan-Smith, and Otto 2005) are

not considered within the scope of this chapter given that the distinguishing feature of this type of evaluation is evaluators collaborating in some fashion with program practitioners and/or stakeholders (non-evaluators) rather than collaboration between evaluators and/or between donors (funders); and joint evaluations vary in the degree to which they use highly participatory evaluation methods (Hageboeck 2009).

Context and Background for Joint Evaluations

Joint evaluation as a construct has evolved most discernibly within the development sector, initially pioneered by donor governments coming together under the auspices of the OECD DAC (Beck and Buchanan-Smith 2008). As early as 1991, the DAC promoted the use of joint donor evaluations "in order to improve understanding of each others' procedures and approaches and to reduce the administrative burden on the recipient" (OECD 1991, 8); and in 1996 the DAC encouraged donors to make joint evaluations a routine approach to conducting evaluations (OECD 1996).

This DAC emphasis on donor-level partnerships in defining joint evaluation (i.e., multi-donor joint evaluations) was broadened in later years to reflect the momentum in development cooperation towards a much broader and inclusive focus on partnerships, including joint evaluations undertaken with non-donor agencies (OECD 2006). This change was reflected in DAC's changing their standard terminology from "multi-donor evaluation" to "joint evaluation" as evidenced in its revision of the DAC series from *Effective Practices in Conducting a Multi-Donor Evaluation* (OECD 2000) to *Guidance for Managing Joint Evaluations* (OECD 2006).

The significant increase in joint evaluations within the development sector during the last decade is generally attributed to a new development paradigm or new "ways of doing business" in development cooperation that emphasize collaborative, multi-donor assistance efforts. The shift towards increased collaboration is often attributed, as least in part, to the Paris Declaration on Aid Effectiveness (OECD 2005b), as well as the Accra Agenda for Action (2008) the Busan Partnership for Effective Development Co-operation (2011), and the Millennium Development Goals Report (United Nations 2012).

Along with increases in the quantity of joint evaluations being conducted in the development sector during the last decade, the literature in the field

suggests that the concomitant trend towards national ownership has also resulted in an increased emphasis on the role and involvement of aid recipients and donor recipient countries in the joint evaluation process. However, there has been some concern expressed that the drive for national accountability in donor countries may negatively impact the trend towards joint evaluation as donors concentrate on their own contributions to the development processes in a particular partner country (OECD 2010; OECD 2012).

Although joint evaluations have a longer and more robust history in the development sector than in the humanitarian sector (Beck and Buchanan-Smith 2008), joint evaluation does have a strong and historic foothold within humanitarian aid as well. Significant groundwork for joint evaluation efforts has been laid during the last two decades with the multi-agency evaluations of international responses to humanitarian crises, including the Rwanda genocide and the Indian Ocean Tsunami. The Rwanda joint evaluation was the first comprehensive evaluation of emergency operations and represented an level of unprecedented international collaboration to learn lessons from the international response to Rwanda genocide (Dabelstein 1996; Borton 2001); and the landmark multi-agency, system-wide Rwanda evaluation continues to inform contemporary joint evaluation debates and practices (Beck and Buchanan-Smith 2008).

Although system-wide joint evaluations were also considered to evaluate the response to Hurricane Mitch in 1998, as well as the Kosovo crisis, neither of these situations resulted in a system-wide joint evaluation (Beck and Buchanan-Smith 2008); and the Rwanda joint evaluation remained the first and only system-wide evaluation until Indian Ocean Tsunami disaster. However, many other joint evaluations occurred within the humanitarian sector in those intervening years and beyond, most frequently smaller groups of donors coming together to evaluate humanitarian actions in a particular country, the performance of a group of agencies or a thematic issue (Beck and Buchanan-Smith 2008). In addition to donors, UN agencies and non-governmental organisations (NGOs) have also expressed recent and growing interest in joint evaluations of humanitarian actions (Telford 2009); but recipient governments, national NGOs and research institutions have been more reticent to become involved (Beck and Buchanan-Smith 2008).

Within the humanitarian sector, the literature reveals two particular approaches that have been most prominent: joint real-time evaluations and impact evaluation. Joint real-time evaluations, or inter-agency real time evaluation, are evaluations implemented concurrently with a humanitarian operation for the purposes of providing formative data for immediate use in supporting decision-making at the field level. A groundbreaking joint evaluation using a real time approach in the humanitarian sector was the *Inter-Agency Real Time Evaluation of the Humanitarian Response to the Darfur Crisis* (Broughton, Maguire, and David-Toweh 2006) commissioned by the UN.

This joint evaluation was heralded as "the first attempt to comprehensively evaluate an ongoing crisis across all sectors and function" (Broughton, Maguire, and David-Toweh 2006, 1), and is indicative of the increasing emphasis during the last decade on real-time evaluations (RTEs) using a joint evaluation approach. The trend towards increasing numbers of joint real-time evaluations is evident from the Inter Agency Standing Committee's (IASC) commitment to a pilot inter-agency RTE program through which three joint real time evaluations were implemented between 2007 and 2008 (Telford 2009); and European Commission's Humanitarian Aid Office (ECHO)'s implementation of a series of joint real-time evaluations with World Health Organisation (WHO) and other donors (Beck and Buchanan-Smith 2008).

The trend towards promoting joint impact evaluations of humanitarian actions aligns with the current humanitarian agenda emphasizing impact, accountability, and evidence-based policy. Joint humanitarian impact evaluation has been discussed by numerous agencies within the past decade, including Emergency Capacity Building (ECB), Office for the Coordination of Humanitarian Affairs (OCHA) and United Nations Children's Fund (UNICEF), and at the interagency level several initiatives have focused on increasing collaboration through joint evaluations of humanitarian actions. The saliency of joint impact evaluations is evident from the OCHA commissioned paper on *Evaluability Assessment for Impact Evaluation of the Humanitarian System at the Country Level and a paper on Joint Humanitarian Impact Evaluation* for discussion at the 25th Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP) meeting in London in 2009 (Beck 2009).

Beyond the development and humanitarian sectors, however, there appears to be sparse evidence of joint evaluations conducted in other fields.

A comprehensive and exhaustive review of the research and literature, including published reports, peer-reviewed journal articles, and on-line data and documents across multiple fields of study and using diverse terminology (e.g., inter-agency, multi-agency, multi-donor, etc.) suggest that joint evaluation as a construct is not common outside of international development and humanitarian sectors. What is less clear is the extent to which the practices and evaluation approaches used in other fields, but perhaps discussed with very different constructs and terminology, may contain similar characteristics and attributes of joint evaluation in the development and humanitarian aid sectors.

For example, large national agencies in the United States such as the U.S. Department of Education and the National Science Foundation frequently fund external third-party entities to conduct large-scale meta-evaluations of specific grant funding streams. The individual grants are implemented by independent agencies (e.g., state education agency, local education agency, non-profit organisation), with diverse programming structures, activities and goals. At the State level, many governments have implemented outcome monitoring programs for various mental health and social services that involve large numbers of diverse agencies participating in the evaluation process (Wall et al. 2005; Mason et al. 1998).

In the current international mission in Afghanistan, military commands with various regional or domain responsibilities regularly collaborate to produce quarterly outcome evaluations. Although the terminology of joint evaluation is not used in these circumstances, there are similarities in the structure and process of these types of large-scale, system-wide evaluations that warrant further discussion to determine the extent to which they should be included in theoretical frameworks of joint evaluation.

Key Findings from Existing Research and Literature

In a previous study by the authors of this chapter (Muller et al. 2012) a comprehensive literature review of the field of joint evaluation was conducted that considered: meta-evaluations of joint evaluations; lessons learned and best practices; methodological approaches; DAC Evaluation Network members' perceptions of joint evaluations; how-to-guides; and theoretical studies published. In this section, we highlight some of the main aspects of that literature review by discussing key findings related to the frequency of joint evaluations,

the quality of joint evaluations, and the benefits and challenges of conducting joint evaluations. Details related to lessons learned and best practices are not discussed given that comprehensive "how-to" guides are readily accessible from OECD (2000, 2006) and ECB (2011); and other findings related to best practices and lessons learned are anecdotal in nature, and not based on systematic and rigorous research.

As evident from the literature review, there is limited empirical research related to the impact or effectiveness of joint evaluations, nor are there any substantial attempts to develop theoretical frameworks that seek to describe and structure the variables of importance, explain the reasons for joint evaluation, and understand how the collaborative process works in practice. The studies that do exist often have significant methodological limitations such as focusing on self-reported perceptions known to have limited reliability and validity; and describing best practices or lessons learned that are based primarily on anecdotal data and dialogue in the field rather than a more systematic and rigorous methodological approach. The need for more systematic and rigorous approaches to studying joint evaluation is clear.

Frequency of Joint Evaluations

In terms of the frequency of joint evaluations, beyond anecdotal data and perceptions of those in the field, the only empirical data that currently exist are from the development sector: an OECD (2010) survey specifically addressing the frequency of joint evaluation within the DAC-OECD member population; and a publicly accessible inventory database of on-going and planned evaluations that are voluntarily submitted by members of the DAC Network on Development Evaluation. While these data provide some insight into the frequency of joint evaluations within the development sector, they are limited in their usefulness due to concerns about the representativeness of the samples. Similar analyses or data do not currently exist in the humanitarian sector.

In terms of the development sector, OECD's (2010) study included DAC-OECD member surveys. Based on analyses of these survey results, OECD (2010) found that joint work makes up for 24% of the 696 evaluations carried out by member departments on an average year, with 15% of these joint evaluations conducted with partner countries, 7% conducted with other agencies/donors, and 2% unspecified. The data also indicate that joint evaluations are more

frequently conducted by bilateral members than multilateral members (29% versus 7% of all evaluations are jointly conducted); although OECD notes that this disparity is likely skewed by the large number of joint evaluations reported by one member, Japan International Cooperation Agency (JICA). However, Andersen and Broegaard (2012) note that these percentages may be inflated due to donors' use of varying definitions of "joint evaluation" that could result in over-reporting. In addition, the self-report nature of the data and lack of anonymity or confidentiality of the members' survey responses may also result in over-reporting, particularly given the strong paradigm emphasizing collaboration in the development sector.

While acknowledging that similar data do not exist for earlier periods, OECD concludes, based primarily on perceptual data (i.e., based on recollections of the numbers of joint evaluations discussed at earlier meetings of the network or published on member websites), that these numbers represent a significant increase from ten or fifteen years ago when the concept of joint evaluation was just emerging in the development sector. However, as noted by OECD, there is currently no empirical data to substantiate this claim.

In addition to the OECD (2010) survey results, we reviewed and analysed the DAC-Development Cooperation Directorate's inventory database of on-going and planned evaluations voluntarily submitted by members of the DAC Network on Development Evaluation.[2] Ten percent of the 517 submitted evaluations indicated that they were joint evaluations, with another 4% potentially becoming a joint evaluation (i.e., represented a planned evaluation that indicated interest in a joint evaluation). Of the joint evaluations, most fell into the topic/sectors of budget support (19%), country-level evaluation (21%), and multilateral aid (15%). Country agencies and multilateral agencies that indicated at least three joint evaluations included: European Union (EU) Commission, Sweden, Denmark, France, the United Kingdom, Finland, Belgium, African Development Bank, and Austria.

It should be noted that this list is not exhaustive, as several agencies were simply listed as "joint" and in many cases only the lead country was listed (although it is likely that in some cases the only other agency would be the

2. See, http://www.oecd.org/dac/evaluationofdevelopmentprogrammes/dcdndep/whatisbeingevaluated-sharingmemberplans.htm. Last updated May 2102.

partner, non-donor country). Comparing this list to OECD's (2010) study, JICA is notably absent from this list. The self-report data may represent a severe limitation on the ability to generalize from these statistics. Ambiguity in the field titles may also be problematic. For example, the "joint with country" field is only reported in six of the cases, while "multi agency" is only reported for 17 of the cases. Furthermore, the data is limited in the sense that responses may reflect intentions as opposed to actions.

Quality of Joint Evaluations

The only systematic analyses of the quality of joint evaluations that appears in the research and literature is Beck and Buchanan-Smith's (2008) assessment of a sample of joint evaluations against eight ALNAP Quality Pro Forma areas: terms of reference, appropriateness of the overall evaluation methods, consultation with and participation by primary stakeholder, the use and adherence to international standards and guidelines, OECD DAC criteria (aggregate), gender equality, protection and advocacy. In six of these areas of the Pro Forma, comparative, longitudinal analyses were used to compare Pro Forma assessment data from the joint evaluations with the results from pervious ALNAP meta-evaluations using the same Pro Forma and protocol; and results were triangulated with agency and evaluator interviews.

Based on these analyses, joint evaluations show higher quality than single-agency evaluations (Beck and Buchanan-Smith 2008). For the six Pro Forma areas included in the analyses, the average percentage of joint evaluations rated good was 27% and satisfactory was 40%, as compared to 9% of single-agency evaluations rated good and 34% rated satisfactory. More specifically, the findings indicate that for joint-evaluations the terms of reference are generally clearer and more usable; consultation with aid recipients and local populations is stronger; adherence to internal standards and guidelines are stronger; and OECD DAC criteria are more rigorously used. In addition, joint evaluations generally had more rigorous methodologies than single-agency evaluations; however, this finding was inconsistent given significant gaps in some of the joint evaluations reviewed, particularly pertaining to attention to gender equality, protection and advocacy.

As noted previously, there are some limitations to this study, particularly in terms of the voluntary nature of report submission to the ERD database

from which the sample was drawn. However, concerns regarding the voluntary nature of the database are minimized by the fact that this limitation applies to both single-agency evaluations and joint evaluations. In other words, potential concerns about the representativeness of the samples given the nature of the database (e.g., are voluntarily submitted reports may be of higher quality than those not submitted) are lessened because in this instance any bias in the submitted reports is likely to be similar for both joint evaluations and single-agency evaluations.

Benefits and Challenges of Conducting Joint Evaluations

There is a general recognition of the major benefits of joint evaluations; and the research and literature also contains numerous descriptions of the challenges or costs of conducting joint evaluations. SADEV's (2008) study represents one of the more systematic and empirical approaches to identifying perceived challenges of joint evaluation, producing findings from a comprehensive survey distributed to all members of the DAC Evaluation Network, supplemented with telephone interviews with individuals representing development programs in twenty two different countries, as well as seven different international organisations and development banks. Survey respondents identified the following challenges: joint evaluations are not well aligned with accountability purposes (more appropriate for learning); joint evaluations are generally more costly than individual evaluations; joint evaluation planning processes are time-consuming and require lengthy negotiations to reach agreement on scope, methodology, evaluation questions, etc.; and the timing of needed formative feedback for decision-making becomes more challenging the more parties involved.

Other discussions of the benefits and challenges in the literature are generally more anecdotal, or based on less formal or systematic evidence. However, OECD (2006) does discuss benefits and challenges based on a comprehensive review of extant joint evaluations; findings and recommendations from Breier's (2005) study; the outcomes of a 2005 Nairobi workshop (OECD 2005a), *Joint Evaluations: Challenging the Conventional Wisdom – the View from Developing Country Partners*; and feedback and input from DAC-OECD Evaluation Network members and partners. OECD discusses the following potential benefits of joint evaluation: mutual capacity development; harmonisation and reduced transaction costs; increased participation of developing countries and alignment

of evaluations with national needs; increased objectivity, transparency and legitimacy of the evaluation and its impact; and an ability to address a broader scope of evaluation questions.

Challenges noted include subjects that are especially suited to joint evaluation being more difficult to evaluate than single agency projects, and the complexity and costs of the processes involved in coordinating joint work. Breier (2005) also addresses benefits and challenges of joint evaluation. Based on the outcomes of extensive discussions at the 2005 Nairobi workshop, Breier provides a long list of both opportunities and benefits of joint evaluation, accompanied by a parallel long list of the problems and challenges associated with each of the identified opportunities and benefits.

Conclusions and Way Ahead

The evidence gathered for this chapter indicates that joint evaluation is a major approach in the discipline, especially in the domain of international development and humanitarian aid. While initial efforts have been made to study the wider context of joint evaluation in the form of meta-evaluations, lessons-learned, and practice guidelines, more systematic theoretically-informed academic research is lacking. This final section develops a research agenda to address some of the deficiencies highlighted, and makes recommendations for broadening the field by considering how evaluation is, or should be changing given the landscape of increasing collaborative governance and inter-organisational collaboration. First, in terms of the need for more systematic theoretically-informed academic research, this chapter identifies several key gaps in the research and literature, including the following:

Studies comparing joint evaluation and single organisation evaluation

Although many joint evaluations have been conducted and several meta-evaluations have looked at quality, there have been few attempts to meta-evaluate joint evaluations compared to single organisation evaluations. Further research on joint evaluations is needed to assess against evaluative frameworks such as Scriven's (2011), which considers the validity, clarity, credibility and propriety of evaluations. Similarly, studies that more systematically examine the impact and effectiveness of joint evaluations as compared to single organisation evaluations would benefit the field. Another important aspect is to consider

how joint approaches affect the utilisation of evaluation results. Sophisticated utilisation models already existing in the literature could be employed (e.g. Balthasar 2009) as a starting point, with some modifications required to take into account the inter-organisational aspects at the point of result utilisation.

Cost-benefit analyses

Cost-benefit studies are required to develop empirical understanding about the advantages and disadvantages of each approach and relative financial, resource, political and "transaction" costs and benefits, assessed against the quality of evidence and its utilisation in the joint and non-joint cases. Breier (2005) provides an in-depth discussion of the calculation of both direct and indirect costs of joint evaluation; and abundant economic evaluations and cost-benefit analyses already exists that could be used to begin to more empirically examine the differential costs and benefits of joint evaluations and single organisation evaluations.

Focused theoretical development and analyses

Our understanding of joint evaluation would benefit from focused theoretical development and analyses. A key starting point is to understand the various combinations of customers, donors, project organisations, evaluating organisations, evaluators, and program beneficiaries that constitute joint evaluation situations. A basic typology expanding on the simplistic scales of "jointness" would be a reasonable starting point.

Future Directions in the Field

In addition to these more targeted areas identified for future research, this chapter also highlights the need to continue to broaden the field by considering how evaluation is, or should be changing given the landscape of increasing collaborative governance and inter-organisational collaboration. Fundamentally, there is a broader link that evaluation theorists can make to the wider literature on multi-organisational collaboration in governance. Thus far, the majority of evaluation literature has focused on issues surrounding the evaluation of a single program either by the organisation that implemented the program, or by external independent or regulatory bodies (Conley and Margaret 2003). With the rise of network government and collaborative policy

implementation, evaluation literature has started to address evaluation of collaborative partnerships; however, the focus has been on evaluating the partnership itself rather than considering how evaluation might change as a result of the partnership (Cross et al. 2009; Woodland and Hutton 2012). Currently, there are no frameworks or theories to guide the practice of evaluation conducted amongst multiple organisations.

There are two distinct cases to be considered in future research. First, to study how multiple organisations collaboratively implementing a single program can conduct a collaborative evaluation of that common program, rather than their own independent activities. Many examples of such programs can be found in the literature (e.g., Ansel and Gash 2007). Second, to study how multiple organisations implementing separate, but related programs in a common sector or "tradespace" (Meharg 2009) can conduct a collaborative evaluation of the overall sector. In this case, there are many examples of collaboration in environmental policy areas such as watershed management programs (Conley and Margaret 2003; Eikenberry, Arroyave, and Cooper 2007; Imperial 2005; McNamara 2008), inter-organisational collaboration in disaster response (Eikenberry, Arroyave, and Cooper 2007; Simo and Bies 2007), and evaluations of international development programs (Picciotto 2003, 2005). Further investigation will likely reveal more permutations of organisational arrangements possible.

In the context of analyzing multi-organisational collaboration, there are a wide list of questions derived from key focal points identified in the literature on organisational collaboration and evaluation. Huxham and Vangen (2005) identified cores themes that are important in collaboration between organisations: organisational learning, membership structures, working processes, resources, common aims, communication and language, compromise, commitment and determination, identify, power, culture, trust, leadership, accountability, social capital, equality, and risk. They identified that all of these aspects are important, to varying extents, in the required antecedents, processes and outcomes of collaboration. Similarly, based on an extensive literature review, Mattessich, Murray-Close and Monsay (2001) identified six clusters of variables that are important to consider: political and social environment, membership characteristics, process and structure, communication, purpose, and resources.

Evaluation scholars have identified many factors—too numerous to cover in this space—that are important to the planning, implementation, and validity of evaluation in organisations. Broadly, evaluation literature identifies the following aspects: methodology, techniques, reliability and validity and their qualitative analogues, policy, uses of evaluation in organisations, environmental factors, goals of evaluation, resources, and the structure of evaluation organisations. Using these above characteristics as a guide, the following list of initial research questions is suggested as a starting point (note, the term "inter-organisational" refers both to interaction between evaluating organisations, and interactions between sponsoring, donor, or customer organisations):

1. What are the factors in evaluation that are relevant to joint evaluation?
2. What are the overall benefits of joint evaluation?
3. What are the incentives for conducting joint evaluation?
4. What are the antecedents for joint evaluation?
5. What are the organisational challenges in conducting joint evaluation?
6. How do environmental (situational) factors affect joint evaluation?
7. What are the inter-organisational processes of joint evaluation?
8. How are organisations structured in joint evaluations?
9. What methods are important in joint evaluation?
10. How are the uses of evaluation affected by collaboration?

Future Directions for NATO

This chapter demonstrates the importance of joint evaluations in the fields of international development and humanitarian aid. Based on recent increasing use and positive perceptions, we foresee continuing significance of joint evaluations. Additionally, the trend towards collaborative governance in the development sectors, and comprehensive approaches in the military and NATO spheres, makes joint evaluations an increasingly important and essential way to measure program effectiveness with credibility, and to increase capacities and relationships between collaborative members. This final section considers the implications of this discussion for NATO.

As captured in the 2010 Strategic Concept NATO has increasingly emphasised the importance of working cooperatively with other actors. Given that success of an overall intervention is interdependent between the various

actors present, NATO recognised that assessment of progress is also interdependent (NATO 2010). Furthermore, there is increasing calls for NATO to reduce overlaps and seek efficiencies through cooperation. This is critical in operational theatres such as Afghanistan: the draw-down of Western presence will reduce the ability of military forces to conduct operations assessment, and NATO may increasingly rely on government aid, development and state agencies, and appropriate international and nongovernmental organisations, to fill data collection gaps.

In the area of operations assessment, few mechanisms are in place to achieve these goals. The ideas about joint evaluation presented in this chapter present a fruitful opportunity for NATO—and other military organisations and alliances—to consider the steps necessary to operationalise the idea of cooperative assessment, as part of a broader comprehensive approach. As a way to start the debate, we present some practical suggestions.

Stakeholder analysis

First, the civilian organisations relevant to assessment must be identified, and different types and circumstances of possible partnerships need to be defined. The challenges of civil-military cooperation are well known and understood (Ankersen 2008), and any interaction needs to be based on clear mutual benefits. The particular level of organisation at which interactions could be conducted depends strongly on the type of evaluation and its purpose. For example, at the theatre level and below, the military emphasises measurement of inputs and outputs, with some consideration of impacts in security domains. This is relevant to the decentralised evaluations for "results-based management" conducted by project implementation units in government agencies. From the operational level upwards, to strategic military (SHAPE) and political-military levels (NATO Headquarters), the thematic, sector, or country-wide evaluations conducted by central evaluation units are of relevance.

Building mechanisms of cooperation

Second, detailed mechanisms that specify how cooperation can be achieved need consideration. Much research and experience demonstrates that inter-organisational cooperation is facilitated by the existence of previous relationships, and mutual trust and respect between cooperating parties (Ansel and Gash

2007). NATO operations assessment personnel should seek to develop their networks and long-term partnerships with civilian evaluation units. Examples include: attendance at civilian evaluation conferences; participation in network organisations such as the OECD or the Interagency Standing Committee; bilateral meetings with individual nations; and extending invitations of civilian evaluation personnel to NATO projects, experiments, exercises, and meetings on Operations Assessment.

Depending on the level of interaction and the type of evaluation, formal agreements (often called "Memorandum of Understanding" or "Memorandum of Agreement") could be established to create a framework for interaction on Operations Assessment and facilitate partnerships during missions. While the design of agreements is entirely contingent on circumstances, they might specify roles and responsibilities, management structures, timescales, security protocols, and legal issues, or delegate certain responsibilities or authority to field personnel. For those partnerships where a high level of interaction is possible, joint mission planning is important to improve the conduct and quality of evaluations. In certain cases, planning and evaluation representatives of national aid and development agencies may be able to jointly plan missions with NATO and military personnel. Where joint mission or programme planning is not possible, evaluation staff may be able to plan directly with operations assessment personnel.

Depending on the type of partnership and type of evaluation, "burden sharing" options could be considered, in the context of a partnership agreement: temporary loan of staff, or expanding survey contracts to cover use by other organisations; pooling and sharing of costs between evaluation partners; coordinating geographic areas of responsibility for evaluations; dividing up surveying of evaluation subjects (i.e. villages, individuals) between partners; dividing tasks between organisations (e.g. data collection, data analysis).

These two concepts, stakeholder analysis and building mechanisms of cooperation, represent two critical steps necessary for NATO to establish a foundational base from which to operationalise the idea of cooperative assessment. Given the increasing use of joint evaluation in the fields of international development and humanitarian aid, and the potential benefits of cooperative assessment for reducing overlaps and increasing NATO efficiencies, the ideas presented in this chapter present a unique opportunity for NATO—and other

military organisations and alliances—to begin to implement joint evaluation as part of a more comprehensive approach that capitalizes on the increased interdependence of various actors. Beginning the process now to establish a foundational base for cooperative assessment will help to position NATO to take advantage of the fruitful opportunities for joint evaluation that are expected to continue to grow for NATO and other military organisations and alliances.

References

Ankersen C. ed. 2008. *Civil-military cooperation in post-conflict operations: emerging theory and practice.* Abingdon, Oxford: Routledge.

Accra Agenda for Action. (2008). 3rd High Level Forum on Aid Effectiveness. Accra: Ghana. Available online: <http://www.oecd.org/development/aideffectiveness/41202012.pdf>

Andersen, O. W., and E. Broegaard. 2012. The political economy of joint-donor evaluations. *Evaluation* 18(1):47-59.

Ansel, C., and A. Gash. 2007. Collaborative governance in theory and in practice. *Journal of Public Administration, Research and Theory* 18(4):543-571.

Balthasar, A. 2009. Institutional design and utilisation of evaluation: A contribution to a theory of evaluation influence based on Swiss experience. *Evaluation Review* 33(3):226-256.

Beck, T. 2009. *Joint Humanitarian Impact Evaluation: Options Paper*. Presented at 25th ALNAP Meeting, London.

Beck, T., and M. Buchanan-Smith. 2008. Joint evaluations coming of age? The quality and future scope of joint evaluations. In Herson, M., J. Mitchell, and B. Ramalingam. eds. *ALNAP Seventh Review of Humanitarian Action*. London: Overseas Development Institute.

Borton, J. 2001. Doing study 3 of the joint evaluation of emergency assistance to Rwanda: The team leader's perspective. In A. Wood, R. Apthorpe, and J. Borton. eds. *Evaluating International Humanitarian Action: Reflections from Practitioners (74-99)*. London, New York: Zed Books.

Borton, J., M. Buchanan-Smith, and R. Otto. 2005. *Learning from Evaluations of Support to Internally Displaced Persons: Synthesis Report of a Joint Evaluation Programme*. Stockholm: Sida.

Breier, H. 2005. *Joint Evaluations: Recent Experiences, Lessons Learned, and Options for the Future*. Paris: OECD.

Broughton, B., S. Maguire, and K. David-Toweh. 2006. *Inter-agency Real-time Evaluation of the Humanitarian Response to the Darfur Crisis*. UN Office for the Coordination of Humanitarian Affairs (OCHA). Available online:

<http://ochanet.unocha.org/p/Documents/RTE_DARFUR_Final_Report_14_Mar.pdf>

Busan Partnership for Effective Development Co-operation. 2011. Fourth high level forum on aid effectiveness. Busan, Republic of Korea. Available online: <http://www.oecd.org/dac/aideffectiveness/49650173.pdf>

Conley, A., and A. M. Margaret. 2003. Evaluating collaborative natural resource management. *Society & Natural Resources* 16(5):371-386.

Cousins, J. B., J. J. Donohue, and G. A. Bloom. 1996. Collaborative evaluation in North America: Evaluators' self-reported opinions, practices, and consequences. *American Journal of Evaluation* 17:207-226.

Cross, J. E., E. Dickmann, R. Newman-Gonchar, and J. M. Fagan. 2009. Using mixed-method design and network analysis to measure development of interagency collaboration. *American Journal of Evaluation* 30(3):310-329.

Dabelstein, N. 1996. Evaluating the international humanitarian system: Rationale, process and management of the joint evaluation of the international response to the Rwanda genocide. *Disasters* 20(4):287-294.

Development Assistance Committee (DAC). 2005. *Joint Evaluations: Recent Experiences, Lessons Learned, and Options for the Future*. Working Paper. Paris: OECD.

Emergency Capacity Building Project (ECB). 2011. *What We Know About Joint Evaluations of Humanitarian Evaluations: Learning from NGO Experiences*. Available online: <http://www.ecbproject.org/resources/library/42-what-we-know-about-joint-evaluations-of-humanitarian-action-learning-from-ngo-experiences>

Eikenberry, A. M., V. Arroyave, and T. Cooper. 2007. Administrative failure and the international NGO response to hurricane Katrina. *Public Administration Review* 67:160-170.

Feinstein, O., G. K. Ingram. 2003. *Lessons Learned from World Bank Experiences in Joint Evaluation*. DAC Working Party on Aid Evaluation. Agenda item 4 ii.

Hageboeck, M. 2009. *Trends in International Development Evaluation Theory, Policy and Practices*. Publication produced for the United States Agency for International Development. Washington, DC: USAID.

Huxham, C., and S. Vangen. 2005. *Managing to Collaboration: The Theory and Practice of Collaborative Advantage*. Abingdon, England: Routledge.

Imperial, M. 2005. Using collaboration as a governance strategy: Lessons from six watershed management programs. *Administration and Society* 37(3):281-320.

Mason, M., S. A. Beck, P. Meadowcroft, and E. S. Kiely. 1998. Multiagency outcome evaluation of children's services: a case study. *The Journal of Behavioral Health Services and Research* 25(2):163-176.

Mattessich, P. W., M. Murray-Close, and B. R. Monsay. 2001. *Collaboration: What Makes It Work*. 2nd ed. St. Paul, MN: Fieldstone Alliance.

McNamara, M. W. 2008. Exploring interactions during multiorganisational policy implementation: A case study of the Virginia Coastal Zone Management program. Doctoral Dissertation in Public Administration and Urban Policy. Norfolk, VA: Old Dominion University. Available online: <http://proquest.umi.com/pqdweb?did=1631167961&Fmt=7&clientId=3505&RQT=309&VName=PQD>

Meharg, S. J. 2009. *Measuring What Matters: In Peace Operations & Crisis Management*. Ottowa, Canada: School of Policy Studies, Queen's University.

Muller, P., A. Williams, A. Ruddy, and M. Moss. 2012. Joint evaluation. An unpublished study produced under contract for NATO Headquarters, Supreme Allied Commander Transformation.

North Atlantic Treaty Organisation (NATO). 2010. Strategic Concept for the Defence and Security of the Members of the North Atlantic Treaty Organisation.

Organisation for Economic Co-operation and Development (OECD). 1991. *Principles for Evaluation of Development Assistance*. Paris: Development Assistance Committee. Available online: <http://www.oecd.org/derec/dacnetwork/35343400.pdf>

——— 1996. *Shaping the 21st Century: The Contribution of Development Co-operation*. Available online: <http://www.oecd.org/dac/2508761.pdf>

——— 2000. *Effective Practices in Conducting a Multi-donor Evaluation*. Available online: <http://www.oecd.org/development/evaluationofdevelopment-programmes/2667318.pdf>

——— 2002. *Glossary of Key Terms in Evaluation and Results Based Management*. Available online: <http://www.oecd.org/dac/evaluationnetwork> on 11 September 2012.

——— 2005a. Workshop on joint evaluations: Challenging the conventional wisdom – the view from developing partners. Nairobi, 20-21 April 2005. Workshop Report. Paris: Development Assistance Committee.

——— 2005b. *Paris Declaration on Aid Effectiveness*. Available online: <http://www.oecd.org/development/evaluationofdevelopmentprogrammes/dcdn-dep/38245246.pdf>

——— 2006. *Guidance for Managing Joint Evaluations*. Available online: <http://www.oecd.org/derec/dacnetwork/37656526.pdf>

——— 2010. *Evaluation in Development Agencies, Better Aid*. Available online: <http://dx.doi.org/10.1787/9789264094857-en>

——— 2012. Workshop on lessons learned from international joint evaluations. Available online: <http://www.oecd.org/dac/evaluationofdevelopmentprogrammes/dcdn-dep/49892763.pdf>

Picciotto, R. 2003. International trends and development evaluation: The need for ideas. *American Journal of Evaluation* 24(2):227-234.

——— 2005. The evaluation of policy coherence for development. *Evaluation* 11(3):311-330.

Pollitt, C. 1999. Stunted by stakeholders? Limits to collaborative evaluation. *Public Policy and Administration* 14(2):77-90.

Scriven, M. 2011. Evaluating evaluations: A meta-evaluation checklist. Available online: <http://michaelscriven.info/papersandpublications.html>

Simo, G., and A. L. Bies. 2007. The role of non-profits in disaster response: An expanded model of cross-sector collaboration. *Public Administration Review* 67:126-142.

Swedish Agency for Development Evaluation (SADEV). 2008. *Challenges, Opportunities, and Approaches for Increasing Joint Donor Programming of*

Evaluations. Available online: <http://www.oecd.org/dac/evaluationofdevelopmentprogrammes/dcdndep/41615876.pdf>

Telford, J. 2009. *Review of Joint Evaluations and the Future of Inter-agency Evaluations.* Final Report. United Nations Office for the Coordination of Humanitarian Affairs (OCHA). Available online: <http://ochanet.unocha.org/p/Documents/Review_of_joint_evaluations_and_future_of_IA_evaluation%5B1%5D.pdf>

United Nations. 2012. *The Millennium Development Goals Report 2012*. New York, NY: United Nations.

United Nations Development Programme (UNDP). 2009. *Handbook on Planning, Monitoring and Evaluating for Development Results*. New York, NY: UNDP.

Wall, J. R., M. Busch, S. M. Koch, G. Alexander, H. Minnich, and S. Jackson-Walker. 2005. Accountability in child welfare services: Developing a statewide outcome evaluation program. *Psychological Services* 2(1):39.

Woodland, R. H., and M. S. Hutton. 2012. Evaluating organisational collaborations: Suggested entry points and strategies. *American Journal of Evaluation* 33(3):366-383.

8

Engaging Local Actors in Operations Assessment[1]

Emery Brusset, Andre Kahlmeyer

Abstract

This chapter discusses the benefits and risks associated with including local actors in NATO operations assessments. It also suggests processes NATO could use to integrate local actors into the assessments. This would require NATO to 1) engage, listen to, and generate mutual respect and trust; 2) build personal relationships; 3) plan with the local actors, ensure local ownership; 4) deliver on commitments; 5) be open and transparent, provide feedback and learning; 6) exercise continuity, patience, and bear in mind what happens after handover and exit. Engaging local actors in operations assessments could help NATO better understand the civilian effects of their operations (a key dimension of population-centric warfare), including objectives such as perceptions of development and security, improvement of governance and the rule of law. It may also involve the development of mechanisms to ensure local ownership of assessment processes and consideration of new 'civilian effects' teams that can focus on non-target related objectives and operate relatively independently from the military. There are also risks both for the local actors that could become targets of the insurgency and for NATO in maintaining the security of its troops and information.

Introduction

NATO has become a lot more involved in asymmetric theatres of operations in the past two decades, which challenge conventional operational procedures as well as the relevance of NATO's doctrine in culturally complex environments,

1. A Note from the Editors: This work is derived from a HQ SACT contract awarded to Channel Research in June 2012 titled: "Initial Research and Analysis on Issues Related to Local Actors in NATO Operations" (RFP-ACT-SACT-12-03). The creation of this article was not part of the contract and was conducted without compensation by the authors.

where state institutions are often too weak for NATO to rely on them. The concept of "enemy" has become more fluid as NATO intervenes in environments which are geographically, socially and politically fragmented, and often faces a myriad of adversaries. In these kinds of operations the assessment lens needs to consider the impact of military operations on local actors in a multi-perspective manner where "local actors" are defined in NATO documentation as elected or non-elected civil servants at the regional or local government level, local authorities in general, village elders, tribal, cultural and religious leaders, and host nation civil society representatives (NATO 2012a). This excludes military personnel and international aid and development personnel—i.e., those actors associated with an international intervention.

In general, stronger involvement of local actors (both governmental and non-governmental) can increase the knowledge necessary for any operation, especially in the socio-economic field. Local knowledge about structures, population composition, demographics, and especially about the perception of the presence and action of foreign soldiers is key information for military operations. The involvement of local actors can increase confidence, trust and contribute to relationship building. By engaging in a direct dialogue with local actors NATO will be able to better identify influential individuals and understand how conflict drivers are perceived locally.

Operations assessments are used to measure the progress of military operations (both military/security and development/governance), and the subsequent development of conclusions and recommendations that support decision-making (NATO 2012b). Assessments are both a tool and a process. As a tool, assessments collect information that helps commanders adjust campaign objectives and verify their relevance and likelihood of success. As a process, assessments can promote relationship building, a better understanding by NATO of the local populations, and, ideally, a better understanding of NATO by local populations. Assessments are critical to managing objectives and plans (both military and development), and ultimately reinforcing the security of personnel, as well as that of the local population in the theatre of operations.

In hybrid environments (e.g., Afghanistan) a special focus should be given to the perceptions of the affected population and their opinions about NATO's presence. Tools are needed to collect data to establish baseline perceptions and to monitor progress against plans, especially for non-security goals. The

assessment of perceptions would include political and socio-economic information about a given area, as well as the perception of the targeted population about the progress being made, both in security and development terms. Often this will require statistical and survey methods, as well as a team of well-trained evaluators that will interact to collect qualitative data directly with the local population. In chapter four of this book, Katherine Banko (2013) describes how to form these teams and conduct the surveys.

This chapter focuses on how the greater involvement of local actors in operational and tactical level assessment processes and their use in planning can increase the chances of success for a NATO mission. Companies that have grasped the importance of actively developing and sustaining relationships with affected communities throughout the life of their projects, and not simply during the initial feasibility phase, are reaping the benefits of improved risk management and better final outcomes. As approaches evolve from a short-term means of meeting regulatory and lender requirements to a longer-term focus on relationship-building and risk mitigation, new business opportunities were identified benefiting everyone. Dialogue around the consequences of operations, positive or negative, allowed a new range of information to enter the decision processes. We adapt this industry approach to the context of NATO and other militaries' needs to engage with local actors in operations assessment.

While this chapter argues that the increased and broader involvement of local actors will contribute to improving assessment processes, one should also keep in mind and respect the huge imbalance between the resources available for military operations as compared to that of local actors. Local engagement in assessment processes will put additional burden in terms of time and commitment on local actors and possibly raise expectations that cannot be met later. Inclusion for the sake of inclusion—"ticking the box"—should be avoided under all circumstances; however, if designed and implemented in a meaningful way the benefits will outweigh the invested resources, both locally and for the military.

Literature on Local Actor Engagement

Current NATO guidance on operations assessment does not discuss how local actors should be involved. In other literature three general categories can be discerned.

First, the involvement of local actors in large industrial projects, in particular oil and mining. This often involves a connection between local communities and vast engineering projects with their own constraints—in other words which have to balance negative impact on local society with mitigation and compensation measures. The main originators of this type of good practices are trade bodies such as International Council on Mining and Metals (ICMM), the global oil and gas industry association for environment and social issues (IPIECA), or the Society of Petroleum Engineers (SPE). Good practices can be retrieved online on the Equator Principles website[2] and in a reference by the International Finance Corporation of the World Bank called Stakeholder Engagement (IFC 2007).

Second, the involvement of local actors in assessments, monitoring and evaluation in humanitarian operations. There has been an increasing body of literature mainly around the Humanitarian Accountability Partnership,[3] as well as the Active Learning Network for Accountability and Performance (ALNAP). Other examples are available from bodies such as Collaborative for Development Action or NGOs.

Third, the academic literature on the connection between military operations and local populations. This has been heavily influenced by the Human Terrain Systems approach (McFate and Fondacaro 2011) and is evolving into a new field of cultural or social anthropology. NATO member states have developed guidance relating to influence (CICDE 2008), and for intelligence and socio-economic conditions.

2. See: http://www.equator-principles.com

3. Humanitarian Accountability Partnership is an association dedicated to establishing best practices in 'downward accountability' to local populations. See http://www.hapinternational.org

Effect on Local Actors and Local Dynamics

Opportunities of Local Actor Engagement

In previous and on-going NATO operations, local actors have been involved in assessment processes at the strategic, operational, and tactical levels. Their involvement has been mainly as recipients of information about on-going operations in their areas, or as "data sources" for use by a variety of international actors. Currently assessments usually take place against pre-designed plans that locals were not involved in developing. We are arguing that involving local actors in the design phase of any intervention will lead to increased relevance, effectiveness, and efficiency of operations and help avoid protracted conflicts. The challenge is to find an information/communication strategy that resonates with local actors.

Aspects the local actors will understand are those that relate to the quality of life, such as social services, transport, production, access to markets, stability, and security. One also has to realise the importance of status and the recognition of the dignity and importance (in local customs) of individual groups, the relation to government and the perceived longevity of authorities in power. At an even deeper and possibly more important level are to be found all the issues that underpin identity, religious values, cultural trust, and loyalty.

These are often the aspects of the local context, which NATO operations will impact. They are directly related to the manner in which local officials are contacted, how the meetings are carried out, what support is given to matters of policing or law enforcement, and how cultural appropriateness is demonstrated. They also relate to the balance to be given to formal justice and to informal justice, and the finer points of the ranking of local actors with which NATO will interact.

In summary, the consideration of local factors—such as customs, cultures, and religion—is crucial to understanding attitudes and point of views of local actors towards the operations taking place on the ground. The accumulated expertise of years of recent operations (i.e., Salmon 2010) should be considered, leading to the adoption of the following principles:

- Listen and connect, generate mutual respect and trust
- Build personal relationships

- Plan with the local actors, ensure local ownership, and deliver on commitments
- Be open and transparent, provide feedback and material learning
- Exercise continuity, patience, and focus on what happens after the handover

Risks Associated with Local Actor Engagement

The stronger involvement of local actors offers opportunities, but clearly also poses risks that have to be taken very seriously in any design for a stronger involvement of local actors in assessment processes. In a number of countries, especially with anti-Western and anti-American sentiments, suspicion about the presence of Western soldiers is very high. They are regularly seen as invaders and a danger for what is considered as local culture and structures. In a country like Afghanistan, being accused of spying for foreign forces can easily lead to death. Even when people do not support NATO's "enemies," collaboration with outside military forces, whether real or perceived, leads to strong reactions, and can undermine human security at personal and social levels.

Another challenge—with the risk of misunderstanding and hence missing the intended objectives—emanates from the complicated system of plans, theories of change, logical frameworks etc. associated with NATO operations. An untranslated version is often incomprehensible for local actors and differs very much from how they see the same situation. For example, a February 2012 study tested a UK Government Department for International Development (DfID) Logical Framework for a District Delivery Programme [(4)] with village elders in a district in Helmand province. The DfID approach was to fund schools through Afghan government channels with the goal of increasing local support to the government representatives at the district level. However, the village elders simply considered schools and education to be an important and good thing for their children, and it had little to no—or only temporary—influence on their loyalty towards the government, which seemed to be much more determined and influenced by local perceptions of the performance, behaviour and character of the district government officials.

4. For more details on the District Delivery Programme see: http://projects.dfid.gov.uk/project.aspx?Project=202190

If they were, for example, considered to be corrupt, parents were hesitant to send their children to school as they feared cheap building material used for the construction of the school that might endanger their children. Perceptions vary from village to village and can only be adequately captured on the lowest administrative level.

Another aspect is that state structures in hybrid conflicts are often viewed with suspicion and mistrust by local actors who instead use parallel social structures to guide their actions. When NATO is involved mainly with government-related local actors, NATO's influence is decreased. Moreover, when NATO plays into social rivalries by supporting individuals that compete with social structures already in existence, they are perceived as intruders, or even parties to these conflicts.

This requires again very careful thinking, planning and design for any broader involvement of local actors in assessment processes. Their position and standing, their influence in their region and their commitment or non-commitment ("spoilers") to NATO's operations has to be taken into consideration. In certain circumstances the simple act of participation can put in jeopardy the lives of the affected populations and actors, or can compromise the integrity of military systems. Some feel the risks associated with employing participatory approaches in crisis-affected populations overshadow the benefits due to the following factors:

- Risk of compromising the lives of persons approached by uniformed personnel, tainted by association and suspect of overt as well as covert cooperation
- Risk of threatening the cultural integrity or customs of the population due to the ignorance or neglect of local norms
- Risk of exposing military personnel in areas where cover is not provided and back-up support is not readily available
- Risk of releasing confidential or secret information to other sides
- These risks need to be put in context. Unlike intelligence operations, the main objective here is to obtain detailed socio-cultural information, not for targeting, but rather for the improvement of assessment processes that will contribute to the strategic, operational, and tactical goals.

Increasing Local Ownership

NATO Definition of Local Ownership and Its Limitations

"Local ownership" according to the NATO definition from draft stabilisation and reconstruction doctrine is an appropriation by national authorities of commonly agreed objectives and their active support their implementation.

NATO will find that the definition limits its own interaction and understanding of the local context, especially in fragile states where people do not relate to national institutions as authoritative sources of power. It is important to point out here that while the definition of local actors rightly includes non-governmental actors; the definition of "local ownership" only includes governmental actors. However, a key characteristic of weak and failed states is that the central government does not rule over the whole territory and is not able to enforce its will upon the whole territory.

Another shortcoming in the definition is the idea that the host nation government, the relevant authorities and the international community and NATO actually fully agree on principles and objectives, while local actors might not even be aware or understand them. In practice, most development and defence policy regarding interventions is decided in Western capitals. While efforts are made at high levels to involve host nations in identifying objectives, questions remain about the ability of nascent governments to adequately represent local concerns, without established democratic procedures in place. Furthermore, the need for fragile and emergent governments to secure international backing and aid may create pressure to agree to externally imposed objectives.

The authors feel that a good definition of "local ownership" should include an assurance that local actors understand NATO's end state, objectives and processes as a political and military organisation. Local perception is likely to depend on the initial military kinetic campaign making local actors reluctant to cooperate. It is for NATO to communicate clearly and simply the reasons for its presence and involvement. It is only as a result of local understanding about NATO that it might gain a gradual license to operate, which is based on the consensus that locals will draw some benefits from NATO's operations.

Local ownership includes the buy-in of influential people in the area of operation; and an alignment of international objectives to local objectives. In

a country like Afghanistan, or Syria, political action depends on individual personalities, their commitment, experience, worldview, access to resources and their relations with donors and foreign nations present in their area of responsibility. A minister, a local politician, a local village elder, businessmen or religious authority are often very influential in determining the course of local politics and in shaping local opinions and perceptions (including about NATO's operations).

The local structures should be analysed and understood *sui generis* and according to whom they are. Trying to see local structures through a dogmatic lens (e.g., "a government representative on the local level is supposed to perform tasks x, y, z") will lead to a distorted picture. In some areas, the government representative will actually fill that role, while in other areas a local non-governmental strongman, religious leader, or village elder will take over this function. It is therefore important to understand structures of influence and communication channels (between local actors, with the capital, communication access to NATO etc.) at local, national, and regional levels. In the case of Afghanistan the local actors are:

- Village elders (priority)
- Tribal, cultural, religious leaders (priority)
- Religious civil society more broadly, for example in Afghanistan the religious infrastructure in the country, including the Ministry for Hajj and Religious Affairs on the central, provincial and local level
- Host nation civil society representatives
- Elected civil servants at regional or local government level
- Non-elected civil servants
- Local authorities in general

An actor and stakeholder mapping should be developed that reflects both the influence of key actors, as well as the relationships between those actors. Once the local actor's structure is established, key actors should be analysed in their capacity, willingness, and motives to cooperate, support, or spoil NATO's operations and objectives.

A special emphasis should be paid to the religious civil society in a country like Afghanistan (but also Libya, Syria, and Somalia). The Ministry of Hajj

and Religious Affairs in Kabul influences the teachings and sermons of many religious clerics in the country. Being a relatively conservative Islamic country, with few people being able to read and write the sermons, advice and behaviour of religious clerics has strong significance for large parts of the population. A theological opinion about any given subject (including possibly about the presence and operations of NATO in the country) may have higher significance than any government legislation or opinion, especially in remote areas. If a military presence and operation or a development intervention funded/implemented by NATO is condemned as "unislamic" by local influential religious leaders, this will seriously and negatively affect the acceptance of any intervention and increase the risk for the staff/soldiers.

When the Goals Are Contradictory to the Intentions of Local Actors

If the intent of NATO and the intents of local actors are contradictory, this seriously hampers NATO's chances of reaching its objectives. In the case of Afghanistan, for example, both the government and the population heavily depend on external support and funding at present. This leads regularly to a situation where most people will not openly reject an intervention, as long as they see some benefit from it (such as getting a new school in a village). It also leads, however, to a situation where statements of loyalty may not reflect reality. In the worst case this could lead to a sabotage of an intervention that will not easily be recognisable. The problem for NATO is that any operation and intervention will ultimately fail, if the real commitment is lacking on the partner side. This commitment, at times described as "winning hearts and minds," is crucial and should be part of NATO's assessment processes. While governance and development indicators are currently already taken into account, assessments would benefit from more information about the population's perception of events and developments.

Improving the Quality of Operations Assessment

A Suggested NATO Methodology

The following methodology, if implemented by NATO, will significantly enhance their ability to engage with local actors.

Preparation and understanding phase

- *Understand the decisive aspects (determinant conditions)*. This includes mapping of actors regarding their societal influence and support for the measures of effectiveness in the assessment. It also involves understanding the non-security related drivers that affect the behaviour of the population
- *Organise for engagement*. This step is designed to develop a posture that will enable ease of access without increasing vulnerability on both sides. It includes identifying and training specific units in negotiation and communication and creating coordination groups and grievance mechanisms
- *Elaborate a game plan*. This involves developing picture plan of how the interaction with local actors could develop, including a model of consultation and data collection

Sector entry

- *Be present*. Develop a strategy of direct proximity with the local that results in real relationships
- *Avoid raising expectations or reacting superficially*. Local actors will require time to understand what is intended, and will not necessarily understand the nature of an operations assessment. They will tend to overestimate the significance of meetings. Similarly military personnel should avoid perceiving normality through the frame of past experience, and learn to suspend their judgment
- *Review other operations assessments*. Establish synergies with civilian organisations (e.g. humanitarian aid) and particularly local governments that are performing monitoring and evaluation assessments

Handover and exit

- *Generate a single narrative*. The operations assessment will generate an overall storyline, which will reflect the mission's implementation, and gaps that remain. For the local actors to be fully included, this will require a mutual understanding of the operations' objectives
- *Prepare handover from the beginning*. The rapid turnover of military units requires the establishment of independent capacities, which may be contractors, local authorities or civil society groups, who will perpetuate the process. Continuity is essential, even more so in an environment characterised by fast-paced and non-linear change
- *Keep your plans confidential*. This calls for a structural separation of information that can be shared with local actors with no risk of disclosure of intelligence, and those that should be maintained secret. This may be accomplished by separating higher-level change (internal) of a more military nature from that which can be "owned" by locals

Feedback and learning

- *Feedback and learning should be a continuum.* It is important to feedback results and assimilates better knowledge so that the campaign plan can be adapted. The learning can also be used as part of an information campaign, to ensure local actor engagement and further feedback loops

How to Engage?

- *Focus groups*. Focus groups run in various locations on regular basis can provide consistent and diversified information
- *Grievance mechanisms*. Grievance mechanisms have been most useful to private sector companies whose activities often heavily impact local communities. While NATO is results-oriented in its operational and tactical objectives, it might be unaware of the damage it causes on the way to completing its objectives. Establishing a complaint mechanism that allows for communication between individuals, communities and NATO could help reduce collateral damage, avoid it, or compensate for it. It first shows that NATO cares about local communities, provides a place for locals to turn to if they are threatened, and can help verify NATO's

relevance, effectiveness and sustainability of its operations. NATO should therefore consider this mechanism as a crucial communication and coordination tool, which can allow for confidence building with the communities that could have a stake in supporting NATO's objectives. In its absence, frustrations and anger on the part of communities can grow, which can easily backfire against NATO's efforts, and undermine local ownership.

- *External contractors*. The tradition of monitoring and evaluation in civilian agencies has led to the creation of a professional community of consultants and researchers which specialise in carrying out field assessments using systematic data collection methods. These have been used to good effect through the use of external contracts, giving these teams considerable autonomy about their logistics and mode of presentation, urging them to exercise their own duty of care.

Practical Measures

These steps can be undertaken by personnel in uniform, but they do point to the need for a broader set of capacities. We suggest the development of tools that will facilitate the task of inclusion of local actors:

- Operations assessments should be influenced by social science, because of the need to focus on contextual factors—for example the status of elders versus that of university educated youth in a given culture. The traditional separation between data collection and data analysis needs to be much reduced, with analysts involved in the data collection
- The strategies of the local actors are the most important factor in their involvement. They will invest time and potentially risk exposing themselves if they feel that the objectives pursued are useful to them, and do not threaten their values. As such, both the civilian effects and the military effects should be included in the goals of intervention, and the aspects that are materially relevant to them should be provided to the local actors in an intelligible communication

NATO may want to consider the use of 'civilian effects' teams that focus on non-target related objectives and operate relatively independently from the military. They could present information back to local actors, and design intervention strategies.

There will be a need to include specific data capture tools that can be analysed systematically, such as stakeholder mapping (which, unlike Social Network Analysis, allows for the capture of influence and attitudes) which captures the relations between all relevant actors in a defined area. An "events and trends mapping" in which local stakeholders, facilitators and international military and civilian personnel engage in a narrative exercise to capture the key drivers may further help understand the local environment. Once these sets of information are generated, they should be linked to a geospatial mapping that includes both military and civilian information such as development indicators and especially perceptions.

In order to capture the complexity and diversity of human, social, and political situations in an area of operation, a typology of various types of actors should be developed. This typology should be flexible enough to enable NATO to tailor their approach towards each actor in a conflict and culture-sensitive way. A proper understanding of structures of influence, hierarchy, and societal relations is needed. The mapping of social structures, and "who fits what where and how" is crucial to lessening casualties, strategic mistakes, and political agendas.

A sustainability standpoint is strongly recommended before any contact with any local actor, ensuring that s/he will not face problems with their environment afterwards, because of collaboration/betrayal/spying accusations for the foreign military. The tactical approach to local actors should be carefully planned and designed before any engagement. Local knowledge should be sought to verify and fine-tune NATO's approach to the involvement of local actors, and to identify possibly negative effects (i.e., UK Ministry of Defence 2012). A team of trusted local advisors should be hired to validate the do-no-harm approach and the cultural and social acceptability of the chosen approach.

Beware of what NGOs have come to call the "gatekeepers." These are the individuals who will seek to gain the monopoly of interaction with the foreigners to negotiate their own privileges, but who are not truly representative of

the local structure, or may even be actively repressing parts of the population. These should be identified and isolated.

Conclusion

Local actor engagement goes further than the collection of information that helps prevent detrimental behaviour. It also entails the development of an ad hoc structure (or methodology) that consists of phases and mechanisms designed to reap the benefits of their involvement. Forming active and sustainable relationships with various communities fulfils various objectives:

- Provide a discussion space that allows NATO to contrast its operational objectives with the reality as it is perceived and conveyed by local actors
- Provide local actors with an opportunity to express their needs, concerns, perceptions
- Allow NATO to ensure human security and contribute to building a societal project which is in agreement with the social structures already in existence (by adjusting objectives to the local needs, aspirations and understanding of the situation, NATO empowers drivers of change within the society as it is)

Through direct engagement with local actors, NATO deconstructs assumptions that provide a basis for its operational rationale, and verifies its veracity against local perceptions and understanding. Thanks to this verification process, NATO minimizes the risks of its operations, both for its personnel and for the populations at hand by managing to better understand conflict drivers, their nature and the motives that influence them.

Finally, a recommendation: NATO should consider including political guidance on the best ways to engage local actors in subsequent versions of the NATO Operations Assessment Handbook. This guidance would include a definition of what is means to engage local actors and methodologies for engagement. This would overcome some of the shortcomings encountered by the current definition of 'local ownership.' These methods or approaches for engaging with local actors should help bridge the gap between the implementation of plans by the NATO forces and international community and the often inadequate approaches used by the host nation governments to resolve local grievances.

References

Banko, K. 2013. Public opinion polling in operations assessment. In Williams A., J. Bexfield, F. F. Farina, and J. de Nijs. 2013. *Innovation in Operations Assessment: Recent Developments in Measuring Results in Conflict Environments*. Norfolk, VA: Headquarters Supreme Allied Commander Transformation.

Centre Interarmées de Concepts, de Doctrines et d'Expérimentations (CICDE). 2008. *Doctrine Inter-Armées des Opérations Militaires d'Influence*, PIA 03.253, n. 69 DEF/CICDE/NP.

International Finance Corporation (IFC). 2007. *Stakeholder Engagement: A Good Practice Handbook for Companies Doing Business in Emerging Markets*. Available online: <http://www1.ifc.org/wps/wcm/connect/938f1a0048855805beacfe6a6515bb18/IFC_StakeholderEngagement.pdf?MOD=AJPERES&CACHEID=938f1a0048855805beacfe6a6515bb18>

McFate, M., and S. Fondacaro. 2011. Reflections on the Human Terrain System During the First 4 Years. *PRISM* 4(2):63-82 <http://www.ndu.edu/press/reflections-human-terrain-system.html>

North Atlantic Treaty Organisation (NATO). 2012a. Bi-Strategic Command Input on Engaging Local Actors: Part II, 5000/TSCGXC0200/TT-7837/Ser: NU0007. Norfolk, VA: NATO Headquarters Supreme Allied Commander Transformation.

——— 2012b. *NATO Operations Assessment Handbook*. Interim Version 1.0. Norfolk, VA: NATO Headquarters Supreme Allied Commander Transformation.

Salmon, A. 2010. Embracing evolution post-Iraq – A response to counter-insurgency concepts. *Global Policy* 1(2).

UK Ministry of Defence. 2012. Joint Doctrine Note 2/12, Assessment, Annex 3A on Conflict Sensitivity.

9

The Use of Private Industry in Conflict Analysis and Assessment[1]

Emery Brusset, Andre Kahlmeyer, Bee Heal

Abstract

This chapter is a follow-on to Chapter 8 on the value of including local actors. It is generally accepted that engaging the private sector can increase access to expertise, save money, and increase mission flexibility. Here we argue that for operations assessments the private sector can: (1) help re-prioritise the information used to understand and subsequently rebuild relations with the population and state, which is particularly important in population-centric conflicts; (2) establish participatory assessment processes that capture the motivations of local actors; and (3) operate independently, under their own duty of care, beyond the strictures of "whole of government" approaches, while still providing the full range of benefits of civilian capabilities. Certain ethical precautions must be taken which revolve around agreement on objectives and procedures for information sharing.

Background

The centre of gravity for counter-insurgency operations at the strategic level usually revolves around reducing popular support for insurgent groups, thus isolating them from the population. This places a high importance on accurately understanding population perceptions and how they may react to changes—and military activities—in the theatre of operation. For example, when the military performs construction work, a simple quantitative measure

1. A Note From the Editors: The use and value of operations assessments has increased significantly in recent years, notably in stabilisation programmes, peace-keeping operations and peace efforts. In some situations NATO may want to leverage private contractors to conduct or help conduct these assessments and associated data collection. This chapter provides one view on how private industry can contribute to NATO assessments. As such, it is not an endorsement by NATO or the editors of any particular application or company. The authors were approached by the Editors to contribute this chapter, and received no compensation or payment.

such as kilometres of roads built probably does not capture the resulting attitudes of the affected population. If the local perception is that corruption was involved the final result may be negative.

The state of the art in this area has been pushed forward considerably by ISAF using opinion surveys. Unfortunately, the current tools and procedures still do not sufficiently capture the perceptions and opinions of populations to enable an understanding of the causes of progress or regress in security and development indicators. An example of this is the industry model constituted by the Helmand Monitoring and Evaluation Plan.[2] This monitoring system contains finely nuanced tracking of shifts in perceptions relating to stability, in time and space. It has, however, been less apt to evaluate specific interventions, by establishing clear causal links between an intervention and a shift in perception.

To help obtain this deeper understanding there are a number of data collection methods in the civilian world that provide "eyes and ears" on the ground. One example is snowball sampling, which is based on gradual penetration of a population group through trust. It relies on local experts that melt into their environment to reduce security threats.

The Place of the Private Sector

The military frequently outsources tasks to the private sector (for example companies involved in logistics) to access experts in a specific field and to reduce the need for standing capacities that can be delegated to contractors, especially when these are not related to war fighting and can be delivered just as professionally by civilian contractors.

Oddly, this rationale to use the private sector has not been widely extended to the field of operations assessment by NATO member states or even by NATO, with the exception of contracting out survey companies. While many firms are used to collect data, the analysis is often not done by them. The design and the final translation of the findings are rarely outsourced. The few examples

2. This is a UK funded programme to monitor progress on the Helmand reconstruction strategy, involving various socio-economic dimensions. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/28106/20110312_HelmandAnnualReview_Final.pdf>

of outsourcing of analysis are generally for situational awareness, except when it touches on intelligence capabilities.

At the same time there is a large community of expertise in the related civilian development fields of monitoring and evaluation. Examples include the web of national and international evaluation society associations, and professional activities such as the Organisation for Economic Cooperation and Development (OECD),[3] the United Nations Evaluation Group (UNEG),[4] and development banks. Other specific examples include evaluation units at the Ministry of Foreign Affairs in Germany, another one for the Ministry of Cooperation, and two others for the Development Bank (KFW)[5] and for GIZ,[6] the technical development agency. Each one of these relies on independence of judgement when conducting their evaluations and is supported by a large number of consultancy and contractor firms, many of which are visible on the internet. In addition there is an increasing body of literature including the Humanitarian Accountability Partnership,[7] as well as the Active Learning Network for Accountability and Performance (ALNAP).[8]

Expanded Expertise Leading to Reprioritisation of Information

The concepts used by civilian firms engaged in evaluation and operations assessment hinge on the chain of effects triggered by an action—going from the action, through the reaction of the population (and not only the target population), to the consequences of this action. They focus on impact (both intended and unintended) and the direct and indirect changes that resulted from the action. They look beyond measures of performance to understand the decisive conditions, or drivers, of the situation on the ground.

Such an understanding is derived above all from an analysis of the social, economic and cultural environment. It must also include knowledge of governance, and of physical geography. Employees of these firms have expertise

3. See, http://www.oecd.org/dac/evaluation/
4. See, http://www.uneval.org/
5. See, http://www.kfw.de/kfw/en/Development_Finance/index.jsp
6. See, http://www.giz.de/
7. See, http://www.hapinternational.org/
8. See, http://www.alnap.org/

in these areas and also are skilled at obtaining information. Their ability to withstand tough travelling and living conditions, ambiguity of mandates, changes in schedule, their patience and empathy, their elicitation skills, are additional important attributes.

With these skills the private sector can quickly seek, identify and convey crucial information about a complex environment needed for high quality operations assessments. This type of sectoral expertise and availability is, at times, lacking in military deployments. As cited by American intelligence specialists quoting the situation in Afghanistan:

> At the battalion level and below, intelligence officers know a great deal about their local Afghan districts but are generally too understaffed to gather, store, disseminate, and digest the substantial body of crucial information that exists outside traditional intelligence channels. A battalion S-2 shop[9] will, as it should, carefully read and summarize classified human intelligence (HUMINT), signals intelligence (SIGINT), and significant activity (SIGACT) reports that describe improvised explosive device (IED) strikes and other violent incidents. These three types of reports deal primarily with the enemy and, as such, are necessary and appropriate elements of intelligence. What lies beyond them is another issue. Lacking sufficient numbers of analysts and guidance from commanders, battalion S-2 shops rarely gather, process, and write up quality assessments on countless items, such as: census data and patrol debriefs; minutes from shuras with local farmers and tribal leaders; after-action reports from civil affairs officers and Provincial Reconstruction Teams (PRTs); polling data and atmospherics reports from psychological operations and female engagement teams; and translated summaries of radio broadcasts that influence local farmers, not to mention the field observations of Afghan soldiers, United Nations officials, and non-governmental organisations (NGOs). This vast and underappreciated body of information, almost all of which is unclassified, admittedly offers few clues about where to find insurgents, but

9. This refers to the branch in a company level or below military headquarters responsible for collation, analysis and management of intelligence support to military operations.

> it does provide elements of even greater strategic importance—a map for leveraging popular support and marginalizing the insurgency itself. (Flynn, Pottinger, and Batchelor 2010, 7)

In summary, beyond the mantra for more specific socio-cultural information (encompassing governance, economics, and crime issues), operations assessments of counter insurgencies also need people skilled at discerning the key drivers, and the causal links of social change. This reprioritisation of information to support an understanding of social phenomena can often be best accomplished by individuals in the private sector.

Access To and Engaging the Population

As highlighted in the previous chapter on engaging with local actors, it is increasingly clear that the success of an operation will depend on maintaining open channels of communication with a wide range of stakeholders—people that can influence, or be influenced by, an operation. This open flow of communication will depend in great part on the sense that host populations have that they are part of the process used by NATO to plan and conduct operations. The process whereby information is obtained for operations assessment is of an importance that compares with the conduct of military operations. It is also increasingly clear that it cannot be only a process of information extraction, as the experiences of the US Department of Defence Human Terrain Systems have shown (McFate and Fondacaro 2012). Information creates a common bond, and is the precondition of trust, that most essential ingredient.

In other words, access is not motivated only by a need to generate knowledge. The very presence of observers changes the environment, and the perceptions of the objects being studied. The presence of evaluators or analysts is, at the very least, a message conveying interest and may result in a relationship with the potential of becoming a force for change in itself. This is the notion of assessment as a form of engagement.

The ease of movement and similarity of outlook often enables civilian contractors to engage local population and carry out a dialogue with the national stakeholders. They are sometimes more accepted than uniformed personnel, or civilian personnel operating under military escort. Some of the ethical debates

that have occurred in the United States when Human Terrain Systems teams accompany military units[10] are an example of this distinction.

At the same time a large challenge with using civilian contractors in operations assessment, revolves around the ethics of information sharing and neutrality. While the well-being of the population will in most cases be increased by the knowledge obtained by civilian contractors, guarantees will need to be given that this information will not be used for targeting purposes. Such use would immediately cancel the unique nature of the access that civilian contractors have to the population. Moreover, the civilian contractors will need to see their work limited to the assessment of progress in civilian effects.

It is generally considered in a military organisation that the least information shared the better for the force, whereas for a civilian assessment contractor the sharing of information is usually so wide that only direct safety precautions constrain it. However, screening information that the military could use for operational targeting may be appropriate, as the military will avoid sharing information on targets. Having an inception stage at the outset of a civilian contractor deployment to ensure alignment of strategic aims and the nature of the objectives pursued may help generate a mutual understanding of roles.

The advantages of civilian contractors in operations assessments are not the same as for the security or logistical services. They lie in the singular type of knowledge provided about social change, leveraging the dynamics of familiarity and loyalty that civilian actors can generate (if well trained). They do not replace large standing non-core military capabilities, nor create deniability—quite the opposite, as they extend the reach of the military, and serve to demonstrate a positive intent. In insurgency situations it is increasingly important to demonstrate presence and intent. Engaging with local actors can help achieve this goal.

Flexibility

The private sector can not only tap into a wide range of expertise through avenues such as subcontracting, it can also perform functions that are not

10. There has been much academic debate about the US DoD Human Terrain System and the ethics surrounding their methodologies in Afghanistan. See http://www.aaanet.org/about/Policies/statements/Human-Terrain-System-Statement.cfm

available to military forces. The capability of a private sector team goes much beyond the actual team deployed and its own management of resources, and includes an "exoskeleton" of knowledge (access to relevant expertise in real time), operational capacities (deployment in good conditions, sub-contracting of personnel if not available immediately), and project control. The latter includes the ability to coordinate and synthesise the inputs of a vast array of specialists, some of which may never before have carried out the operations assessment tasks.

In addition, the private sector often has a degree of flexibility that cannot exist for civilian personnel operating under military mandates. The private sector employees are usually less hampered by bureaucratic restrictions on movement, handling of resources, safety and security guarantees, reporting lines, etc.

Characteristics of and Access to the Private Sector

The existence of a large community pre-dates its employment for operations assessment. Since the early 1990s there has been a growing involvement of the bilateral and multilateral evaluation units within each aid and development ministry and agency, dealing with conflict related programming. This culminated in the Joint Evaluation of the Emergency Response to the Crisis in Rwanda (Eriksson et al. 1996) but has since spread to programmes and initiatives in all "fragile" countries.

The private sector assessment community is usually accessed through public tenders by public authorities, NGOs, and even increasingly by the private sector (mining and petroleum—although in the latter case there is still a prevalence of negotiated contracts). The writers estimate that the total budget spent annually is approximately US$25 million, if one assumes an average budget of US$ 2 million per OECD, EU, development bank and UN member equipped with an evaluation unit (some of the member states have two or three). This represents full time employment for many individuals around the world and is a solid basis on which to begin to develop capacities specifically aimed at military operations assessment.

The community is organised into for-profit or non-profit entities. The smaller firms usually have a specific technical depth (typically numbering

less than 30 staff), while the larger ones have the ability to tap into a number of sectors (these contractors typically implement programmes but also have assessment units). All are required by contract to have adequate insurance cover, communications and mobility. Military security requirements (security infrastructure, communication, security clearances) can usually be accommodated where required.

The firms will hire individuals for specific bodies of knowledge based on the scope of the assessment: area expertise (local personnel and academic specialists such as historians or anthropologists), sector expertise (small scale infrastructure compatible with the local environment and anti-corruption measures, etc.), institutional expertise (of NATO, of the national government), and finally analytical capacity (carry out surveys, test for bias, analyse large amounts of qualitative data, etc.).

A civilian team is typically composed of a single team leader, who is responsible for the delivery of a quality product. This person coordinates a team of experts in three or four different domains: area knowledge (culture and history mainly), institutional knowledge (e.g., public administration), and technical knowledge (infrastructure, financial management, food security, etc.). In some cases the team leader is a specialist in evaluation methods, although other team members could have this as their dominant or secondary expertise.

Conclusion

Properly used, the private sector can improve the conduct and content of operations assessment in counter-insurgency and state-building operations where the population itself is the centre of the analysis. This chapter has identified three unique contributions the private sector can make to advance the cause of operations assessment: they can obtain information outside the adversary-centric focus of the military and rapidly turn it into knowledge; they can engage the civilian population in ways that are conducive to better relationships and create opportunities for coalitions; and they can move around in a way akin to journalists to access areas not open to the military. One major risk remains which touches on the duties and rights of civilian contractors, issues that also exist for other civilian organisations, even within the same member state.

References

Eriksson, J. R., T. Sellström, H. Adelman, A. Suhrke, B. Jones, J. Borton, and K. Kumar. 1996. *The International response to conflict and genocide: Lessons from the Rwanda experience*. Steering Committee of the Joint Evaluation of Emergency Assistance to Rwanda.

Flynn, M. T., M. Pottinger, and P. D. Batchelor. 2010. *Fixing intel: A blueprint for making intelligence relevant in Afghanistan*. Washington, DC: Center for a New American Security.

McFate, M., and S. Fondacaro. 2012. Reflection on the human terrain system during the first 4 years. *Prism* 2(4):63-82.

PART 4:

NEW APPROACHES, IDEAS AND CHALLENGES

10

Complexity Theory: Implications for Measuring Peacebuilding Progress

Dr. Cedric De Coning

Abstract

When crisis situations become so dynamic that we are no longer able to keep track of the effects of all the specific initiatives underway we commonly refer to them as 'complex.' What does it mean when we say a particular conflict, or the international response to it, is complex? What can we learn from applying the knowledge generated by the study of complexity to the peacebuilding context? Could insights from complexity theory assist us in improving our understanding of some of the core challenges experienced by peacebuilding systems? For the purposes of this chapter, complexity is described as a complex system that has the ability to adapt, and that demonstrates emergent properties, including self-organising behaviour. It comes about, and is maintained, as a result of the dynamic and non-linear interactions of a large number of its elements, based on the information available to them locally, and as a result of their interaction with their environment, as well as from the modulated feedback they receive from the other elements in the system.

Introduction

In this chapter we will introduce complexity and then consider some of the implications of it for peacebuilding policy and practice. When referring to policy we will address the way we think about peacebuilding, and especially what we regard as the appropriate roles and relationships between the local and international actors in any given peacebuilding context. And when we refer to practice, we will address the analysis-planning-implementation-assessment cycle, and we will consider what implications an application of complexity may have for our understanding of conflict analysis, design and planning, management and coordination, and assessments and metrics. We will conclude this

chapter with a discussion on the implications of complexity for the unintended consequences of peacebuilding and what can be done to manage them.

Complexity

Paul Cilliers describes complexity as a system that has the following characteristics:

- It consists of a large number of elements
- These elements interact based on the information available to them locally (none of the elements are able to comprehend the complexity of the system as a whole)
- At least some of the elements also interact with the environment (it is an open system)
- The interactions are rich, non-linear, dynamic and they feed back on each other (recurrence)
- The conditions under which such a system operates are far from equilibrium, i.e. the elements are under sustained pressure
- The combined result causes such a system to spontaneously organise itself, maintain itself, and adapt (there is no external, controlling agent)
- Over time, this process develops a history, i.e. complex systems evolve over time and the past is co-responsible for the present behaviour of the system, i.e. a complex system cannot be understood as a snapshot of the present, without also taking its evolving history into account (Cilliers 1998)

A number of other thinkers engaged in the study of complexity have identified similar characteristics, and some have added additional attributes (e.g., Mitleton-Kelly 2003;[1] Clemens 2001;[2] Ramalingam and Jones 2008[3]).

1. Self-organisation, emergence, connectivity, interdependence, feedback, far from equilibrium, space of possibilities, co-evolution, historicity and time- and path-dependence.

2. Fitness, co-evolution, emergence, agent-based systems, self-organisation, self-organised criticality, punctuated equilibrium and fitness landscapes.

3. Interconnectedness and interdependence, feedback processes, emergence, non-linearity, sensitivity to initial conditions, trajectory in phase space, relevance of chaos and the edge of chaos, adaptive agents, self-organisation and co-evolution.

When comparing these different sets of characteristics, it is clear that there is great degree of convergence among complexity thinkers about the core characteristics that constitute complex systems. In this introduction to complex systems three of these core characteristics, namely a whole-of-systems approach, non-linearity and self-organisation, will be discussed. In the process, a number of related concepts that form the basis of our understanding of complexity, including emergence, adaptation and feedback will also be explored.

A Whole-Of-Systems Approach

The concept of complexity is embedded in a whole-of-systems approach. A system can be defined in a very general sense as a collection of interacting elements that together produce, by virtue of their interactions, some form of system-wide behaviour (Mitchell 2009). In other words, a system is a community of elements that, as a result of their interconnections, form a whole. In complex systems, the interaction is dynamic, i.e. a complex system changes with time (Cilliers 1998).

Complexity is not, however, interested as much in the agents as nodes in the system, as in the patterns of their interconnections and how that generates meaning or purpose in the system as a whole (Cilliers 1998). In other words, complexity is interested in how the elements interact and how this interaction develops into the system as a whole having new capacities that did not exist within the individual elements.

In complex systems, the whole has properties that cannot be found in the constituent elements or in the sum of their properties. In social systems, for instance, the society as a whole develops and maintains norms and identities that serve the common needs of the community. In some ways this results in suppressing some of the interests and needs of the individual and of special interest groups in the interest in the general wellbeing and survival of the society as a whole. Morin points out that not only is the whole more than its elements because new qualities or properties emerge due to the organisation of the elements in a whole, but the whole can also be less than the sum of its parts because "a certain number of qualities and properties present in the parts can be inhibited by the organisation as a whole" (Morin 2005, 11).

The concepts 'social' and 'society' conjure up images of systems made up of people that share a common socio-cultural, national or civic bond. When

studying people in the context of them being part of a society, as opposed to studying them as individuals, a different side of their being—including aspects related to their role in society as well as aspects related to the restrictions that conforming to the society places on them—is revealed. These are aspects of their being that could not be revealed by studying them in isolation from their place in a social system.

By studying the society as a whole made up by the patterns of activity of the individuals and the various networks and sub-systems—such as family, clan and tribe that develop out of these patterns—we reveal insights into the way individuals derive meaning from their roles in a community and how the interactions between these individual roles shape, sustain and transform both the society as a whole and the individuals that make up that society. These are insights that could never be identified by studying only the individual.

In moving from the individual to the community and society, we come across organisation. Complex systems cannot do without hierarchy and structure, but in complex systems hierarchy is not hard-wired or externally determined and controlled; the hierarchy of a complex system is emergent and self-organised and thus changes with the system as it adapts and evolves in response to its environment (Cilliers 2001). The vitality of the system depends on its ability to transform itself, including its structure and hierarchy. Hierarchy thus is a typical characteristic of complex adaptive systems, but in complex systems hierarchies themselves exhibit complex adaptive characteristics (Chapman 2002).

The last aspect of a whole-of-systems approach that should be discussed is the role of boundaries and borders in complex systems. Complex systems are open systems and this implies that interactions take place across their boundaries (Cilliers 2002). These interactions take place with other systems and the environment, for instance there is a flow of information and/or energy between the system and its environment through its boundaries. Systems consist of interrelated subsystems, and some boundaries can thus fall within larger systems or share borders with them (Chapman 2002). Not all sub-systems are neighbours physically; some are virtually linked—in social systems agents far away from each other may link up via social media, for instance, and collaborate, coordinate and otherwise influence each other's systems and in this way interpenetrate such systems.

Complexity thus builds on and is grounded in a whole-of-systems approach. However, it is concerned with a specific type of system, namely 'complex' systems, and to gain more understanding of that differentiation we turn to another set of important properties of complexity, namely non-linearity and self-organisation.

Non-Linearity

In the previous section, a whole-of-systems perspective was introduced and it was explained that complexity is interested in the patterns of interconnections among the elements, and how this dynamic interaction generates properties beyond those that exist in its constituent parts. In this section, the second characteristic of complexity is introduced, namely that in complex systems the causal patterns of these interactions are non-linear—the outputs are not proportional to the inputs (Hendrick 2009).

Jervis argues that mathematical linearity involves two propositions, namely: (1) that the changes in a system's outputs are proportional to changes in its inputs and thus, (2) that the system's outputs corresponding to the sum of two inputs are equal to the sum of the outputs arising from the individual inputs. He goes on to point out that we often intuitively expect linear relationships. For example, if a little foreign aid slightly increases economic growth, it is expected that more aid should produce greater growth.

However, complex systems often display behaviour that cannot be understood by extrapolating from the units or their relations, and many of the results of actions are unintended (Jervis 1997). Thus, an important characteristic of complex systems is that non-linear variables may have a disproportionate impact at one end of its range (Byrne 1998). Non-linearity thus refers to behaviours in which the relationships between variables in a system are dynamic and disproportionate (Kiehl 1995).

The first characteristic of non-linearity is that the outputs it generates are not proportionate to its inputs, that is they are asymmetrical. The second aspect of non-linearity is that non-linear systems do not follow a pre-determined, and thus predictable, cause-and-effect path. Nor can such a path, once traced in hindsight, be replicated to generate the same effect. A third aspect of non-linearity that sets complex systems apart from complicated ones are that they cannot be reduced to something simpler, like a set of laws or rules

that can help us to predict the behaviour of the system. Cilliers explains that "a large system of linear elements can usually be collapsed into an equivalent system that is much smaller" (Cilliers 1998, 4). Non-linear data sequences and non-linear system processes cannot be reduced to formulae or rules that can compress the amount of information necessary to manage them, or to make them otherwise predictable and controllable.

A fourth characteristic is that linear logic cannot be used to explain the behaviour of complex systems. Non-linearity generates outputs that are not necessarily proportional to the inputs; in complex systems it is possible for two inputs to generate an outcome that is larger, or smaller, than the sum of the two inputs together. In complex social systems we often talk of indirect or unintended consequences. For instance, one may organise a training course with the aim of imparting a skill, but then it turns out that the most important benefit that the participants gain from the training is not necessarily the skill, but the team-building and social networking.

As these four characteristics have demonstrated, our common-sense understanding of non-linearity is often closely associated with the concepts of disorder, chaos and randomness because we typically explain non-linearity as the opposite of the linear, the logical and the orderly. It is thus important to emphasise that in the context of complexity non-linearity is not associated with disorder. In fact, non-linearity is an essential ingredient in the processes of emergence and self-organisation that generate order in complex systems.

Non-linearity has been presented as the element that distinguishes a complex system from a linear-deterministic mechanical system. The latter is fully knowable, predictable and, therefore, controllable in principle. It, therefore, is also unable to do anything that is not pre-programmed or designed if it is man-made system or new in the sense that we could not know of it in advance if it is a natural system. In contrast, the non-linearity in complex systems is what makes it possible for these systems to adapt and to evolve, i.e. to create something new that goes beyond what is pre-programmed in the parts that make up the system. Non-linearity is thus an essential part, in fact a pre-condition, for emergence, self-regulation and adaption in complex systems (Cilliers 1998).

One of the ways in which complex systems use constraints to maintain themselves within certain parameters is through the use of feedback mechanisms. When certain thresholds are crossed, positive or negative feedback

is used to correct the system back to within the parameters. While complex systems may thus theoretically be capable of a huge variety or range of actions, their behaviour is typically constrained within a fairly limited range of options. While individuals may thus be theoretically free to choose any action, their behaviour is typically constrained to within a fairly limited range of options by influences such as what would be regarded as legal, moral and appropriate by an individual's society, family and friends. When an individual acts outside of these parameters, feedback is applied through a range of social sanctions that, in most cases, serve to direct the individual back to within the social norm.

At this point, the first two complex-systems characteristics were introduced, namely the whole-of-systems approach and non-linearity. Let us turn now to the third characteristic, namely self-organisation.

Self-Organisation

Self-organisation refers to the ability of a complex system to organise, regulate and maintain itself without needing an external or internal managing or controlling agent.

Take for example the economy of any reasonably open economic system. An economic system is a self-organising system in that it continuously responds to a large number of factors without requiring a controlling agent (Cilliers 1998). The economy is often discussed as if it were an organism, but we need to think of it more as an ecosystem because it is not the economic system as a whole but rather the individuals and organisations that constitute the economic system, that individually consider and respond to the factors that matter to them. It is the cumulative and collective effect of their actions that determines the overall behaviour of the system.

The state of the economy in any given country or region depends on a large number of dynamic factors. As these conditions vary, the individuals and organisations in the system continuously adjust their actions so that they can reap the most benefit from the prevailing conditions. Each individual or organisation acts in its own self-interest, but sometimes their actions can have significant implications for the system as a whole, especially when a series of individual actions aggregate into swarm behaviour—where the actions of some trigger behaviour by others that result in large swarm-like fluctuations in the system as a large number of individual agents respond similarly in what appears

to be coordinated behaviour. For instance, a large number of people may start fleeing when a rumour spreads that an attacking force may be approaching. Or a large number of investors may start flocking to a certain market or stock as rumours spread of its good prospects.

There are also some economic agents that are trying to influence the system in what they perceive to be in the best interest of their sub-system or even the system as a whole. Governments, central banks, and multilateral institutions like the International Monetary Fund or the World Bank may, from time to time, try to act in ways that they perceive to be in the interest of the world economy or the economy of a region or a specific country. Their actions, however, only constitute another input into the system, and they do not have control over how the system responds to their inputs. We can thus not regard them as controlling agents. At best they are some of the more influential agents in the system.

The organisation of the economic system as a whole thus comes about as a result of the interaction between the various agents that constitute the system and its environment (Cilliers 1998). There is no single agent or groups of agents that controls the economic system, but there are many agents that try to influence the behaviour of the system, and there are many more who simply respond to what they perceive to be the current state or future direction of the economy. The economy self-organises spontaneously, and this is an emergent process that comes about as a result of the cumulative and collective interaction of all the agents in the system.

As discussed in the previous section, this process is non-linear and dynamic and thus cannot be predicted or controlled. So many causal reactions are happening simultaneously that no one agent or group of agents working together can control the system.

Although a complex system like the economy is too complex to model deterministically (Cilliers 1998), it is possible to influence it at various levels. As mentioned earlier, many organisations, like central banks, exist explicitly for the purpose of trying to influence the economy. Non-linear causality generates asymmetrical relations, which implies that relatively powerless agents can sometimes have a disproportionate effect on the system. However, the effects of any such interventions, regardless of the relative power of the agent, usually only influence the system in the short to medium term because the rest

of the agents in the system will respond to any new developments, and these responses will impact on each other and result in further waves of reactions. The cumulative and collective effect of these responses will result in the system as a whole responding in unpredictable ways.

Another important property of complexity in general, and self-organisation in particular, that has been referred to several times before is known as emergence. Emergence is important because it explains how the elements in the system are not just merely interacting with each other in order to maintain themselves. In complex systems, the interactions of the elements generate a new collective effect (or effects) that would not have occurred if the different agents acted on their own. New system characteristics 'emerge' through the process of interaction (Cilliers 1998). Morel and Ramanujam (1999) explain self-organisation as a process of spontaneous creation of complex structure that emerges due to the dynamics of the complex system, which makes self-organisation an emergent phenomenon.

The dynamic and non-linear relationships among the components in complex systems generate new emergent properties, namely properties that cannot be predicted merely by analysing the individual components of the system. Complicated systems do not have emergent properties, and the way in which they work can potentially be fully understood, and predicted, by analysing their components and the rules that govern their interactions. In a complicated system, disorder is understood as entropy, namely as the loss of energy in the system that, if unchecked, will result in the gradual collapse of the system into disorder. In contrast, non-linearity and dynamism play a critical role in creating and sustaining order in complex systems, that is to say in enabling order to emerge (Cilliers 1998). This change over time—the way in which a system adapts on the basis of its own internal processes as well as its interaction with its environment and the way in which it generates new structures, forms and functions—is what is meant by emergence. A key characteristic of complex systems is thus that they emerge and maintain themselves spontaneously, without the intervention of an external designer or the presence of some form of internal or external controlling agent (Cilliers 1998).

Three of the core characteristics of complexity, namely a whole-of-systems approach, non-linearity and self-organisation have now been introduced, and key concepts such as feedback and emergence have also been discussed. In the

next sections the focus is on the implications that non-linearity, self-organisation and emergence have for complex operations.

Implications of Complexity for Peacebuilding Policy

What insight can be gained from applying complexity to peacebuilding? The most fundamental is probably the realisation that the ability of external agents to gain knowledge of the complex social systems we are dealing with in the peacebuilding context is inherently limited. Complex systems—which include all social systems—are non-linear, and this means that we are not able to know enough about these systems to predict their behaviour using a linear cause-and-effect science model.

Concepts like statebuilding and peacebuilding convey the assumption that we are able to 'build' the state and 'build' peace, in the same way we can design and build a bridge or a tunnel. But social systems are part of the organic world, not the material world, and we can't build them, we can only nurture and influence them. Social systems are not like a machine where the parts have a specific pre-designed role in a closed causal loop, with one ultimate purpose and only one pre-determined way to achieve that purpose. When a machine is stressed it breaks down and needs to be repaired. In complex systems the elements relate in open non-linear ways and this enables the system to evolve, so that it can find new ways to pursue its goals and reach its objectives, despite obstacles, stressors and constrains.

When we approach conflict as if it is something in a society has been broken and needs to be repaired, or if we approach peace as if it is a problem that needs a design solution, then we fail to grasp the complexity of the social system we intend to engage with. It is only when we recognise that it is impossible to 'fix' or 'build' complex systems from the outside, and that at best we can try to influence them, that we are starting to understand the complex nature of the phenomenon we intend to engage with.

What are some of the implications that flow from these insights? First, we have to change our theories of change. We have to acknowledge that there are no off-the-shelf solutions and no one theory of change, or model of state transformation, that can claim universal applicability. We have to come to terms with what it really means when we say that something is context-specific. It means that a sustainable social-political order can only emerge from that

context. It means that we cannot import a universal model and simply make a few adjustments for the local culture and context.

Second, complexity theory has shed light on how complex systems self organise. Self-organisation in the statebuilding and peacebuilding context refers to the various processes and mechanisms a society uses to manage its own peace consolidation process, that is to say the overall ability to manage its own tensions, pressures, disputes, crisis and shocks without relapsing into violent conflict. For statebuilding and peacebuilding the implication is that interventions have to be essentially about stimulating and facilitating the capacity of societies to self-organise.

Seen in this context, peacebuilding is a very delicate and self-contradictory process fraught with built-in tensions. There is an inherent tension in the act of promoting a process of self-organisation; external interference undermines the ability of the 'self' to develop (to take responsibility, to learn from failure and successes) sufficiently for self-organisation to emerge. Understanding this tension—and the constraints it poses—helps us to understand why peacebuilding is so complex. It should also free us from illusions of easy solutions and grand models and help us to focus on case-by-case transactions that seek to reflect the interface between local context and international interests.

Many, if not most, international peacebuilding missions to date have made the mistake of interfering so much that they ended-up undermining the ability of the local system to self-organise. External peacebuilders impose neoliberal political and judicial norms and model institutions according to their own ideal types. In the process we deny these societies the room to develop their own institutions which are emergent from their own history, culture and context. External peacebuilders fail to recognise the degree to which their own norms and institutions are the product of their own history, culture and context. Consequently, they underestimate the challenge of transferring these norms and institutions to other cultures and contexts. The key to successful statebuilding and peacebuilding lies in finding the appropriate balance between external security guarantees and resources, on the one hand, and the degree to which the local system has the freedom to develop its own self-organisation, on the other.

We may be able to identify and agree on some broad principles in, for instance, the form of a statebuilding and peacebuilding code of conduct.

Ultimately, however, what is appropriate has to be determined in each specific context, as in the articulation a compact between the local authorities and representatives of the international community. As these processes are dynamic and non-linear, what is appropriate will be continuously changing and such compacts would thus also need to ability to evolve.

Applying complexity theory to peacebuilding we can conclude that self-sustainable peace is directly linked to, and influenced by, the extent to which a society has the capacity and space to self-regulate. For peace consolidation to be self-sustainable it has to be the result of a home-grown, bottom-up and context-specific process. The robustness and resilience of the self-organising capacity of a society determines the extent to which it can withstand pressures and shocks that risk a (re)lapse into violent conflict. Peacebuilding should thus be about safeguarding, stimulating, facilitating and creating the space for societies to develop robust and resilient capacities for self-organisation.

International peacebuilding interventions should provide security guarantees and maintain the outer parameters of acceptable state behaviour in the international system, and they should stimulate, facilitate and create the space for the emergence of robust and resilient self-organised systems. International peacebuilding interventions should not interfere in the local social process with the goal of engineering specific outcomes, such as trying to produce a neoliberal state. Trying to control the outcome produce the opposite of what peacebuilding aims to achieve; it generates on-going instability and dependence, and it undermines self-sustainability. The art of peacebuilding thus lies in pursuing the appropriate balance between international support and home-grown context-specific solutions.

Implications of Complexity for Peacebuilding Practice

In this section we consider the implications of complexity for peacebuilding practice, and we will focus on the analysis-planning-implementation-assessment cycle, with sub-sections on (i) conflict analysis; (ii) design and planning; (iii) assessments and metrics; (iv) management and coordination; and (v) unintended consequences.

Conflict Analysis

When something is complex it cannot have one definitive problem-set. We should thus not attempt to solve complex social problems with methodologies designed to identify such limited problem-sets, so that such analysis can then lead to a neat matching set of solutions. This insight from complexity stands in stark contrast with the dominant determined-design approach in peacebuilding practice. The current best practice in international peacebuilding is to start any new mission or programme, or any serious review of a current mission, with a conflict analysis that is aimed at determining the root causes of the conflict, with a view to then generate a set of actions that will address, and ultimately resolve these root causes.

This dominant approach is deterministic and it is based on two beliefs: first that there is problem that can be discovered and solved, and second that the international peacebuilding system has the ability to diagnose such a problem and to design and administer a cure (the solution to the problem). The main problem, from the perspective of the deterministic approach, is lack of resources. Because the potential resources for any particular intervention are limited, the purpose of the needs assessment and conflict analysis is to determine priorities. If the problem seems difficult to diagnose then the problem is related to a lack of resources dedicated to the analysis—all problems are solvable by the international peacebuilding system provided you devote enough attention (read resources) to it.

In fact, most peacebuilding agencies have a pre-determined supply driven response capacity, and the purpose of the analysis is simply to link the supply and the 'perceived' need. If your agency deals in food security, there is a high probability that your needs assessment, or your representative in a joint needs assessment, will find that there is a food security problem. If your agency delivers peacekeeping missions, or mediation services, you are likely to find that the problem requires a peacekeeping or mediation solution.

The analysis is thus not aimed so much at diagnosing the problem, because the 'solution' has already been pre-designed by the mandate of the agency, the only real question is how big the need is. There is thus an assumption that a prior generation has done a kind of macro-analysis that has resulted in the current generation of institutions, and that we should not re-invent the wheel, we should just focus on how best to apply, and perhaps further refine, these

solutions. How much of what we are designed to deliver does this particular situation require? How big should our operation be? It is thus assumed that the international intervention will not be able to deal with all the problems and needs identified, so the analysis must identify the most critical problems and needs, so that the plan can be designed to focus on the priorities identified.

The insights from complexity for peacebuilding we have discussed earlier suggest an alternative approach to conflict analysis. First, a complex systems approach reminds us to be mindful not just of the specific aspect we may be focussed on, but to consider it in the context of the wider system of which it is a part, including the context and environment within which that system operates. Second, it reminds us of the importance of the dynamic and non-linear interconnections among the agents in the system, and the way in which these interactions generate context specific meaning in the system. It reminds us that any given manifestation of a problem is unique to that context. What may explain one context may not necessarily work in the next.

Third, it tells us that there is no one single 'state of affairs' or set of 'root causes' that can be identified and solved. We can not intervene in one system in isolation, and even it we were able to do so, that system will not be static. We will always have to deal with multiple systems that are constantly in the process of interacting with each other and the larger international and regional environments of which they (and us) are part. To understand the system we need to see it in motion, and in relation to its environment, not in freeze-frame and not in isolation. Fourth, complex systems do not follow any pre-determined causal design where certain root causes determine certain outcomes that can be discovered through an analytic method.

Conflict analysis thus has to be an on-going process of exploration and self-critical analysis, informed by an awareness of our inability to fully understand the complex systems we are dealing with, and an awareness of the fluidity and change in the system. Most current pre-mission analysis are informed by short field visits that generate a once-off report, on the basis of which a mission is planned, and both the analysis, and plan, is only revisited periodically, e.g. annually as part of a new budget planning and reporting cycle. In the UN peacekeeping context, the original plan is partly locked-in because of the multi-year assessed Results-Based Budget (RBB) cycle, and it requires considerable political and bureaucratic will to make any fundamental changes to it.

Conflict analyses of complex systems thus need to be informed by our inability to understand the system in all its complexity, and should be guided by an approach that takes account of the highly dynamic, non-linear and context specific properties of complex systems. Any analysis would thus have to be limited in scope and relevant for a relatively short period of time, and information gathering will need to take a multi-pronged and highly adaptive approach, so that it can be open and sensitive to feedback and changes in the system and environment. A complex systems approach cautions us against processes, tools and mechanisms that generate simplified 'actionable priorities' that we tend to then blindly implement. It informs us that if a period has passed without us adjusting our analysis then we are likely to have missed something because the one thing we can predict is that the system will be constantly changing.

Design and Planning

Ricardo Wilson-Grau argues that in situations with high levels of uncertainty "intentional design tends to tie down the capacity to respond and innovate, above all when the social change or development organisation is bound to achieving those predefined results in order to demonstrate success to its stakeholders, notably donors" (Wilson-Grau 2008, 2). He argues that the "alternative to full-fledged intentional design, as well as to the more conventional modes of strategic planning, is for the social change or development organisation to keep its planning process light and imaginative…do not be concerned about precisely what changes you expect to see…invest the time and energy you save, in monitoring the outcomes to which you have contributed" (Wilson-Grau 2008, 2). Similarly, Harry Jones argues that "recognising uncertainty heightens the importance of building flexibility into projects, and adapting to the available signals about performance and progress as you go along" (Jones 2008, 3).

The insight we have gained from the application of complexity thus far, suggest that it is not possible to definitively identify a problem and design a solution at the outset, i.e. in the initial planning phase, using an analytical problem solving methodology. Instead we have to use an alternative planning methodology, what I will broadly term a continuous process of operational experimentation and adaptation. "A pragmatic and resilient policy that makes allowance for change events is the only evolutionary stable strategy for survival" (Rihani 2002, 83).

Design and planning cannot be limited to the start of a peacebuilding mission or programme, but need to be part of an on-going process of adaptation, throughout the lifetime of the operation or programme. The planning process should involve the broadest possible representative group of agents in the peacebuilding system, so that it can be informed by the widest possible cross-section of information. It should be as distributed as possible, meaning that central planning should be limited to broad strategic direction, so that the various agents can interpret and apply that direction as they know best in their respective contexts. The design process should generate multiple options, and the planning process should experiment with those that it thinks may be most likely to have the desired effect. As discussed earlier, this is essentially an evolutionary process of variation and selection.

However, we need to understand this as a continuous process—the situation will keep changing, and therefore the interventions also need continuous adjustments. We need to be open and willing to abandon old solutions, including those that worked well, when adapting them does no longer seem to generate the desired effect. Planning should thus not be seen as identifying a path to and end, and sticking to it, but rather as a continuous process that is aimed at helping the peacebuilding system to adapt to its environment, in pursuance of its peace consolidation objective. The process will generate milestones, such as an annual planning document, and intermediate reviews or updates, but the really important aspect is the continuous process of engagement by as many peacebuilding agents as possible in the dialectic process of making sense, together, of the how the system functions, how it can be influenced, how it is responding to earlier attempts to influence it, how it is changing, etc. An analytical methodology aims at generating solutions; a complex systems approach generates processes for continuously managing our adaptation within complex systems.

We will only know if a specific approach is having the intended effect on the basis of the feedback we receive, and this implies that our methodology has to designed in such a way that we are able to monitor and process the feedback generated by the systems we are trying to influence and the environment (Wilson-Grau 2008, 2). In the next section we will consider the implications of complexity for monitoring and evaluation.

Assessments and Metrics

In a linear system it is possible to explain an outcome in terms of a sequence of causes, for example A caused B, and that B caused C. Complex systems are non-linear. "Small changes can cause, through feedback and effects multiplying rather than just adding, very large changes elsewhere in the system. When effects are multiplicative rather than additive, it is not convincing to attribute one change to a single other change. The richness of interconnections means that any one change has several prior causes and itself may contribute to further changes in these causes" (Chapman 2002, 43).

As pointed out earlier, Morin (2005) stressed that emergence is indeductible from the qualities of the parts, and thus irreducible. This is why are unable to attribute causation when studying interventions and effects, and why considering contribution—how certain activities contributed to a certain effect—is a more humble and realistic alternative in complex systems. Morin also considers the role of feedback in what he describes as a circular system where the effect itself intervenes in the cause, or in other words "feedback is a process which complexifies causality" (Morin 2005, 15). In complex systems the behaviour of the system is influenced by its own internal processes and external causes, and its internal system have evolved as a result of its particular history.

Traditional results based ideas of accountability and responsibility are associated with a simple linear theory of causation (Chapman 2002). In order to monitor whether progress is being made, it is regarded as essential to provide measures of performance and targets. In complex systems the pursuit of any single target is likely to distort the operation of the system and thereby reduce its overall effectiveness. "One of the significant dangers of specifying targets and simple measures of performance is that the result will be sub-optimisation. Emphasising a single measure of performance leads to a decrease in overall performance. A specific target can encapsulate only one element of a complex organisation, and its dominance is likely to undermine other aspects of the organisation that are crucial to its general and long-term effects. For complex systems the only effective judge of performance is the end-user" (Chapman 2002, 46).

Chapman's point is that in order for us to measure progress, we need to monitor the overall effect of the system, rather than monitor specific goals that we have set for ourselves. For instance, our theory of change may be that if we

invest in increasing the number of police officers, we will reduce crime. If we monitor the number of police officers we may report progress as we achieve the milestones we have set for ourselves, but Chapman is arguing that if we want to measure progress in this case we have to monitor the overall changes in the level of crime. The causal relation between increasing the number of police and crime is our hypothesis.

There are a large number of other factors that may influence the levels of crime. The overall effect is best measured from the perspective of the society we wish to benefit, and in this example, the feedback that will be generated from measuring the level of crime as perceived by the community, as opposed to the level of crime as monitored by the police, is likely to reveal interesting differences. Police statistics are usually incident and report related, and are thus likely to be limited to what the police monitoring system has been designed to see.

However, a community based feedback system may reveal not only what the system anticipated to find, but it may also reveal how the system is adapting. For instance, increased policing may drive crime away from certain areas but into others. This kind of 'undirected' feedback is thus also useful to show how our interventions result in costs being externalised, for instance shifts from one part of the system to another. The point is that complex outcomes is best monitored by looking at overall system effects, from the perspective of the end-user or, in the peacebuilding context, from the perspective of the communities that are intended to benefit from peace consolidation. This insight has important implications for the type of monitoring and evaluation systems we should design in complex systems, and the kind of feedback processes we should use to inform our decision-making processes.

Feedback is critical for adaptation, and monitoring for feedback and effects should thus be a very important part of the peacebuilding process. Current peacebuilding practice neglects monitoring and evaluation, and where it does take place it is usually programme specific, and output and outcome focussed. There have been very few attempts to monitor and/or evaluate the system-wide effects of peacebuilding interventions to date. This is one of the areas that will require significant investment in the future.

Management and Coordination

As complex systems are dynamic, our methodology for generating knowledge about the system and the environment needs to have the ability to keep up with the rate of change in the environment. A management or command approach that is sensitive to the temporary nature of the approaches we are experimenting with, and to the need to continuously adapt our decisions based on new information, is more likely to cope with complex systems.

Management or command in this context refers to directing the experimentation, assessment and adaptation process. By contrast, a management style that is based on the belief that the manager or the organisation has the knowledge and/or agency necessary to correctly identify the problem and to develop the 'right' solution for the problem, and that it is somehow noble to stick to that decision and 'stay the course,' regardless of the feedback, is likely to be less effective when dealing with complex systems. According to Cilliers following such a management strategy constitutes "an avoidance of complexity" (Cilliers 1998, 112). Rihani agrees and argues that "command-and-control methods are useless. Complex adaptive systems respond better to light-touch styles of management based on constant monitoring of overall patterns of performance coupled with judicious small-scale incremental adjustments" (Rihani 2002, 93).

Real time decisions will always be a compromise between the urgency of the action, the information available at the time, and the time available and opportunity for pursuing coherence. In this highly dynamic and non-linear environment, most policy decisions are taken in the absence of crucial information, which prevents informed policy decisions, and in the presence of major uncertainties about the potential cost, benefits and risks involved in certain courses of action (Cloete 2004). Being aware of the limits of our knowledge in this context is very useful, because it reminds us not to overestimate our ability to analyse a conflict, identify root causes, and prescribe medium to long-term solutions.

Our study of complex systems have found that distributed, decentralised control makes a system more flexible, and therefore increases its ability to cope with highly dynamic and non-linear changes in its environment and within the system itself. If we apply this notion to the management of complex social systems, it would argue against an over-concentration on centralised and top-down control. The critique against centralised management or rigid leadership

is based on the knowledge that centralised control mechanism will, in complex systems, lead to the degeneration of the system (Cilliers 1998). "Command-and-control methods and detailed forecasts and plans, effective for linear systems, are inappropriate as it is not possible to select sensible actions…where results cannot be traced back to specific causes" (Rihani 2002, 9).

Unintended Consequences

The autonomous, self-organising character of the peacebuilding system does not ensure automatic, guaranteed positive impact (Cloete 2004). Although some may accept that political systems are complex, even fewer seem ready to recognise that many of their outcomes are the unintended consequences of complex interactions (Jervis 1997). You do not need a complex system to have unintended consequences, but when you are dealing with complex systems, unintended consequences should be understood as a natural outcome of the dynamics of such a system and the phenomenon should therefore not come as a surprise (Aoi, De Coning, and Thakur 2007).

Jervis goes on to point out that in a system, the chains of consequences extend over time and many areas and they are always multiple. He argues that disturbing a system will produce several changes (Jervis 1997). As the result of the interconnectedness of systems one can never merely do one thing. These kind of unintended consequences occur frequently in the context of the international response to conflicts and humanitarian emergencies, where the introduction of a new element to a highly interconnected complex system generates unintended consequences. Aoi, De Coning, and Thakur (2007) provide a number of examples in their Unintended Consequences of Peacekeeping. They point, for instance, to the problems that are created when the effects of incentives cannot be limited to a specific target group, for instance when the level of assistance to a refugee or internally displaced camp creates tension with the surrounding host community because the availability of food inside the camp is better than outside, or when ex-combatants are perceived to benefit more from the international community, through a disarmament, demobilisation and reintegration (DDR) programme, than those that did not participate in the violence.

An awareness of complexity informs us that it is not possible to interfere in a complex system like a human community, and have only one effect. Whenever

we attempt to change something in a complex system, the system responds to our intervention in a number of ways. We can anticipate that the system will respond in some of these ways, and some of these responses will have been the intended response that we wanted to elicit.

The system, however, is also likely to respond in other ways that we did not anticipate (Aoi, De Coning, and Thakur 2007). This does not mean that we are powerless in the face of unpredictable and unstable system effects. On the contrary, an improved understanding of the dynamics of complex systems should improve our ability to anticipate that there will be unpredictable and unintended consequences, and this should enable us to be more sensitive to such consequences when they occur, and to take steps to mitigate their effects or to adjust our actions accordingly.

Those responsible for the planning, management and assessment of peace-building missions thus need to recognise that unintended consequences are a normal consequence of the complex dynamic nature of complex systems. Aoi, De Coning and Thakur conclude that the UN and other institutions that undertake peace operations need to develop institutional mechanisms for addressing unintended consequences, and should institutionalise planning and assessment mechanisms that will enable it to anticipate and respond to emerging unintended consequences.

Conclusion

In this chapter we explored the implications of a complexity perspective on peacebuilding assessments. We started off by introducing complexity, and we specifically looked at three of the core characteristics of complexity, namely a whole-of-systems approach, non-linearity and self-organisation. In the process we also touched on key concepts such as feedback and emergence.

We then considered some of the implications of complexity for peacebuilding policy and practice. The most fundamental implication of complexity for how we understand and approach peacebuilding is probably the realisation that the ability of external agents to gain knowledge of the complex social systems we are dealing with in the peacebuilding context is inherently limited. In other words, we need to recognise that international actors do not have the agency to analyse a conflict, design a solution, and apply that solution with a reasonable likelihood that such an externally designed intervention can result in sustainable peace.

First, in complex systems, there is no one definitive problem that can be solved. Second, for a peace process to be sustainable, any complex social system will need to develop its own institutions to manage its own conflicts peacefully, and for that to happen it needs enough space and time to allow its own self-organising processes to emerge. International peacebuilders can assist and facilitate this process, but doing too much will undermine and delay the self-organising process. The key to successful peacebuilding thus lies in finding the appropriate balance between international support and local self-organisation, and this will differ from context to context.

The implications for peacebuilding practice is derived from our understanding of how complex systems function. Most of the technical models we rely on for conflict analysis, planning, management and assessments, are based on linear cause and effect assumptions that do not fit with our experiences and knowledge of how complex social systems function. In this chapter we have suggested a number of alternative approaches for analysis, planning, management and assessments that address the non-linear and highly dynamic nature of the social complex systems that we are attempting to influence when we undertake peacebuilding missions.

The chapter concluded by considering the inevitability of unintended consequences when attempting to influence any complex system. When we

intervene in such systems they will respond in several ways. Some of these we may anticipate, but a complex system, because it is non-linear and dynamic, will also respond in ways that we can not anticipate. We now know that this will be the case, so NATO and other institutions that undertake peace support operations need to develop institutional mechanisms for addressing the inevitable unintended consequences of its interventions, and should institutionalise planning and assessment mechanisms that will enable it to anticipate and respond to such emerging unintended consequences.

The core finding of this chapter is that international peacebuilding interventions should not interfere in complex social systems with the goal of engineering specific pre-determined outcomes, such as trying to produce a neoliberal state. Trying to control the outcome produce the opposite of what peacebuilding aims to achieve; it generates on-going instability and dependence, and it undermines self-sustainability. A complexity informed approach to peacebuilding should be about safeguarding, stimulating, facilitating and creating the space for societies to develop robust and resilient capacities for self-organisation.

References

Aoi, C., C.H. De Coning, and R. Thakur, eds. 2007. *The Unintended Consequences of Peacekeeping Operations*. Tokyo: United Nations University Press.

Byrne, D. 1998. *Complexity Theory and the Social Sciences: An Introduction.* London: Routledge.

Chapman, J. 2002. *System Failure*. London: Demos.

Cilliers, P. 1998. *Complexity and Postmodernism: Understanding Complex Systems*. London: Routledge.

——— 2001. Boundaries, hierarchies and networks in complex systems. *International Journal of Innovation Management* 5(2):135-147.

——— 2002. Why we cannot know complex things completely. *Emergence* 4(1/2):77-84.

Clemens, W.C. Jr. 2001. Complexity theory as a tool for understanding and coping with ethnic conflict and development issues in post-Soviet Eurasia. *International Journal of Peace Studies* 6(2):1-16.

Cloete, F. 2004. *Chaos and Quantum Complexity Approaches to Public Management: Insights from the New Sciences*. 2004 Conference of the Association of South African Schools and Departments of Public Administration and Management, Pretoria, 13-14 May.

Jones, H. 2008. Outcome mapping and complexity. Outcome Mapping Learning Community, Newsletter 2.

Hendrick, D. 2009. *Complexity Theory and Conflict Transformation: An Exploration of Potential and Implications.* Working Paper 17. Bradford: Department of Peace Studies, University of Bradford.

Jervis, R. 1997. *System Effects: Complexity in Political and Social Life*. Princeton: Princeton University Press.

Kiehl, D. 1995. Chaos theory and disaster response management: lessons for managing periods of extreme instability. In Koehler G. A., ed. *What disaster management can learn from chaos theory?* Conference Proceedings, May 18-19. <http://www.library.ca.gov/crb/96/05/over_12.html>

Mitchell, M. 2009. *Complexity*: A Guided Tour. New York: Oxford University Press.

Mitleton-Kelly, E. 2003. Ten principles of complexity & enabling infrastructures. In Mitleton-Kelly, E., ed. *Complex systems and evolutionary perspectives of organisations: the application of complexity theory to organisations.* London: Elsevier.

Morel, B., and R. Ramanujam 1999. Through the looking glass of complexity: the dynamics of organisations and adaptive and evolving systems. *Organisation Science* 10(3):278-293.

Morin, E. 2005. Restricted complexity, general complexity. Presented at the Colloquium "Intelligence de la complexité: épistémologie et pragmatique," Cerisy-La-Salle, France, June 26th. Translated from French by Carlos Gershenson. Also available as a chapter in Gershenson C., D. Aerts, and B. Edmonds. 2005. *Worldviews, Science and Us.* Liverpool: University of Liverpool.

Ramalingam, B., and H. Jones. 2008. *Exploring the Science of Complexity Ideas and Implications for Development and Humanitarian Efforts.* Working Paper 285. London: Overseas Development Institute.

Rihani, S. 2002. *Complex Systems Theory and Development Practice.* London: Zed Books.

Wilson-Grau, R. 2008. When is intentionally designing your work not useful? Outcome Mapping Learning Community. Newsletter 2.

11

Operations Assessment and Planning for Transition Stages

James N. Bexfield

Abstract

This chapter provides insights on the conduct of assessments that support decisions when NATO is disengaging from a counterinsurgency/stability operation. It is based largely on a NATO study that supported transition planning in the security, governance, and economic areas in Afghanistan in 2010/11. The chapter begins by providing some background on the Afghan planning process and study, reviews the methods used to develop metrics and concludes with some insights and suggestions for those organising for and/or conducting assessments. Since many of the ideas and principles contained in this chapter apply to the practice of conducting operations assessments in general, it may prove useful to the entire NATO assessment community.

Introduction

Many of the other chapters in this book have focused on the best ways to perform operation assessments to support an on-going stability operation. Toward the completion of an operation there is a period when NATO is withdrawing forces and transitioning leadership and control of operations supporting economic development, governance, and security to the host government. When to transition various regions of the country will depend on many factors, including the capability of local security forces, the capability of the local government and economic conditions in the region.

This chapter is concerned with the conduct of operation assessments that support transition decisions. It is largely based on activities that took place during the early planning stages for NATO's withdrawal from Afghanistan in the Summer and Fall of 2010. Many of these activities were part of a major NATO Research and Technology Organisation (RTO) System Analysis and Studies (SAS) Panel Specialist Team study (hereafter referred to as *SAS-091*)

that was documented in *NATO Operations Assessment: A Case Study Based on Planning for Transition in Afghanistan* (NATO 2011a). This study was commissioned in May 2010 at the request of the International Security Assistance Force (ISAF) to initially address the collection and management of data to support operation assessments in Afghanistan. It was expanded to include metrics for the transition in early July 2010.

This chapter begins with background on the Afghan transition planning process followed by a section describing some major activities conducted by the NATO study to support this planning. The third section contains some of the metric results emanating from these activities. Sections four and five provide an overview of two major products generated as part of the RTO study that apply to operation assessments in general. The chapter concludes with some observations and suggestions.

Background: Joint Framework for *Inteqal*

NATO and the international community (IC) convened over the past few years a series of conferences with the Government of the Islamic Republic of Afghanistan (GIRoA) to coordinate mutual activities. At the Kabul International Conference in July 2010, the GIRoA and the IC committed to working together to further strengthen Afghan ownership and leadership across all the functions of government. While transition from NATO to GIRoA leadership was briefly mentioned at the January 2010 London Conference, the Kabul Conference produced significant planning documents. The major document it produced was the *Joint Framework for Inteqal(1)* (Kabul International Conference on Afghanistan 2010a) that aimed at providing guidelines for the transition of responsibility. The *Kabul Conference Communiqué* was issued after the conference endorsed the *Inteqal* paper and stated:

> Within the framework of Afghan sovereignty participants endorsed the Afghan Government's plan, developed in coordination with the North Atlantic Treaty Organisation (NATO), based on mutually-agreed criteria and phased transition to full Afghan responsibility for security, as set out in the technical Inteqal (transition) paper, and endorsed a

1. *Inteqal* is the Dari word for "transition."

> decision-making process of the Government of Afghanistan and the North Atlantic Council (NAC). (Kabul International Conference on Afghanistan 2010b, 7)

The *Joint Framework for Inteqal* contains general transition policy and guidance in the areas of security, governance, and socio-economic development. It created the Joint Afghan NATO *Inteqal* Board (JANIB), co-chaired by the GIRoA transition lead, the NATO Senior Civilian Representative (SCR), and the International Security Assistance Force (ISAF) Commander, to assess the readiness of Afghan locales (provinces, districts, or municipalities) for transition.

The transition policy consists of two phases: phase 1 involves recommending locales for transition informed by assessments performed by members of GIRoA, ISAF, and NATO and endorsed by the JANIB. If GIRoA accepts the recommendation, then transition begins in the selected locale. Phase 2 guides the locale selected for transition through a four-stage process, culminating with "full Afghan ownership of security, strengthened governance and more effective development" (Kabul International Conference on Afghanistan 2010a, 5) with sustaining support from the international community, where appropriate.

This continued support was reaffirmed at the November 2010 Lisbon Conference, where NATO stated an intent to maintain a long term-relationship with GIRoA—beyond the 2014 deadline when GIRoA accepts responsibility for security throughout Afghanistan.

The *Joint Framework for Inteqal* contained specific guidance by phase. For phase 1 *security* the framework states that:

> ANSF capability is sufficient to assume, with ISAF support, lead security responsibility to contain residual and potential insurgent threats and ensure adequate levels of public security and support. […] Assessments will include: the state of the local insurgency, violence trends, freedom of movement and security of the populace; ANSF operational effectiveness and institutional capacity; […] provincial/national command and control relationships; […] (Kabul International Conference on Afghanistan 2010a, 3)

For phase 1 *development* it states that:

> Assessments will include foundations for attracting private sector investment; local engagement in development initiative; development programs are aligned with national priorities and provide a reasonable attention to all development levels; […] (Kabul International Conference on Afghanistan 2010a, 4)

Phase 2 puts an emphasis on "thinning out," vice "handing off." It involves an area progressing through four phases with each stage building off the successes of the previous stages. The conditions set during stage 4 ("sustain") are the culmination of all the prior stages and should result in a condition that is as close to irreversible as practicality allows. The four stages are:

- Stage 1: Civilian lead for Provisional Reconstruction Teams (PRTs), and smaller Afghan National Security Force (ANSF)/ISAF partnering ratio; increased IC funding commitments for building skilled labour force
- Stage 2: Local government provides adequate justice and public services; ISAF focus is operational mentoring and liaison teams (OMLTs)
- Stage 3: Local institutions enhance service delivery capacity; IC funding channels through central budget. Less OMLTs
- Stage 4: Afghanistan sustains security, government and development.

Criteria for each stage in phase 2 are specified by area (security, and governance and development). For example, under stage 3 (governance and development), the *Joint Framework for Inteqal* states:

> Local institutions exhibit sufficient technical capacity to plan, design, implement, and monitor a more comprehensive range of service delivery, including rule of law, while ensuring adequate accountability and transparency. […] International financial support occurs primarily through the core budget and existing national programmes […] (Kabul International Conference on Afghanistan 2010a, 6)

This is some of the material NATO had available for *SAS-091* (NATO 2011a) when it began its deliberation in the Summer of 2010.

A Brief Description of the NATO *SAS-091*

NATO *SAS-091* produced many useful insights regarding the conduct of operational assessments. The overview of the study in this section provides both some useful insights for those producing metrics for operations assessments and provides context for some of the insights appearing later in the chapter. It used focused workshops to bring together the right expertise to address the major issues related to the development of the transition policy. More specifically, two workshops were used to connect NATO scientists with International Security Assistance Force (ISAF) operators, NATO civilians, and members of GIRoA. Approximately 175 people attended the first workshop at Joint Force Command Brunssum (JFCBS) in the Netherlands (August 29 – September 3, 2010). The second Workshop was hosted by NATO Consultation, Command and Control Agency (NC3A) at The Hague, Netherlands (December 5-9, 2010), with approximately 115 participants, including 15 from GIRoA. Both workshops focused on developing metrics in the areas of security, governance, rule of law, and socio-economic development to support decisions associated with the transition in Afghanistan and included sessions focused on data collection and data sharing.

The first day of both workshops contained introductions, information briefs, and guidance. The remainder of the week was organised around six syndicates (working groups). Four of the syndicates focused on the *Inteqal* issue areas: security, governance, rule of law, and socio-economic development.[2] The remaining two syndicates examined issues associated with data collection and data sharing. At the request of the JFCBS sponsor, all of the syndicates expanded the data inventory by including polling results, demographic information, development projects, and other activities and measures in their deliberations. The last day of each workshop was devoted to briefs from the syndicate co-chairs on their findings and recommendations. An integrated presentation on workshop results was briefed to senior officials in JFCBS (the NATO Headquarters for operations in Afghanistan) and ISAF, the senior sponsors of the study, within ten days of the conclusion of the workshop.

2. Many similar efforts, including the *Joint Inteqal Framework*, include rule of law (RoL) under governance. Due to the unique RoL problems in Afghanistan, the study leads felt RoL deserved special emphasis.

At the beginning of the second workshop the GIRoA participants provided a list of what they referred to as indicators which they defined as "key criteria for initiating" (NATO 2011a, 3.3, 3.9, 3.22) transition. During the workshop the syndicates refined the draft on JANIB indicators and suggested enhancements to the process and procedures for collecting and sharing data. Four GIRoA representatives presented keynote presentations that helped clarify key issues and provide guidance for the security, rule of law, governance and socio-economic development syndicates.

Figure 11.1 below illustrates the general process that individual syndicates used to develop metrics as well as to identify potential data sources to feed those metrics. While the ability to set thresholds for a metric was considered by the syndicates during metric development, suggesting a specific threshold value for individual metrics was viewed as a political decision and hence outside the scope of the study.

Developing Metrics

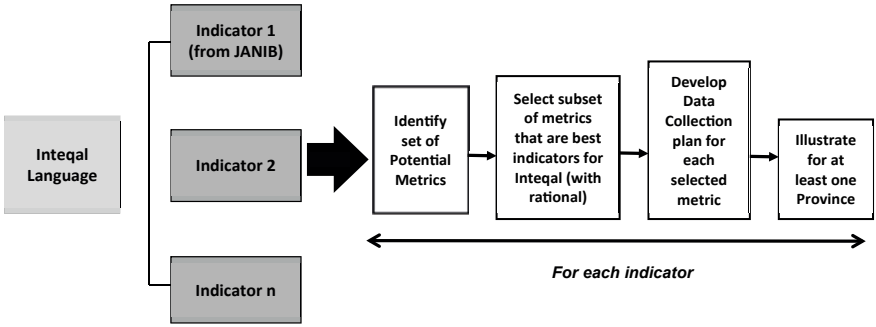


Figure 11.1: NATO study process for developing metrics

This section provides an overview of the kinds of metrics needed to assess a regions readiness for transition in two areas: governance and socio-economic development.[3] The approach begins by identifying potential measurable quantities for the indicators developed via the JANIB process. The collaborative processes used at the workshops enabled all parties to better understand

3. NATO *SAS-091* contains more details and includes metrics in the security and RoL areas.

each other's viewpoints, thus enhancing the final products. This was especially important regarding the insights provided and gained by the host nation participants. However, it should be noted in hindsight that the order in which regions should transition is a very "political" decision made at the highest levels in the Afghan government. It is not clear how much of the planning for transition described below was actually used in making decisions. This does not argue against the planning function as the learning that occurs indirectly impacts many future decisions.

Governance

The phase 1 condition in *Inteqal* is:

> There is adequate public confidence in provincial and district government to deliver basic rule of law, public administration, and essential services. Access to provincial institutions and resources is fully inclusive. (Kabul International Conference on Afghanistan 2010a, 2-3)

Assessing the readiness of locales to transition with respect to governance and RoL conditions is particularly challenging. Basic governance problems in Afghanistan include the absence of channels to express needs, bottlenecks in the delivery of services, poor reach and unclear relationships between sub-national and national entities, subversion of public finance rules, and neglect of human rights. In addition, as of July 2011, there were no Afghan government institutions present in approximately 100 of the country's 401 districts. The governance indicators for phase 1 (when to begin transition) were:

G.1 Provincial governor's office capWable of basic administration
G.2 Key line departments present, active and capable of delivering basic services
G.3 Functioning provincial council
G.4 Plans for essential government administration facilities
G.5 Available mechanisms for reintegration and reconciliation
G.6 Access mechanisms for legal and traditional dispute resolution
G.7 Available complaints mechanisms

The syndicate focused on identifying ways to measure the current state of these indicators. As an example, they suggested the following for G.2:

- Provincial and district development plans established at the local level, shaped by provincial councils and credible informal actors, and approved in Kabul
- Percentage of *tashkil*[4] filled commensurate with the ability to deliver services
- Disbursement of funding to key line departments by their respective ministries in accordance with Provincial and District Development Plans
- Provision of basic services in accordance with Provincial and District Development Plans

In addition, the syndicate stressed the importance of the ability of the governor's office to communicate vertically and horizontally, linking the population to the national government and coordinating across the provincial line ministries. Training plans should be in place at the start of transition and, as phase 2 progresses, the percentage of the staff that has received training should be tracked. Merit-based appointments of district governors and other key officials both improve productivity and enhance public confidence. Finally, the formulation of credible provincial strategic plans and district development plans are critical. Spoiler metrics[5] in this area are the level of corruption and criminality of governors and key officials and lack of plans for essential government administrative facilities.

They also suggested several data sources including ISAF, Provincial Reconstruction Teams (PRTs), United Nations Assistance Mission – Afghanistan (UNAMA), and GIRoA, especially the Independent Directorate for Local Governance (IDLG). In some cases qualitative evaluations may be the preferred

4. An organisational document which dictates force structure, personnel and strength, command relationships, and unit/staff functions and mission descriptions for the Afghan National Police.

5. A metric that demonstrates a necessary condition for transition has not been met. These metrics serve as show-stoppers if they reach certain levels. Attempting to transition the government lead to the host Nation under such conditions will likely result in effectively transferring control of that province to the insurgents instead of to the host Nation. A good approach to identifying spoilers is to consider the reasons critics use to prove the campaign is failing. A spoiler metric for security could be a lack of trust in the local police and army units. For rule of law, a spoiler might be evidence of significant intimidation (kidnapping of judges) or presence of a Taliban shadow court in a district.

method. Despite these numerous potential data sources, obtaining reliable governance data in most areas of the country remains a significant challenge.

Socio-economic development

Development encompasses much more than just the economy. It also includes the development of schools, medical capacity, and infrastructure for household utilities. It is about developing social capacity that includes issues such as equality of women in the workforce. An example of the inclusion of social issues in the development section of the *Joint Framework for Inteqal* is:

> Development priorities and principles will continue to include creating social and economic opportunities, promoting the principle of equity, demonstrating that the GIRoA works honestly for its people, providing people with opportunities to participate in their own development, ensuring representation of all, and addressing social inequities. (Kabul International Conference on Afghanistan 2010a, 3)

The syndicate did note that development is an area where GIRoA may have a competitive advantage over the insurgents. Taliban offer the population dispute resolution but few other essential services. The Afghan government can undertake and complete major development projects at the sub-national level, some taking months or years to complete. Properly focused, these projects can have a significant positive impact on peoples' lives and their perception of the local and national government. The development indicators provided via the JANIB process were:

D.1 Access to employment opportunities
D.2 Access to market, basic public services, and utilities
D.3 Agriculture land brought into licit production
D.4 Access to education for both boys and girls
D.5 Conditions in place to enable District and Community development programs
D.6 Access to higher education and/or vocational training courses/ skills building

The syndicate strived for a comprehensive and compact set of enduring metrics that do not demand sophisticated resources to collect. For instance, for indicator D2 they suggested the following metrics:

- % of households and business within urban areas with functioning electric grid
- % of villages with roads that connect district centres to provincial capital (Ministry of Rural Rehabilitation and Development (MRRD) rural access program standard)
- % of population's food requirements satisfied by Afghanistan agricultural efforts
- Level of access to and supply of food and its nutritional value
- Proportion of population rated nutritionally poor (minimum nutritional consumption)
- % of population with access to adequate sanitation and potable water
- Number of families without permanent access to shelter
- Infants (0-12 months) mortalities; Child (13-60months) mortalities; Mortalities of all above 60 months in age
- Number and type of fully functioning health care facilities
- % of population with access to basic health care package (by time to reach nearest facility)
- % of infrastructure preparation to objectives (i.e., drought and flood)

Concerning data for these phase 1 metrics, the syndicate suggested the UN, various Non-Governmental Organisations (NGOs), and international institutions such as the World Bank as potential sources. They regarded GIRoA data sources such as the Central Statistics Office (CSO) and line ministries (especially Provincial Development Plans held by the MMRD) were the most comprehensive. They also suggested the use of polling data in some cases, while stressing the need to identify alternative methods as polling is expensive and the results can be fragile. One limitation is direct knowledge of data accuracy but, coupled with details of project priorities assembled by Community Development Councils and, recorded in District Development Plans, this data should be sufficient to allow transition planning.

The syndicate felt that additional details were needed when developing metrics for phase 2. They identified desired conditions and then developed

both metrics and sub-metrics. For example, under D.1 ("access to employment opportunities") they identified the following conditions:

- Regulatory environment attracts new businesses
- Adequate access to capital
- Access to business enabling commodities
- Adequate labour pool

Then for each condition they developed one or more metrics. For example, for "adequate access to capital" they identified two metrics:

- Ability to acquire a licit loan
- Amount of loans (capital) provided to provincial entities
- Finally, they developed sub-metrics for "ability to acquire a licit loan":
- Number and type of licit financial institutions in the province
- Change in total amount of loans disbursed by all licensed financial institutions (NATO 2011a, K13)

Discovering Data

It is usually more cost effective to identify and use existing data sources than it is to develop new collection programs. In addition to being less expensive to collect, existing data may be able to provide historical perspectives usually not obtainable with new collection programs. There are several reliable military sources of security data available to NATO, but the availability of military developed data in other areas important to transition is often lacking. There are multiple sources of governance and socio-economic development data in NATO engagement areas such as Afghanistan, but how does one learn about them and assess their creditability? Many data collections are known and accessible only to their "owners," the ones who gathered the data.

The challenge is to make these data sources discoverable and accessible so that the data could be shared more broadly to support local and national assessments. NATO could commission a study to identify these sources and evaluate their accuracy, but even with a focused effort, it would take an inordinate amount of time to identify all these sources, catalogue them, and evaluate their utility. There needs to be a self-sustaining capability consisting of a community of interested parties of analysts and data owners seeking to increase the awareness of the data they are generating or using in their studies.

This was recognised in the NATO *SAS-091* mentioned earlier, spawning an initiative called "DataCards" (NATO 2011a, J1-5) a mean to identify and catalogue in an easily searchable formant as many of the existing data sources (both government and non-government) as possible.

The goals of the DataCards project were to 1) make sources of data more readily discoverable and accessible; 2) reduce the costs associated with searches for data; 3) provide a means for the operator to access non-government data; and 4) enable the resulting capability to be easily sustained by a community of interest. It uses a structured wiki concept that enables data collectors to quickly provide metadata (a "card") about their data (a summary description of the content, quality, intended purpose, and potential uses). Another goal of the developers was to make it as easy as possible to populate the wiki with new data sources.

The DataCards wiki enables anyone with knowledge of a data source to contribute new information on that source. The contributor can either fully or partially create a new DataCard, or add to an existing DataCard. It typically takes about 5 minutes for the data owner to complete a card. The tool includes a search function that will allow the user to find out if a data source has already been entered. When possible, there is a hyperlink to the data. It also contains a keyword based search feature to allow data users to quickly identify potential data sources.

This concept provides the best characteristics of a searchable database without a restrictive interface, thereby enabling many organisations to share information on their sources. DataCards is being maintained by the National Defense University (NDU) in Washington DC. At the beginning of 2013 the web-hosted catalogue contains over 1700 data sources (cards) with U.S Defense department network can access a web-enabled version.[6] A more open version of DataCards is available that uses username and password controls to validate users, so that anyone with a need can access the resource.[7] In addition, NDU distributes an Excel spreadsheet containing the cards on a periodic basis.[8]

6. See https://datacards.osd.mil

7. See http://www.datacards.org; there is also a group on Linkedin (group-digests@linkedin.com) that discusses current data issues.

8. For email distribution contact Dr. Brian Efird at brian@DataCards.org

In summary, the DataCards catalogue tool allows anyone needing specific data to rapidly sort through a comprehensive list of sources of data to find data that might be relevant to their work and quickly obtain an initial assessment as to the suitability of the data for their purposes. The database initially focused on Afghanistan and has recently been expanded to include other social science databases of potential interest to NATO. In addition NDU is in the process of instituting a quality evaluation system. The DataCards project has the potential for improving data sharing beyond the assessments world in NATO. The concept and technology are available for others to use that have similar data sharing goals.

Best Practices Guide for Conducting Operation Assessments for Counter-Insurgencies

During the course of the NATO *SAS-091* many challenges common to most operational assessments were encountered, including vague, confusing, and occasionally contradictory guidance documents inadequately defined terms, and data sources in the non-kinetic non-traditional military engagement areas that were hard to find, and when discovered often were unreliable or poorly matched with measurement needs. Debates raged over the value of subjective judgments versus quantitative data and methods of aggregating data varied in quality and consistency.

As a result, NATO *SAS-091* produced a guide with a wide set of best practices for dealing with these challenges that can be applied to any assessment. Its purpose is to complement, not replace, the more detailed planning or instructional documents contained in official instructions and field manuals. Instead, the guide attempts to close the gap between the ideal and the reality of assessment by providing insights into the "philosophy" of assessment, highlighting the challenges, and sharing best practices from the field.

This section contains a brief description of the 20 articles in the guide.[9] The articles in the guide are broadly categorized as either assessment tenets (Part One) or assessment methods (Part Two). The assessment tenets seek to clarify the assessment's purpose and objectives. By reminding practitioners of how their assessments can be used to influence an overall campaign strategy, it

9. For the full version of the guide, see NATO 2011a, Annex I.

becomes easier to make the right choices between sources and methods. It also helps practitioners understand how to build and communicate an assessment to influence strategic decisions. Part Two is more tactical in nature, with some of the most common assessment pitfalls, reminders of some fundamentals, and suggestions for dealing with intransigent players or intractable obstacles.

Part One: Assessment Tenets

Article one: Remain true to the assessment's objective. The objective of an assessment is to produce insights pertaining to the current situation, and to provide feedback to support decisions by the leadership. This article discusses how key elements of this objective should guide the assessment development process.

Article two: Take a multi-dimensional perspective. This article describes why it is essential to build the assessment by looking at the environment through multiple perspectives that cross lines of operations and time periods. It also highlights some errors that may arise if the assessment lacks a broad perspective.

Article three: Serve as the bodyguard of truth. Assessment teams develop what may, by default, become the only publicly available, official picture of the campaign. Therefore, assessment teams must serve as the bodyguard of truth and never compromise the integrity of their reports. This article outlines nine key practices that help preserve the integrity of assessments.

Article four: Ensure independence and access. Assessment teams need access to a wide array of information and people in order to perform their job properly. This article describes how to secure independence and access through a partnership between the senior sponsor of the assessment team and individual line of operation owners.

Article five: Nurture the Knowledge Development (Intelligence) – Assessment Partnership. The activities related to knowledge management and assessments often seem remarkably similar, thus generating the potential for confusion or duplication of effort. This article discusses how to build a mutually supporting relationship between the two activities.

Part Two: Assessment Methods

Article six: Establish a terms of reference document. Unclear terms generate confusion in the design of the assessment framework, the analysis of data, and the reporting of insights. Thus, it is in the team's best interests to develop a Terms of Reference document that includes an assessment design and a data collection strategy as soon as possible.

Article seven: Build the assessment framework iteratively, incrementally, and interactively. The assessment framework should be built in stages through a collaborative process to minimize complexity, allow for effective learning, and retain clearly established priorities. It also allows the assessment team to refine the focus and scope of the assessment framework based on lessons learned.

Article eight: Discriminate between indicators and metrics. Most people use the terms "indicator" and "metric" interchangeably and normally suffer little or no consequences or confusion. However, there are times when it is useful to discriminate between the two. This article offers a useful approach regarding when and how to discriminate.

Article nine: Use each class of indicator properly. Grouping indicators into classes that share a common set of characteristics may be beneficial or detrimental to the assessment process. Several of these broad classes are described in this article, including those that measure input versus outcome, those that indicate failure to achieve a condition (spoilers), metrics that can indicate positive or negative effects depending on context (bipolar), and those that serve as substitutes for other hard-to-measure indicators (proxies).

Article ten: Beware of manipulated metrics. Some metrics can be manipulated by the subjects under observation to send misleading signals to observers, rather than reflecting the reality of the current conditions. This is a particularly high risk for metrics used to promote or demote, or directly redistribute resources. This article contains several examples and suggests ways to detect and minimize such distortions.

Article eleven: Develop a manageable set of metrics. There are hundreds of metrics available at any point in time. Thus, it is necessary to establish rules to help select the metrics contributing the most to the operation. This article discusses several screening filters that help practitioners develop a manageable and effective set of metrics.

Article twelve: Retain balance in both metrics and method. Debates on the merits of the narrative report versus summary graphics, the organisational level at which assessments should be performed, and how to preserve the front-line commander's views within higher level summary assessment products persist in the assessment world. This article describes how to achieve a balanced blend of each alternative that captures the best features of each.

Article thirteen: Use field assessment teams. In order to provide actionable information to the decision-maker, assessment insights must be relevant and credible. For critical issues, the only way to achieve this standard is to directly interview front-line units in the field. This article offers an approach that augments the traditional process with the use of field assessment teams dispatched from appropriate levels.

Article fourteen: Use eclectic marginal analysis to bound estimates. When a desired metric is difficult to measure directly, it might be possible to measure the factors that drive the value of the same metric. Under such conditions, marginal analysis can be used with an eclectic set of related metrics to generate a reasonable estimate of the target metric. This section explains the technique and provides some examples of marginal analysis.

Article fifteen: Anchor subjectivity. A degree of subjectivity in assessments is unavoidable. This article discusses methods to minimize the degree of subjectivity, make that subjectivity transparent, and maintain consistency in the way subjective assessments are captured.

Article sixteen: Share data. Every coalition effort faces information sharing challenges. This article discusses important reasons for sharing information and offers some guidelines that promote effective sharing.

Article seventeen: Include host nation data. This article addresses the challenges of using host Nation data and ways to work around the challenges, including host nation data collection systems and the ability of assessment teams to interact with this system.

Article eighteen: Develop metric thresholds properly. This article discusses key guidelines for developing metrics thresholds, including when to adjust levels, how to develop clear definitions of the thresholds, and how to determine when observances of metrics represent a significant change in underlying conditions.

Article nineteen: Avoid substituting anecdotes for analysis. Anecdotes are a useful component of assessments when used properly. Unfortunately, they are

often used as substitutes for a solid assessment. The best rule to keep in mind when using anecdotes is that they are generally the starting point for analysis, not the closing argument of an assessment.

Article twenty: Use survey data effectively. Questions about motivation, satisfaction, degrees of trust or fear, as well as intentions regarding future actions are difficult to measure by monitoring physical activities. Often, this information must be captured by interviews or broader surveys. This article addresses how to manage some of the major concerns associated with using survey data in assessments.

Some Observations for Those Conducting Assessments

This section provides some insights that may prove useful for those conducting assessments supporting transitions from NATO led to host government led activities. It is largely based on the work done by the NATO study team to support transition planning in Afghanistan.

Flexibility

The initial planning documents in Afghanistan had transitions occurring by province with the challenge being to determine when provinces were ready to begin transition. In practice, transitions did not occur exclusively by province; instead, districts in a province or even municipalities transitioned separately. In other words, plan transition by areas which could be as large as a province or as small as a city. In addition, the metrics selected may depend on the entity being transitioned. In other words, transition metrics for a province may differ from those used by a district or city.

Applicability of metrics across phases

Many of the phase 1 metrics on when to transition may be applicable to phase 2 metrics measuring the progress the region is making toward achieving full transition of leadership.

Dependence on adjacent areas

The success of a transition in one region may depend on the state of adjacent regions; an adjacent region that is insecure may significantly impact the neighbouring region's chance for success for transition. Similarly, insecurity on the border could affect a neighbouring region's stability.

Impact of metrics on behaviour

When generating metrics, the developer and leadership should take into account the impact metrics may have on the behaviour of affected organisations and individuals. This includes many dimensions: internal to NATO, host nation, and insurgency.

Interdependence of metrics across areas

Metrics for governance, RoL, security, and socio-economic development are interconnected, each depending on the other for validity. Each can be assessed independently, but a holistic approach is required for a complete assessment of readiness for transition. For example, socio-economic development does not occur in a vacuum—e.g., security is needed for people to feel safe to travel to markets, governing regulations are needed to enhance trade. Thus, measuring the number of newly formed markets may be an indicator of enhanced security, a growing economy, and an active governing body providing supportive regulations. Similarly, understanding how corruption levels impact economic growth may result in identifying a useful metric that covers both areas. How to incorporate these interdependences is a challenge for those developing transition metrics.

Some Suggestions for Organisations Conducting Assessments

We conclude with some advice for those leading assessments teams in support of transition. Most of the suggestions apply to assessment teams in general. Again, much of the material is based on the NATO *SAS-091* (NATO 2011a).

Build an assessment team with a mix of capabilities

Since assessments require a multi-dimensional perspective, the assessment team should include experts from all lines of operations (LOOs) and assessment disciplines. Military officers may be a good source for security

expertise. Individuals with experiences in diplomatic missions usually have the expertise and background needed for governance and RoL. Social scientists from development agencies, such as USAID, can help guide the search for socio-economic development metrics and data sources. These functional experts support the analyst who ensures that appropriate standards and rigor are part of the metric development process and the assessors who report directly to the decision-maker and manage the process. In the end, it is all the members of the assessment team, working together, that integrate and synchronize the resulting products for leadership's use in decision-making.

Establish a common lexicon and insist everyone use it

Transition assessments typically involve many people from a variety of organisations, all with their own assessment terminology and definitions. Much time can be wasted if a common lexicon is not established early in the process.

Improve the processes used to share data among the stakeholders

When conducting assessments it is much easier to use existing data than it is to develop and implement a new data collection process. Reviewing and suggesting changes to policies and procedures for sharing data from the top (requirements identification phase) to the bottom (data collection phase) should be an early assessment team activity.

Develop sustainable data collection processes and monitoring capability

In cases where data do not already exist, the assessment team may need to establish a data collection process using assets in the field. There needs to be a balance between overburdening a unit with too many requirements to collect data and collecting the assessor's most critical data. When units are overburdened data collection will lapse, errors will be more pronounced, and reports will be delayed or not provided at all. Using simple, standardized templates may also help lighten the burden of collecting data and support the sharing of data. NGOs or International Organisations (IOs) cannot be expected to provide data that does not provide them utility. Frequent monitoring will help maintain discipline in collection, thus improving data quality.

Involve the leadership early in the process and keep them informed

Including leadership from the requesting organisation in the planning process should help ensure guidance and feedback is available at each critical stage of development. It will also help maximize the customer's ability to integrate the resulting assessment work into their own decisions and activities.

Identify and involve all key players, including those outside of NATO

It is important to get the perspective from all participants in an operation, which often involves engaging people outside of NATO. Examples include USAID, host governments, embassies, NGOs, etc.

Consider using a workshop approach to capture expertise

A workshop can be an effective way of bringing together operational, regional/cultural, and technical/analytical perspectives and expertise. The resulting face-to-face discussions can lead to a common understanding of a complex, dynamic problem space. The workshop approach is most appropriate when team members are separated geographically and there is a desire to bring in outside expertise, with periodic meetings used to facilitate coordination and product development. Initial workshops may be led by outside experts as the additional people they bring with them may suggest innovative new approaches that could increase the efficiency and effectiveness of the assessment process.

Use spoiler metrics that identify conditions that would disqualify a locale for transition

Spoiler metrics can make it easier to spot the problem by looking for proof/evidence that a condition does not exist instead of looking for proof/evidence that it does exist. There is no need to review other metrics if a key condition for transition is not met. For example, if there is no host government justice institution (e.g., functioning prosecutors and judges) in the district, then there is a rule of law vacuum that anti-government forces can exploit, making transition reversible. These districts are not ready to begin transition.

Adapt to changing conditions

Conditions are constantly changing in any NATO operation, which will likely change the metrics desired by the leadership. In complex operations many organisations may be involved, each with their own objectives and plans. Any one of these organisations could insert new guidance into the assessment process. If the assessment team does not consider this new guidance, its products may become irrelevant.

Don't over apply lessons learned from previous transitions

For example, Afghanistan is quite different from Iraq in terms of human capacity, literacy, authority and legitimacy of the central government, infrastructure, culture, and the role of customs in the everyday life of the people. Apply lessons from other transitions cautiously.

Help build host nation assessment capability

This may be in the form of including them in planning meetings, workshops, and data collection activities. Holding courses that train host nation assessors can both help the host nation manage better manage their affairs after NATO departs and ensure that there are good sources of data available for post departure assessments, especially in the governance and economic development areas.

Conclusion

Operation assessments should be, and often are, major contributors to operational and strategic decisions associated with NATO activities. The better the assessments, the more informed the decisions. One primary goal of this chapter is to help educate and guide NATO personnel responsible for performing operational assessments, with a focus on counter-insurgency environments that are in the planning stage for transitioning from NATO to local responsibility. It is intended as a supplement to official NATO assessment guidance, such as the *NATO Operations Assessment Handbook* (NATO 2011b). Another major goal is to point the reader to other reference material that can help guide their assessment activities such as the NATO *SAS-091* (NATO 2011a) and articles by Flynn, Pottinger, and Batchelor (2010), Ghani (2010), Ghani and Lockheart (2008), and Schroden (2009) listed in the references.

References

Flynn, M. T., M. Pottinger, and P.D. Batchelor. 2010. *Fixing Intel: A Blueprint for Making Intelligence Relevant in Afghanistan*. Washington, DC: Center for New American Security.

Ghani, A. 2010. Afghanistan opportunities and risks. *PRISM* (1)4:15-24. Washington, DC: NDU Press.

Ghani, A. and C. Lockhart. 2008. *Fixing Failed States: A Framework for Rebuilding a Fractured World*. New York, NY: Oxford University Press.

Kabul International Conference on Afghanistan. 2010a. Joint Framework for Inteqal: A Process for Strengthening Peace and Stability in Afghanistan and in the Region.

——— 2010b. Kabul Conference Communiqué. Available online: <http://www.unodc.org/documents/afghanistan/Kabul_Conference/FINAL_Kabul_Conference_Communique.pdf>

North Atlantic Treaty Organisation (NATO). 2011a. NATO Operations Assessment: A Case Study Based on Planning for Transition in Afghanistan, RTO-TR-SAS-091. Paris: NATO Research and Technology Organisation.

——— 2011b. *NATO Operations Assessment Handbook*. Interim Version 1.0. Norfolk, VA: HQ Supreme Allied Command Transformation.

Schroden, J. 2009. Measures for security in a counterinsurgency. *The Journal of Strategic Studies* 32(5):715-744.

12

Operations Assessment: Decision-Based Evidence Making?

Bruce Pennell

Abstract

Decision-based evidence making is usually seen as a pejorative term that refers to the deliberate use of certain evidence to support a decision or policy already decided. In this chapter, the varying roles evidence plays in making, informing or supporting decision-making will give operations assessment staffs better insight into how to develop effective assessments at the operational, theatre and strategic levels. At the strategic level, operations assessment has an essentially political, rather than military, characteristic. The symbolic use of evidence, the exploitation of cognitive biases and approaches from the field of policy analysis are also introduced.

Decision Making and Evidence

When the Victorian-era British General and military thinker Edward Hamley declared that tactics is "the opinion of the senior officer present," (Luvaas 1964, 153) it is usually assumed he was cynically observing that arguments forwarded by captains seldom change the strongly-held views of generals. It is also possible, however, he was reflecting sagely on the roles and responsibilities that should be properly held by those same generals. Specifically, that generals should trust their own judgement, and not be distracted by juniors who are unlikely to have the same ability to sense the essence of the situation.

Hamley's words could also be describing current challenges in operations assessment, which are also highlighted elsewhere in this volume. Specifically, how should analysts and operations assessment staffs reconcile the tensions that arise when views formed by the *senior officer present*, based mainly on gut-feel and an intuitive grasp of the campaign obtained from a privileged position, challenge what appears (to the analyst, at least) to be rational, evidence-based analysis?

Military culture still arguably idolises the intuitive commander, capable of grasping the essence of a complex situation and conducting incisive decision making (DM) under acute pressures of time and information overload. In tactical military operations it is clear that a commander who can execute quick, high quality DM through gut-feel and intuition will be at an advantage. The price of failure is clear: tactical military defeat. Above the tactical level, the advantages of intuition are less evident. The need for quick DM is less apparent because there is more time and resource available to process conflicting information. Yet current NATO doctrine claims: "Intuition, experience and military judgement remain paramount…to support commanders' decision making at the strategic and operational levels" (NATO 2010a, 1-3).

Nevertheless, it is also clear that operational and strategic DM has a strong, if not dominant, political dimension. In the NATO context that means higher level DM is heavily dependent on deliberation, social bargaining and the politics of alliances. In alliance member states the electorate have a voice too, in setting domestic agendas that will exert influence on decision stakeholders. A further challenge is that, in a comprehensive intervention (i.e., with military, political, economic, and social dimensions) it is much less clear whether a single person, even the commander, has the expertise to make a credible assessment across all dimensions. Operations assessment has become, if it was not always, a contested space.

This chapter does not, sadly, describe a new paradigm that resolves these tensions. It is instead a more personal piece based on my experience working the past ten years in various operations assessment settings. It reflects what I feel is the most salient issue; how the operations assessment community should think about evidence. However, the arguments put forward are mine, and certainly not an official view from my employer or NATO.

Use of Evidence in Policy Formulation and DM

The interaction between evidence and policy formation, and the DM process is far from straightforward, as indicated by the growing academic field of policy analysis. In an important critique of how university-level research is used and constrained by government, Boden and Epstein defined for the first time the term "policy-based evidence" (Boden and Epstein 2006, 226).

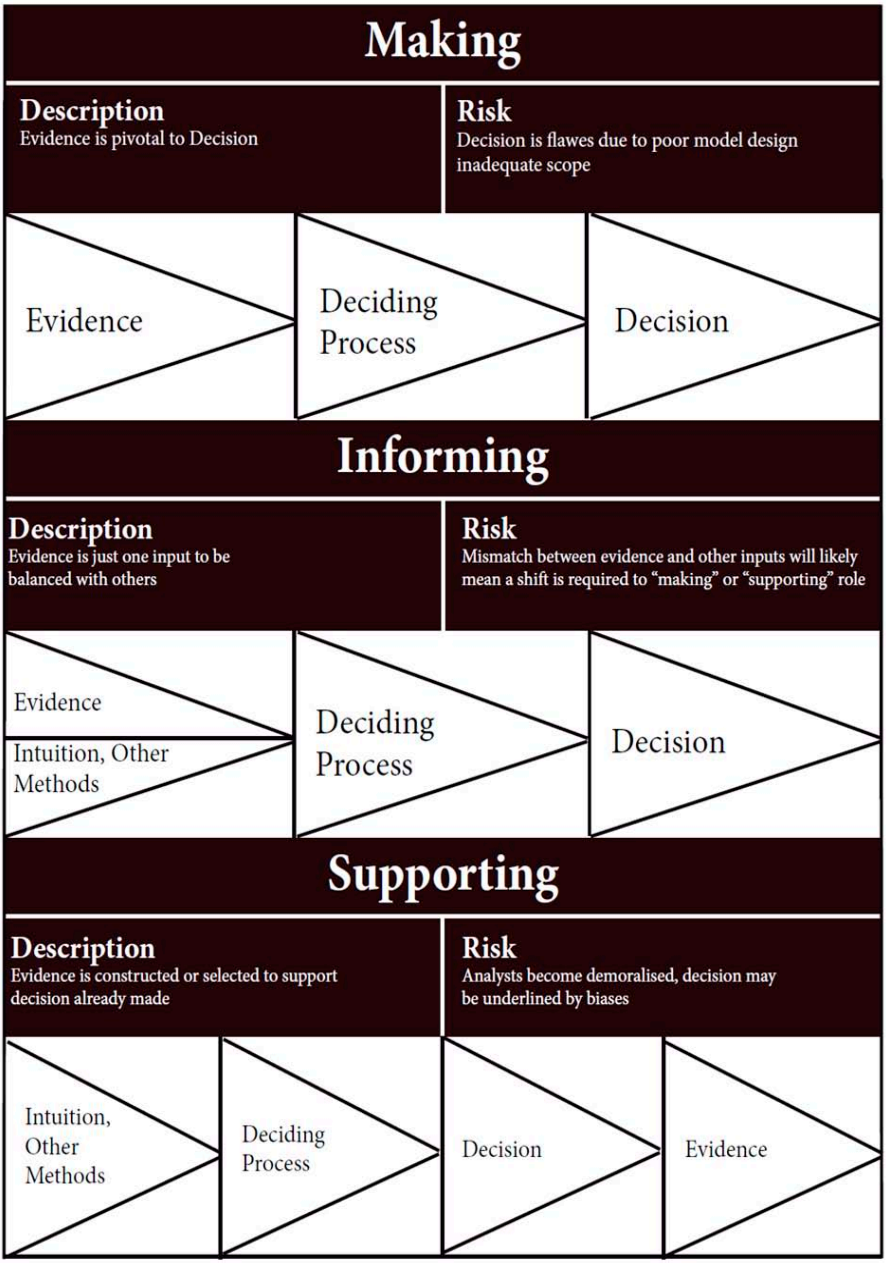


Figure 12.1: Roles for Evidence in Decision Making
(adapted from Tingling and Brydon 2010)

According to Boden and Epstein, policy based evidence is defined as evidence selected specifically to support an already made policy decision.

Their conclusion was that such a process was inherently flawed and fatally undermined the integrity of much social science research. However, a subsequent study by Tingling and Brydon more critically evaluated the phenomenon they termed "*decision-based evidence making*" (DBEM) (Tingling and Brydon 2010, paragraph 3). Tingling and Brydon used the term 'evidence' as all data and analysis brought forward to support a line of argument. They assert that contempt for disconfirming evidence—and its deliberate de-selection as described by Boden and Epstein (2006)—represents just one extreme example of the varying roles evidence plays in DM. These roles are characterised by the terms *making (deciding)*, *informing*, and *supporting* (see Figure 12.1).

Understanding these roles lies at the heart of gaining insights into the use of operations assessments in comprehensive crises.[1] Using Tingling and Brydon's typology, my personal impression is that many operations assessment staff believe that what they do should support assessment in a making or, at worst, in an informing role. Yet, especially when their evidence challenges other perceptions (especially the opinion of the *senior officer present*) they can quickly find their analysis relegated to a supporting role, or rejected entirely. A further challenge faced by analysts is that the role played by evidence does not appear to be the principal determinant of decision quality. Using evidence only in a supporting role, especially in complex situations, does not necessarily mean that decisions are badly made.

Analysing Decision Making

Tingling and Brydon do not offer a theoretical model of DM itself; they simply describe how they believe evidence is used to support it. Although a comprehensive analysis of DM theory is beyond the scope of this paper, it is useful to reflect on some broad characteristics of common DM models which I believe are pertinent to the operations assessment area. Rational approaches

1. NATO's Strategic Concept, adopted at the Lisbon Summit in November 2010, endorsed a comprehensive approach to crisis management involving political, civilian and military instruments (NATO 2010c)

to DM, which include ideas of satisficing [2] and game theory,[3] are founded on the notion that, even in the presence of uncertainty, the evidence and analysis supporting DM is inherently objective, drawing on the techno-rational scientific tradition. We can see this as consistent with the "making/deciding" role in the Tingling and Brydon typology.

The main issue with rational DM models is that, in reality, they fall short of capturing actually how DM takes place. For example, in game theoretic approaches it is possible to calculate an optimal strategy for many two-player adversarial games in terms of costs and payoffs. When observed experimentally, however, adversaries often behave differently and the strategies they implement are sub-optimal. This does not mean that rational models of DM are without value; it is just that they do not always capture important aspects of DM in reality. Returning to the Tingling and Brydon classification, it is hard to reconcile rational DM models when evidence is used in informing and supporting roles.

Stakeholder Views and Value Curves

Alternative DM models try to reflect the subjective views and biases of decision stakeholders to better reflect DM as it is actually executed. Kahneman and Tversky (1979), both economists, observed in experiments with human participants that, instead of being equally indifferent to losses or gains of equal amount, losses appeared to 'hurt' more. Moreover, participant reaction to small losses and gains was overdeveloped compared to larger results. The illustrative outcome of their experimentation was the so-called Prospect theory 'value curve' in Figure 12.2. A positive change (gain) along the horizontal axis delivers a smaller change in perceived 'value' than a negative change (loss) of the same magnitude. Additionally the steepness of the curve close to the reference position indicates the tendency to overvalue small losses or gains compared to larger ones.

2. 'Satisficing' describes the problem solving strategy of seizing the first solution to a problem that appears to offer success rather than searching for an optimal solution. This occurs as a result of human cognitive limits, or 'bounded rationality,' as humans cannot generally rationally evaluate all possible alternatives (Simon 1997).

3. Game theory is a mathematical approach to structured problems where players adopt strategies that have costs and payoffs, first elaborated by von Neumann and Morgenstern (1944).

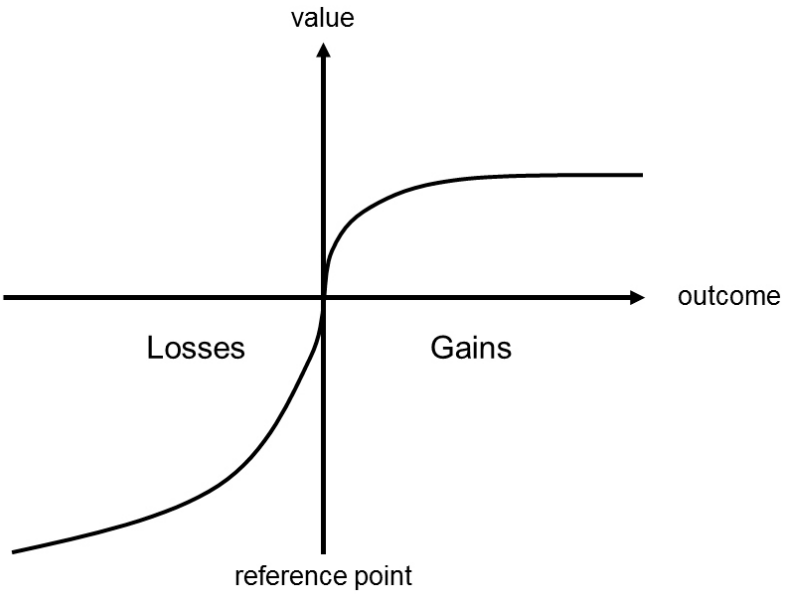


Figure 12.2: Prospect Theory Value Curve (Kahneman and Tversky 1979)

Cognitive biases are inherent, systematic mental behaviours that lead to prejudice in DM. In this case, prejudice does not necessarily mean "bad," but simply a systematic deviation from the rational. For example, Kahneman and Tversky's Prospect theory clearly exhibits a systematic asymmetry of the value of loss compared to a gain of equal magnitude. This is known as the 'loss aversion bias.'[4]

Bias is important to the analyst because it is so often seen as a failure of method; the analytical literature is threaded through with a belief that the best analysis is de-biased. This is, I believe, a cultural artefact inherited by operations assessment staffs working within a mainly rational-analytic

4. And also other biases such as 'scope insensitivity' (value perception changes more slowly at extremes of change).

tradition, with a strong influence from quantitative methods. Very often, this leads to a privileging of the quantitative over the qualitative, and a desire to apply reductionist approaches to complex problem analysis, all with the aim of eliminating bias. I believe, however, we should be very sceptical of claims that such approaches eliminate bias, and—perhaps more controversially—that analysts should instead embrace and exploit it.

The Art of Political Decision Making

None of the authors thus far cited worked in operations assessment, and it is evident the challenge of adequate models, evidence and bias is faced in diverse fields. The academic branch of political science known as policy analysis has developed a wealth of literature on the issue. Given that in higher level crisis operations DM has a significant political dimension, it is perhaps useful to look to that field for insights.

One source of inspiration is Majone's (1989) concept of trans-scientific problems, which cannot be wholly described or confronted with scientific methods alone. Majone describes how approaching such issues with decisionist approaches—aimed at developing rational, calculated choice between clear alternatives—will inevitably be flawed.

Table 12.1 describes an influential model of political DM first presented by the political scientist, Deborah Stone in 2002. Stone argued that a more authentic alternative to the rational model for political DM was what she termed the 'polis model' (Stone 2002). Stone argues that DM solutions (policy strategies) are formed with inducements, rules, rights, and powers as the driving forces, as reflected in the four main tenets in her model.

It is easy to dismiss such a model as cynical and manipulative, or merely a feature of Machiavellian politics. However, I believe it is also possible to frame the polis model more generally within Tingling and Brydon's typology. This may give operations assessment staffs a useful framework within which their efforts may be situated. Stone regards the 'supporting' role for evidence as the pre-eminent mode. Evidence is almost always symbolic, adding credibility and authenticity to a process that is much more focused on mobilising power and undermining opposition.

Stone's *Polis* Model
State goals ambiguously and keep some secret; be prepared to shift and redefine goals as the political situation dictates.
Keep undesirable alternatives off the agenda; present the preferred alternative as the only feasible one; selectively project consequences that make your decision look the best.
Focus on one part of the causal chain and ignore politically difficult ones; avoiding issues around which opponents can coalesce.
Choose the action that hurts powerful constituents the least, but portray your decision as creating the maximum social good.

Table 12.1: Stone's Polis Model of Political decision making (adapted from Stone 2002)

Propositions

The work of Majone and Stone may better reflect non-rational characteristics of policy development and DM, and offer a useful conceptual framework. Additionally, it does suggest a number of further propositions.

Proposition 1

Evidence is rarely objective. Evidence often gives the impression of being objective, incontrovertible, unbiased fact. Even when this is the case, analysts have to make decisions on what evidence to select, whether to seek out additional evidence, what methods to use to analyse it and how to resolve inconsistencies. These decisions and selections are usually implicit, introducing subjectivity, which is often overlooked in the reporting of the subsequent analysis. This is not inherently sub-optimal, but sets the reader on course for the subsequent propositions.

Proposition 2

Counting is not inherently rational. "The dominance of numbers as a mode of describing society….is only a recent, and perhaps temporary, phenomenon in

cultural history—not the result of some underlying reality of numbers" (Stone 2002, 187). Many practitioners among operations assessment staffs come from a background that subscribes to the superiority, or inherent truth, of quantitative approaches and their utility in decisionist approaches. Yet Stone and many others argue that counting is not inherently rational, and is very often itself a political act. The boundary between what is and what is not counted in a certain category may be fuzzy. Conversely, the resulting mapping onto numerical scales may give a false impression of precision. The use of numbers in general can support the illusion that the complex is simple, and that data collection is complete and comprehensive. It may also introduce ambiguity, not least through aggregation.

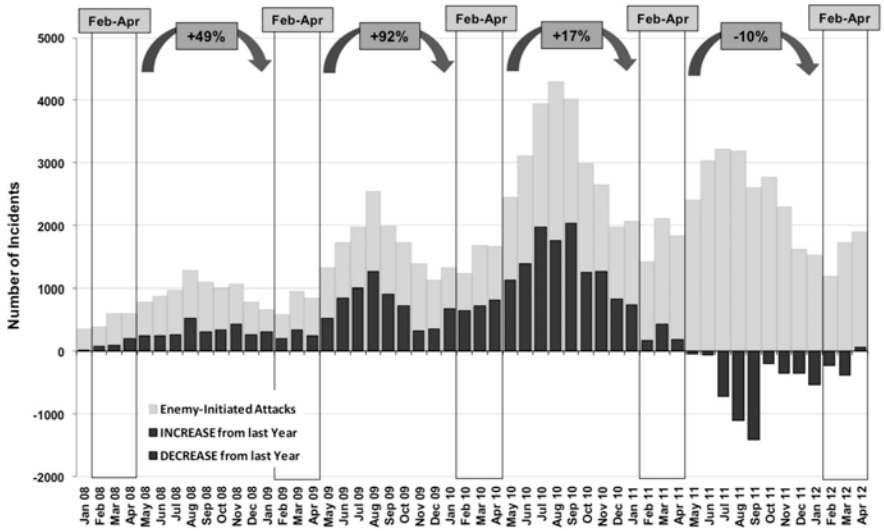


Figure 12.3: Quantitative Data Representation

The tools and techniques for manipulating quantitative data are highly developed, ranging from statistical methods, through optimisation algorithms to time series charting. By contrast, equivalent techniques for manipulating qualitative data are less well-developed in the operations assessment field. An example quantitative output is shown in Figure 12.3.

Most operations assessment staffs would probably be familiar with interpreting such a time-series chart as it ostensibly represents a more insightful piece of analysis than qualitative data such as media reports, personal observations or photographs. Certainly, the chart captures a greater range of data and an evolution in time that a single observation or photograph does not. However, the chart itself conceals the subjective decisions and assumptions that were used to build it, which may include confounding factors such as changes to categorisation rules in the underlying data, or simply the absence of relevant data. Equally, though, unless a photograph is fabricated (and, unfortunately, even sometimes when it is) it cannot always be regarded as inferior evidence, or less influential when the DM stakeholders take notice of it.

Another challenge associated with quantitative analysis occurs when there is unequal access to the underlying data, for example, for reasons of security classification. Whilst it may be insightful to make assessments based on this information, these may not have much influence on stakeholders with access to the underlying evidence. This is a particular issue in coalition operations where national caveats to data sharing may apply, or when there is suspicion (justified or not) that those with privileged access are suspected of manipulating evidence in ways which reflect best on themselves.

Proposition 3

Opinion polling is no easy shortcut to assessing attitudinal change in most conflict areas. This proposition is a corollary of Propositions 1 and 2. During recent complex crises, planners and operations assessment staff, faced with the challenge of assessing attitudinal change within populations, have made increasing use of opinion polling and other survey methods. An understanding of changing attitudes is deemed critical to assessing progress in comprehensive interventions, especially when population attitudes towards an Alliance intervention become, in themselves, key objectives.

Inaccuracy in opinion polls and surveys is well documented, however, even in stable well-developed countries, and even when questions are well-defined (e.g., 'how will you vote in the next national election?"). The chapter by Katherine Banko (2013) in this volume explores some of the specific issues surrounding opinion polling in the insecure areas of Afghanistan, including the difficulty

of finding a representative survey sample, and identifying response anomalies such as social desirability bias.[5]

Despite these shortcomings, the evidence generated by opinion polling and other survey methods during complex crises can appear very persuasive and compelling. As well as the methodological issues highlighted by Banko and others, I believe it is important that operations assessment staffs recognise this evidence is subjective (framed by the values and beliefs of the questioner) and symbolic. This does not mean it has no value, but simply that it is not inherently more truthful than other evidence.

Proposition 4

Symbolic evidence is most persuasive when it exploits relevant cognitive biases. Except in those relatively few situations when quantitative evidence can be exploited and used directly in a making/deciding role, evidence is symbolic to some degree. Arguably, its use becomes more symbolic as one moves through informing to supporting roles. As with the photographic example above, the power or weight of symbolic evidence is found in the influence it exerts on stakeholders, not from some inherent truthfulness of the evidence. This influence is determined in large part by cognitive biases.

Operations assessment staffs will therefore benefit from a better understanding of cognitive biases. In the Prospect theory/ loss aversion bias case above, for example, evidence presented in a way that emphasises the potential for loss will likely have greater influence than the same evidence expressed as a potential gain of the same magnitude.

Intuitively, operations assessment staffs already understand that the way evidence and analysis is presented makes a difference to the level of influence it has on the decision maker. The challenge for the analyst is to identify and account for the cognitive biases that are likely to be in play by assessment stakeholders.

5. Social Desirability Bias describes the phenomenon in polls and surveys where respondents answer in accordance with how they believe the questioner wishes them to respond, or other social norms.

Proposition 5

Overcoming strongly held views rarely depends on more evidence. When evidence challenges the decision or opinion of the senior officer present (which has been, or is being, reached by other means), it is a natural reaction of the analyst to provide more rational evidence of the type already presented. My experience suggests such an approach rarely achieves its aim; partially validating the Hamley quote that opened this chapter. In effect, trying to push DM stakeholders towards greater use of evidence in making roles may in fact lead to a move in the opposite direction. Faced with this problem, analysts may despair, or simply resign themselves to finding only helpful supporting evidence.

Alternatively, they may attempt to understand and analyse the cognitive biases in play and exploit them. For example, the bias of illusory correlation refers to events or data that are randomly correlated, but are interpreted—particularly at first glance—to have meaningful correlation. If an analyst suspects that illusory correlation bias has played a part in the decision being made, then it may be helpful simply to highlight the phenomenon and scrutinise the reasons why it is in play, rather than trying to strengthen a counter-argument.

Proposition 6

Effective comprehensive operations assessment must embrace the political aspect within its area of interest. Majone and Stone stress that higher level DM is essentially political and socially constructed. Above the tactical level, therefore, I believe operations assessment staffs cannot simply ignore the dominant political/social character of comprehensive interventions. The lack of well-developed tools to work in these areas is not a good enough reason to remain disengaged.

A useful analogy can be drawn from the intelligence domain. NATO intelligence doctrine (NATO 2010b) describes the difference between an area of intelligence responsibility (AIR) and a larger area (or areas) of intelligence interest (AII). The assigned unit has responsibility for providing intelligence for the smaller AIR. However, intelligence is also needed for the AII to give the appropriate context for the knowledge gained in the AIR (see Figure 12.4 below). Without an understanding of the AII the knowledge developed from the AIR alone will be flawed. The AII concept gives the assigned unit's Intel staff *permission* to exercise its analytical efforts more broadly.

I believe the same concept applies to operations assessment staff at operational or theatre levels. The broader area of assessment interest that lies beyond the area of assessment responsibility cannot be neglected. In any comprehensive intervention the area of assessment interest must incorporate the political dimension, which means that operations assessment staffs at theatre levels and above need to broaden their skills. Improved understanding of techniques used in the field of policy analysis would be a useful first step.

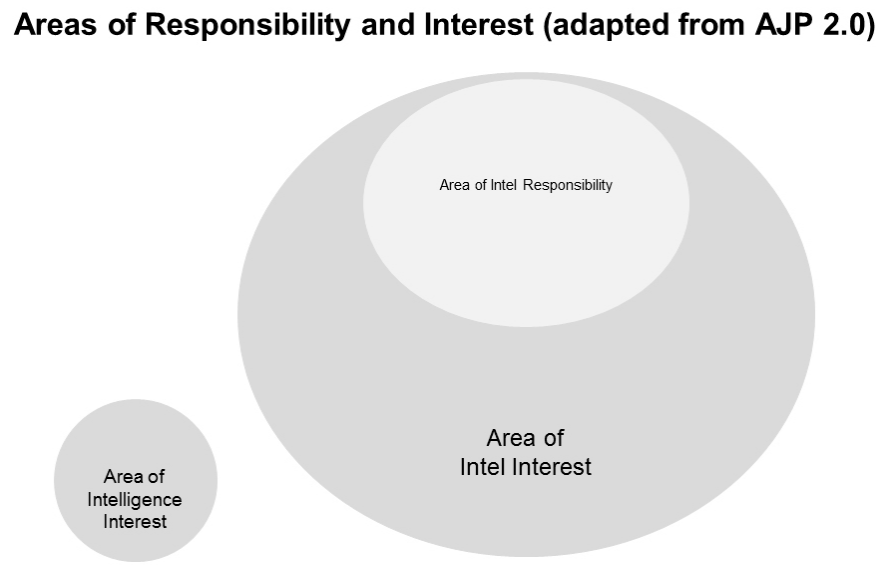


Figure 12.4: Intelligence Areas of Responsibility and Interest

Proposition 7

Systems modelling for comprehensive operations may still offer insight to assessment staffs, despite recent criticism. A complete description of systems approaches is beyond the scope of this paper. In essence, however, systems approaches to modelling views the problem space as made up of interconnected subsystems, within which causal loops and feedback takes place, allied with the idea of delay between cause and effect (Gallo 2013).

From an operations assessment perspective, systems-modelling approaches have been criticised (Connable 2012) because the data to populate formal

systems models of complex conflict is never available. Nevertheless, at the heart of the comprehensive operations dilemma is the challenge of identifying how military, political, economic, social and other interventions interact. It seems to me that, in the absence of anything demonstrably better, operations assessment staff may find it still to be a useful approach. Systems-diagramming offers the important possibility of incorporating multiple possible objectives into the assessment, which may allow greater flexibility when end-states and goals may have to change as the political climate dictates, as suggested in Stone's polis model.

The value of a systems-diagram may be realised without it being completely specified and populated with data. In this respect an understanding of the symbolic nature of evidence is helpful to the operations assessment staff; the systems-diagram may simply be a visual metaphor—"this is the problem as I see it at this time." This may help frame the assessment and therefore does not need to be completely validated by data.

Conclusion

This short chapter provides some insights into the current challenges faced when assessing comprehensive operations. These mainly arise from the difficulty in adapting or applying rational-analytic approaches to problems which are essentially socially constructed. Tensions arise when evidence, especially quantitative evidence, is relegated to decision-informing and supporting, rather than decision-making, roles. Acknowledging that 'counting is not inherently rational', understanding socially constructed DM models, and a better appreciation of relevant cognitive biases, are useful first steps for analysts responding to these challenges.

Assessments at the operational and strategic levels must incorporate the political dimension in their "area of assessment interest." This requires operations assessment staffs to develop broader skills. The academic field of political science and policy analysis specifically may yield useful tools and techniques for assessing change, particularly attitudinal change and handling other qualitative evidence. Although they are far from a panacea, systems diagrams permit accommodation of some challenging aspects of more politically influenced assessment.

References

Banko, K. 2013. Public opinion polling in operations assessment. In Williams A., J. Bexfield, F. F. Farina, and J. de Nijs. eds. *Innovation in Operations Assessment: Recent Developments in Measuring Results in Conflict Environments*. Norfolk, VA: NATO Supreme Allied Commander Transformation.

Boden, R., and D. Epstein. 2006. Managing the research imagination? Globalisation and research in higher education. *Globalisation, Societies and Education* 4(2):223-236.

Connable, B. 2012. *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*. Santa Monica, CA: RAND Corporation.

Gallo, G. 2013. Conflict theory, complexity and systems approach. *Systems Research and Behavioral Science* 30:156-175.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263-291.

Luvaas, J. 1964. *The Education of an Army: British Military Thought, 1815-1940.* Chicago: Chicago University Press. Quoted in I.B. Holley. 1983. Of saber charges, escort fighters, and spacecraft. *Air University Review* September-October 1983.

Majone, G. 1989. *Evidence, Argument, and Persuasion in the Policy Process*. New Haven, CT: Yale University Press.

North Atlantic Treaty Organisation (NATO). 2010a. *Allied Command Operations: Comprehensive Operations Planning Directive*. Interim Version 1.0. edited by Supreme Headquarters Allied Powers Europe. Mons, Belgium: NATO.

——— 2010b. Allied Joint Doctrine for Intelligence, Counter-Intelligence and Security [AJP-2.0]. Brussels, Belgium: NATO Standardization Agency.

——— 2010c. Strategic Concept For the Defence and Security of The Members of the North Atlantic Treaty Organisation: Adopted by Heads of State and Government in Lisbon. The North Atlantic Treaty Organisation.

Neumann, J. von, and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Simon, H. A. 1997. *Administrative Behavior: A Study of Decision-making Processes in Administrative Organisations*. 4th ed. New York, NY: Free Press.

Stone, D. 2002. *Policy Paradox: The Art of Political Decision-making*. New York, NY: Norton.

Tingling, P. M., and M. J. Brydon. 2010. Is decision-based evidence making necessarily bad? *MIT Sloan Management Review* Magazine: Summer 2010. Available online <http://sloanreview.mit.edu/article/is-decision-based-evidence-making-necessarily-bad/>

Weiss, C. H. 1979. The many meanings of research utilisation. *Public Administration Review* 39(5):426-431.

White, L. G. 1994. Policy analysis as discourse. *Journal of Policy Analysis and Management*, 13(3):506-525.

13

Everyday Peace Indicators: An Alternative Form of Assessment

Professor Roger Mac Ginty

Abstract

This chapter provides insights into why we measure things the way that we do. It summarises many of the criticisms made against existing attempts to measure peace and transition and seeks to understand why these methods persist. It then proposes an alternative approach to measuring peace and transition called Everyday Peace Indicators. These indicators are developed from the bottom-up and identified by local communities.

Introduction

Most indicators of transitions from conflict to peace and peace operations are top down. That is, the indicators of change are usually identified by personnel in an international organisation, a national government or an international non-governmental organisation (INGO) and are then rolled out into communities or applied to particular projects or operations.[6] The funding, design, and interpretation of the data are usually conducted centrally, with little reference to local communities. To the extent that local individuals and communities are involved, they are usually survey enumerators, translators or mere data.

There are, of course, many good reasons why survey, monitoring and evaluation (M&E), and operations assessment processes are top-down. Yet, it is worth asking if we can amend current practice, or complement it with new practice, in order to increase the effectiveness of the knowledge that we accumulate. In theory, better information should result in better policy and so

6. For classic 'top down' indicators see the World Bank's World Development Report or the United Nations Development Programme's Human Development Index, both published annually and often based on aggregates of nationally-gathered statistics.

it makes sense to reconsider the ways we gather information, and the extent to which our current practices can accurately gauge local opinions.

This chapter is written from a peace studies tradition[1] and so it may not immediately resonate with those working within military organisations and associated institutions. But it is offered in the spirit of debate and in the understanding that different paradigms have much to learn from each other. The chapter begins by critiquing current approaches to indicators and surveys of change. In its second section it asks the question: why do we organise our surveys and assessments in the way that we do? The answer often lies in organisational inertia and technocracy. The third section outlines a proposal for bottom-up or alternative indicators of peace and change. It relies on individuals and local communities to identify indicators of change rather than using a list of indicators developed by outsiders.

Before beginning the chapter proper, let's start with an anecdote. The author's father owned a store in a small town in Northern Ireland in the 1970s. The town centre was regularly bombed and the shop window was repeatedly smashed. There came a time when my father could no longer afford to replace the glass, and insurance companies would no longer give cover. So my father, and the other storekeepers in the street, replaced their windows with wood. This situation persisted for a number of years. Then, after a few years, my father and the other storekeepers in the street had the confidence to replace the wood with glass. They had picked up, on the grapevine or just by reading the everyday evidence in front of them that the chances of town centre bombs had diminished and so it was worth investing in glass again.

From the point of view of this chapter, the interesting aspect of the story is that orthodox, top-down indicators of conflict and change probably would not have picked up on this everyday aspect of transformation and greater confidence in the community. The authorities in Northern Ireland doubtless had a range of indicators at their disposal. The policing and military authorities

1. Peace Studies takes peace as its starting point. Political Science and International Relations, on the other hand, take—respectively—the state and conflict as their starting points. Peace studies are based on humanity's ability to work through problems and are often normative (in being actively pro-peace). It tends towards conflict transformation or the acknowledgement that the underlying causes of conflict must be addressed. These include the 'structural violence' of discrimination and the belief systems that maintain them. See Smoker 1981.

would have kept track of the number and severity of shooting and bombing incidents. The government would have had a range of statistics at its disposal, many of which were routinely collected: unemployment figures, economic trends, public expenditure by sector etc. What these statistics were unable to collect, however, was the everyday opinion of people on the ground and how these opinions translated into action. What was it that made my father, and the other storekeepers, have the confidence to invest again in glass?

This chapter seeks to make the case for everyday indicators that are genuinely bottom-up. This involves a fundamental rethinking of how we collect and disseminate information. It also relies on a humility (that does not come naturally to 'experts') or recognition that we do not have all of the answers and that we need to reconsider our ability to ask the right questions. Patrick Chabal's recent work *The End of Conceit* (2012) seems to have summed up the problem rather nicely: there is a conceit among western 'experts' (many of them academics) who believe that our way of thinking and processing information is superior to that used by others. As Chabal notes "… we Westerners need to accept that [our] theories might have reached their limits, and for this reason have now become so many obstacles to thinking" (Chabal 2012, 34).

Critiquing Current Indicators

There is a widespread acknowledgement that the current suite of indicators and assessment tools available to international organisations, governments, INGOs and donors and donors are deficient. As a result, there are a number of initiatives, by the G7+ group of states, the US State Department, and INGOs such as the Catholic Relief Services and the Alliance for Peacebuilding, to experiment with new ways of gathering information in conflict-settings (Kawano-Chiu 2011; International Dialogue 2012; Catholic Relief Services 2010). Four criticisms are commonly made of orthodox approaches that measure change in relation to peace and transitions. Not all of these criticisms apply to all cases, and it is worth remembering that this chapter is written from the perspective of assessing peacebuilding and development interventions.

Perspective too narrow

A first criticism is that many indicators of change are linked to specific projects, programmes and operations rather than the wider context in which

the project, programme or operation might function. There are, of course, good reasons for the focus on project-level assessments. At a minimum, the funder will want to know if the project or operation's aims have been met, on time, and on budget. A military commander will want to know the progress towards meeting specific objectives. This is a basic auditing requirement and it fulfils an important role. However, by concentrating on the project, programme or operation there is a danger that the terms of 'success' or 'failure' are set too narrowly, that they don't consider the wider picture of the economic, political and social trends in the society at large. It is possible to have a successful project or operation in a context that is, overall, suffering from a deteriorating humanitarian situation.

Consider, for example, a civil war situation which is demonstrably deteriorating. It is possible for INGOs or international organisations to run successful projects in such a context. The project might involve training human rights monitors or instituting governance reforms within a branch of government. Those who run the project can report success to their donors that the project aims have been met: number of human rights monitors trained over a prescribed period or number of governance regulations introduced in a ministry. But such an approach is dangerously like measuring the 'burn rate' of a project without considering what the project achieved in changing. The essential problem is that the project or operation level is often much too narrow. It reports a partial picture that might be very separate from the actual conditions on the ground. Problems may arise if those involved in the monitoring are tempted to extrapolate beyond what the specific project might tell them.

Perspective too general

A second criticism of many current indicators of peace or transition is that they use national-level statistics. These have many uses and allow for cross-national comparison. But this level of analysis is also quite unwieldy. Conflicts and transitions never occur with uniformity across a national territory. The diplomatic district of a capital city might be an oasis of calm. In some provinces, however, there may be a serious security situation. Moreover, government and effective administration may be absent from these conflict-affected districts and so the gathering of statistics may be interrupted or non-existent. The danger is that national-level statistics mask provincial and local-level differences.

Conflicts, and transitions away from conflict, are primarily experienced at the local level: individually, through the family, household, workplace and peer network. National-level statistics cannot represent the nuanced and highly localised nature of such experiences.

Inappropriate use of proxies

The inappropriate use of proxies arises because peace is incredibly difficult to define: it is usually wholly subjective and accompanied by a bewildering array of caveats. Many measures of social and economic phenomenon have an agreed benchmark against which measurements can be made. The most obvious example is currency: there is general agreement what a dollar is worth and this is translated into more or less universal statistics: 2.44bn people live on less than US$2 a day or a barrel of oil will cost US$80 (Economist 2012). There is no agreed 'currency' for peace, with the result that proxies are often used. For example, since civil war is usually associated with economic deterioration, it is tempting to use a decreasing GDP as an indicator of the proclivity for the onset of war (Collier and Hoeffler 2004). But the use of GDP as a proxy for peace may not be appropriate as there may be many reasons for changes in GDP. Part of the reason GDP and similar statistics are used as proxies of peace is that they *can* be measured. Such statistics are available and are usually dutifully collected by national governments and international financial institutions. But the fact remains; they are rarely good indicators of peace.

Too much emphasis on top-down approaches

Top-down approaches for identifying indicators usually do not account for the actual lived experience of those living through a conflict or transition. As discussed earlier, indicators and assessments are frequently designed, funded and directed from outside of the conflict-affected area and results are typically not shared with the research subjects. Indeed the word 'subject' is worth dwelling upon.

Many research processes include a power relation in which some actors have more power than others. Such research processes may reinforce the subaltern position of local actors and the superior position of external actors. Local communities and external actors might define and describe the same conflict in very different ways. For example, external actors may use the language

of 'terrorism' or 'insurgency.' Local actors may see such terms as alien, and instead define the conflict in religious, nationalist or resource-related terms. Yet, external (including international) actors often have the 'framing power' to make their version of the conflict stick. In a way, this is a process of 'subjectification,' which turns the inhabitants of a conflict-affected zone into subjects and strips them of the agency to define their own conflict.

These criticisms are not intended to underestimate the very real difficulty of measuring peace and other social phenomena. Those faced with this task often have to use the tools that they have at hand. Security and access problems, as well as deficiencies in local administration, can be difficult obstacles to overcome when collecting data in conflict areas. Moreover, local perspectives are not always relevant to monitoring and evaluation tasks. Some assessment exercises are 'self-contained' and do not require reference to local opinion. A review of an organisation's internal operating procedures may not need to refer to anyone outside of the organisation. Yet, in many other cases, local populations, 'end users,' and recipients are relevant and are not included in evaluations or included in tightly prescribed ways. The next section attempts to explain why we persist with using the wrong tools.

Why Do We Measure Things the Way We Do?

There are, of course, some eminently sensible reasons why we measure peace, conflict and transitions in the way we do.

Methodological rigor

Perhaps the main reason relates to methodological rigor. Many of the orthodox current approaches means of measurement favoured by international actors in relation to intervention are methodologically defensible. They follow prescribed patterns and allow for comparison with other cases and projects. Yet, such approaches might be *precisely wrong*. They are methodologically robust (on their own terms) but do not actually help us capture the phenomenon we are interested in (De Vries 2001).

Donor insistence

In the development sector, one reason for the persistence of orthodox ways of measuring transitions is that donors sometimes insist that conventional methodologies be used. For understandable reasons, donors (whether governments, international organisations or INGOs) demand transparency and accountability. They want to be able to see that value for money is delivered and that projects meet their stated aims. Donors are often very prescriptive about the M&E techniques to be used, and often these are traditional and quantitative. They are driven by an 'audit culture' which in turn is driven by political and financial pressures.

Convenience and inertia

People are trained in the way that it is done and come to believe it is the way that it should be done. Many organisations have their own M&E units, or can draw on external M&E consultants. There are few incentives for change. M&E has developed into something of an industry, and like all industries it has its own political economy. Often there are few reasons to bite the hand that feeds us. As a result, innovation and criticism may be muted as those charged with monitoring and evaluation simply get on with the job in hand. More fundamental reviews that ask why are we measuring a social phenomenon in the way that we do, might be a risky endeavour that endangers the political economy of service provider and customer.

Rise of technocracy

Technocracy is the bureaucratic imperative: the privileging of bureaucracy above other guiding principles such as political, moral or kin-based frameworks (Mac Ginty 2012; Centeno 1993). Of course, there is nothing new about technocracy. Great empires of the past depended on administrative competence and sophisticated systems to collect and interpret information. Opportunities for technocracy, however, are markedly increased in international peace-support operations.

Statebuilding and governance reform agendas often centre on establishing or reforming bureaucracy. Concepts and practices that were developed by large accounting conglomerates in the 1980s (benchmarking, total quality management, audit trail etc.) have become mainstreamed into many aspects

of life and these terms are now commonplace in universities or INGOs (Box 1999). While technocracy leads to the better shepherding of resources, it can also camouflage power-shifts where administrative and cost control systems assume dominant positions in organisations. Organisations that were originally established for humanitarian, security or moral purposes have become subjugated to technocratic imperatives. In many cases, this power-shift has been subtle and unquestioned.

Crucially, technocracy is also a mind-set espousing the superiority of technocratic approaches to solving problems. Technocracy is, according to some (e.g., Centeno 1993), an ideology and as such intolerant of alternative approaches that might draw on indigenous, traditional or customary ways of doing things. This is not an argument against technocracy per se. It is, however, a caution against an unthinking adoption of technocracy and its application to all aspects of life. The danger is that technocracy is insufficiently responsive to local culture and aspirations.

Given our interest in the measurement of peace and transitions, the rise of technocracy is crucial. This may result in financial administrative positions driving the measurement process with a subsequent bias towards monitoring the 'burn rate' rather than more meaningful outcomes. For example, such approaches might measure the number of government personnel trained (easy to measure) rather than the outcome of that training in terms of better service (more difficult to measure).

Technocracy often leads to closed thinking where the problem and the solution become mutually reinforcing and protected from innovative alternatives. We see this in conflict analysis when international organisations, governments and INGOs use standardized conflict analysis models (themselves a product of technocracy). These models tend to diagnose conflicts in similar ways and recommend largely similar responses involving statebuilding and good governance reforms.

In the closed loop of thinking, the conflict 'cause' (state failure or weakness) demands an obvious conflict response (statebuilding or reform). Yet, for those living through the conflict, this diagnosis of state failure or weakness might be wide of the mark. For example, the state has been mostly absent for people living in rural Democratic Republic of Congo (Kabama 2010) with potential conflict causes being tribal rivalry, resource wars, or criminality. External

actors, with their technocratic mind-set, however, may be unable to see these conflict drivers and instead concentrate on the absence of the state and the need to create or rebuild a functioning and robust bureaucracy. The chief point is that external actors are often unable to 'see' the real conflict. Technocracy has stripped them of the skills required to look beyond the default position of each society needing a functioning state.

An Alternative Way of Measuring Peace

As mentioned before, international organisations and governments are using innovative ways to augment their existing survey platforms. For example, household surveys have played a key part in the data-gathering of organisations like the World Food Program (WFP) (e.g., Hirotsugu and Dhur 2006) and the United Nations Department of Economic and Social Affairs (UNDESA 2005) for many years.

The WFP regularly surveys the price of bundles of firewood in rural Kenya as a way of measuring the cost of cooking fuel and as a way of gauging household income. Surveying the price of bundles of firewood can indeed be revealing. But the accessibility and value of firewood will change in from locality to locality. In some areas, it will be plentiful, or augmented by animal dung or another fuel source, or culinary traditions necessitate a different rate of fuel usage. The chief point is that local knowledge needs to be factored into the choice of indicator and the design of the survey.

Along with colleagues, the author of this chapter has developed a methodology for Everyday Peace Indictors[2] (EPI) (Mac Ginty 2013). Crucially, these bottom-up indicators rely on individuals and communities in localities to identify their own indicators of peace or transition, rather than having them imposed by external international actors. This marks a significant departure from standard research practices and entails a number of methodological challenges. The EPI methodology[3] employs participatory action research

2. See http://everydaypeaceindicators.org/

3. The Everyday Peace Indicators methodology will be tested in a project that begins in July 2013 and will operate in four African states: South Africa, South Sudan, Uganda and Zimbabwe. It is funded by the Carnegie Corporation of New York, and undertaken by the author with colleagues from the Kroc Institute of International Peace Studies and the Institute for Justice and Reconciliation in South Africa.

techniques implemented in conjunction with local NGOs (Krimerman 2001) and is influenced by innovative projects in sustainable development (Parkins, Stedman, and Varghese 2001; Reed, Dougill, and Baler 2008).

EPI is based on the premise that those who have lived through a conflict or a transition are often the best placed to make judgements on change. They are most likely to notice the nuances that outsiders might miss unless equipped with ethnographic skills (something that requires time, resources, access and security). The key notion is that conflict and transition are experienced locally: in the family, neighbourhood, village and valley. Thus EPI operates at the local level and can serve to offer more nuanced information than that proved by national-level surveys.

The EPI methodology consists of seven stages. In the first stage, local NGOs or consultants identify suitable survey locations. This depends on prior NGO links with communities and on security and access. In the second stage, NGOs operate focus groups among community members to identify indicators of change and peace. This is where EPI deviates from standard measures of peace and transition. Individuals and communities in the focus groups are able to suggest any indicators that they want. So, for example, they may mention the adoption of stray dogs (a sign of a more steady food supply), a better mail service (a sign of a more effective state), or the painting of storefronts in towns (a sign of increased business confidence).

Of course, focus group participants may suggest standard indicators used by INGOs and international organisations or by political leaders. This is the great unknown of the EPI. But the very act of identifying local indicators is a useful exercise in that it reveals issues that are important to community members. In the third stage of EPI, local NGOs collate the indicators into a manageable list and then turn them into a questionnaire. Then, in the fourth stage, the questionnaire is run among the wider community by the NGOs who use traditional face-to-face methods and a very simple mobile phone application. In the fifth stage the survey results are collected and reported back to the community. In the sixth stage, they are translated and fed back to the research project. In the seventh stage the exercise is repeated after an interval of six months to track changes.

There are a number of methodological problems that face the EPI approach, and doubtless more will become apparent as the methodology gets to the field.

A first issue is that of gatekeeping. All societies have gatekeepers, or individuals and institutions that want to control information. In societies affected by violence or division, these gatekeepers are usually linked to a militant or political project and are anxious that a single narrative is reproduced. Such gatekeepers are usually wary of alternative narratives that might dissent from their preferred narrative. EPI will have to overcome this problem and hope that individuals in focus groups are comfortable enough to identify indicators that are meaningful to them and their everyday lives, rather than identify indicators that they think they should use.

A second issue relates to the indicators that participants identify. EPI does not intend to romanticise or depoliticise individuals and communities. It may be the case that survey respondents do not identify local concerns and indicators (such as the adoption of stray dogs, the resumption of a mail service, or an increase in the number of tourists in the area). Instead, they may identify more national-level and political concerns. If this is the case, then so be it. It is important that local input into the survey is as untrammelled as possible by the survey team.

A third methodological issue relates to the representativeness of the survey. EPI is explicit that it wants 'good enough' rather than perfect methodology. It recognises that it operates in fragile and unique environments that are not laboratories. Partner NGOs will endeavour to ensure that the surveys are inclusive (of genders, minorities etc.) but there are few guarantees that conditions on the ground will allow for this.

In summary, the EPI methodology offers a number of opportunities to improve on existing assessment models. Importantly, it should not be seen as a replacement for existing surveys. Instead, it is a way of augmenting them and adding greater nuance and localized information to what is already available. The primary advantage of the Everyday Peace Indicators is that the indicators are identified at the level of the neighbourhood or village. The fact that they may differ from locality to locality may limit comparison between districts. But districts *are* different and we should not expect perfect comparability. Moreover, the act of identifying locally significant indicators will be revealing about local priorities. In violence-prone societies there is often few ways for communities to identify the issues that are important to them. This role is

often monopolized by actors who seek to speak 'on behalf' of communities ('strongmen,' religious leaders, etc.).

There is a chance that the identification of the indicators by local communities could assist in conflict transformation. Unlike conflict resolution or conflict management, conflict transformation is interested in addressing the underlying issues that cause and sustain conflict (Lederach 1995). It places emphasis on education about one's self, one's group, and other groups. It is based on the idea that individuals and groups need to understand themselves and others in order to deal with conflict. The process of identifying indicators may allow communities to realise that they share certain aspirations and issues. These 'indicators +' go beyond the mere collection of data (Mac Ginty 2013) to sparking conversations within communities that can help them solve problems.

Concluding Discussion

At the centre of the issue of collecting information lie issues of epistemology and positionality. These terms relate to the observation that where we stand dictates what we see. Our current modes of thinking and gathering data are often so deeply ingrained (in educations systems, in organisations, in the language and methodologies that we use) that we are often not aware of them. Think of how we measure distance: it is usually in kilometres or miles, or perhaps in terms of time. But not all cultures measure distance in the same way.

When, in 1921-24, British expeditions were trying to find a route to, and up, Mount Everest they would ask local communities to estimate the distances for them (Davis 2012). Of course, local communities in this period had little concept of miles and kilometres, or indeed of formal time. So distances were measured in 'cups of tea' or the distance one would walk before breaking for a cup of tea. The central point is that ways of thinking that seem intuitive to us may not be obvious to others. Moreover, our ways of thinking and measuring may not be able to convey an accurate picture of the society we are interested in.

James C. Scott (1992) has identified societies as having 'public' and 'hidden transcripts'. Public transcripts consist of the easily observable evidence that comes from government ministries, press statements and what people may say in interviews. But all societies have a hidden transcript—the actions and discourses that take place behind closed doors, or when the boss is out of the room or the police patrol has passed. Accessing this hidden, or perhaps more

accurately described as a non-obvious, transcript may be crucial to understanding the needs and aspirations of a society. Allowing individuals and communities to take charge of evaluating their own experiences seems like an important step towards revealing the hidden transcript. Better information might, in turn, allow for more tailored intervention policies.

References

Box, R. C. 1999. Running government like a business: Implications for public administration theory and practice. *The American Review of Public Administration* 29(1):19-43.

Catholic Relief Services. 2010. *GAIN Peacebuilding Indicators*. Baltimore: Catholic Relief Services.

Centeno, M. A. 1993. The new Leviathan: The dynamics and limits of technocracy. *Theory and Society* 22(3):307-335.

Chabal, P. 2012. *The End of Conceit*. London: Zed Books.

Collier, P. and A. Hoeffler. 2004. Greed and grievance in civil war. *Oxford Economic Papers* 56(4):563-595.

Davis, W. 2012. *Into the Silence: The Great War, Mallory and the Conquest of Everest*. London: Vintage.

De Vries, W. F. M. 2001. Meaningful measures: Indicators on progress, progress on indicators. *International Statistical Review* 69(2):313-31.

Economist. 2012. Global poverty: A fall to cheer. *The Economist*, May 3.

Hirotsugu, A. and A. Dhur. 2006. Measuring household food insecurity in emergencies: WFP's household food consumption approach. *Humanitarian Exchange Megazine* (32). Available online <http://www.odihpn.org/humanitarian-exchange-magazine/issue-36/measuring-household-food-insecurity-in-emergencies-wfps-household-food-consumption-approach>

International Dialogue on Statebuilding and Peacebuilding Working Group on Indicators. 2012. *Progress Report on Fragility Assessments and Indicators*. Available online <http://www.newdeal4peace.org/wordpress/wp-content/uploads/progress-report-on-fa-and-indicators-en.pdf>

Kabama, P. 2010. 'Heart of darkness': Current images of DRC and their theoretical underpinning. *Anthropological Theory* 10(3):265-301.

Kawano-Chiu, M. 2011. *Starting on the Same Page: A Lessons Report from the Peacebuilding Evaluation Project*. Washington, DC: Alliance for Peacebuilding.

Krimerman, L. 2001. Participatory action research: Should social inquiry be conducted democratically? *Philosophy of Social Sciences* 31(1):60-82.

Lederach, J. P. 1995. *Preparing for Peace: Conflict Transformation Across Cultures*. Syracuse, NY: Syracuse University Press.

Mac Ginty, R. 2012. Routine peace: Technocracy and peacebuilding. *Cooperation and Conflict* 47(3):1-22.

——— 2013. Indicators +: Everyday peace indicators – a proposal. *Evaluation and Program Planning* 36:56-63.

Parkins, J., R. Stedman, and J. Varghese. 2001. Moving towards local-level indicators of sustainability in forest-based communities: A mixed-method approach. *Social Indicators Research* 56(1):43-72.

Reed, M., A. Dougill, and T. Baler. 2008. Participatory indicator development: What can ecologists and local communities learn from each other? *Ecological Applications* 18(5):1253-69.

Scott, J. C. 1992. *Domination and the Arts of Resistance: Hidden Transcripts*. Yale: Yale University Press.

Smoker, P. 1981. Small peace. *Journal of Peace Research* 18(2):149-57.

United Nations Department of Economic and Social Affairs (UNDESA) (2005). Household Sample Surveys in Developing and Transition Countries. *Studies in Methods, Series F/96*. New York, NY: United Nations. Available online < http://unstats.un.org/unsd/hhsurveys/pdf/Household_surveys.pdf>

North Atlantic Treaty Organisation
Headquarters Supreme Allied Commander Transformation
7857 Blandy Road, Suite 100
Norfolk, Virginia. U.S.A.
23511