

Designing Behavioural-Based Measures of Effect at the Community Level for COIN Operations: Challenges of Design and Framework Integration

Maj John Leahy

1st Canadian Mechanised Brigade Group Edmonton
P.O. Box 10500 Stn Forces
Edmonton Alberta T5J 4J5
CANADA

John.Leahy@forces.gc.ca

ABSTRACT

Assessing the effectiveness of counter-insurgency operations and campaigns within a comprehensive context is exceptionally difficult. Confounding variables such as insurgent presence, limited governance, and competing power groups or factions compete with moderators such as trust between groups, risk to people who support the government, and cultural differences to cloud the reliability, validity, and utility of measures of effect. Despite these difficulties, the Canadian Task Force that deployed in early 2009 attempted to build an assessment framework that provided an integrated approach to understanding and measuring the impact of the operation. The assessment framework was integrated in that it attempted to combine traditional quantitative metrics of output with qualitative metrics that described the effects in concrete definable aspects of local community behaviour, attitudes, and perceptions. Though successful in terms of methodology development and local community assessment, this approach failed to incorporate other metrics and assessment frameworks. Suggestions for future integrated assessment of effects are made.

1.0 INTRODUCTION

There is little debate that assessing the impact of Whole of Government (WoG) or comprehensive approaches to operations in counter-insurgency conflicts such as Afghanistan is difficult from a scientific and a collaborative perspective. Yet the need to collect specific and accurate metrics of assessment in these complex systems such as Afghanistan is captured, for instance, in the Canadian counter-insurgency manual, which notes that:

“With influence activities and effects on the psychological plane, [measures of effectiveness] are applied to activities and the resulting changes in understanding, perception and the will of the target audience... The results of these influence activities require as defined a set of indicators as possible in order to detect changes in perceptions, understanding, attitudes and behaviours.” (p. 6-27)

The technical issues of designing specific reliable and valid assessment metrics in complex environments are immensely complicated by the structural challenges and demands of working in a WoG Task Force. Developing an integrated framework that allows for intertwining strategic and operational effects with operational and tactical needs for assessment may not be easily applied or integrated into existing assessment frameworks (Dobias, Innis, & Caruana, 2009). In fact such a problem may resist solution by traditional approaches in which problems are defined, analysed, and solved in sequential steps; in effect, such a problem may be a so-called “wicked problem” (Rittel & Webber, 1973).

Despite the challenges implicit in multi-level effects assessment in such complex environments, the Canadian Task Force that deployed to Afghanistan in early 2009 (TF 5-09) attempted to develop just such an integrated framework. The assessment framework aimed at incorporating classic effects and metrics drawn from previous work done in theatre, while also defining, quantifying, and ultimately measuring how Canadian operations would impact the local people. In effect, the objective of this effort was to gather the range of metrics needed to assess how Canadian counter-insurgency doctrine was playing out from the perspective of the local population through understanding the changes in the behaviour of local people flowing from Canadian military operations.

This paper explores the strengths and limitations of the Canadian attempt to build an integrated assessment framework, with a focus on the development and application of community-level metrics and methodologies. The focus is on community-level assessment here due to space constraints and the fact that community-level assessment offers a fairly complete example of the challenges in attempting to apply and integrate different methodologies and approaches to understanding a population.

Structurally the paper begins with an outline of the major concepts used to build the integrated assessment framework in 2009, followed by a short summary of the various methodological tools used to collect data at the community level. Lastly the paper addresses the problem of how the data collection methods fared in the complex environment of Afghanistan.

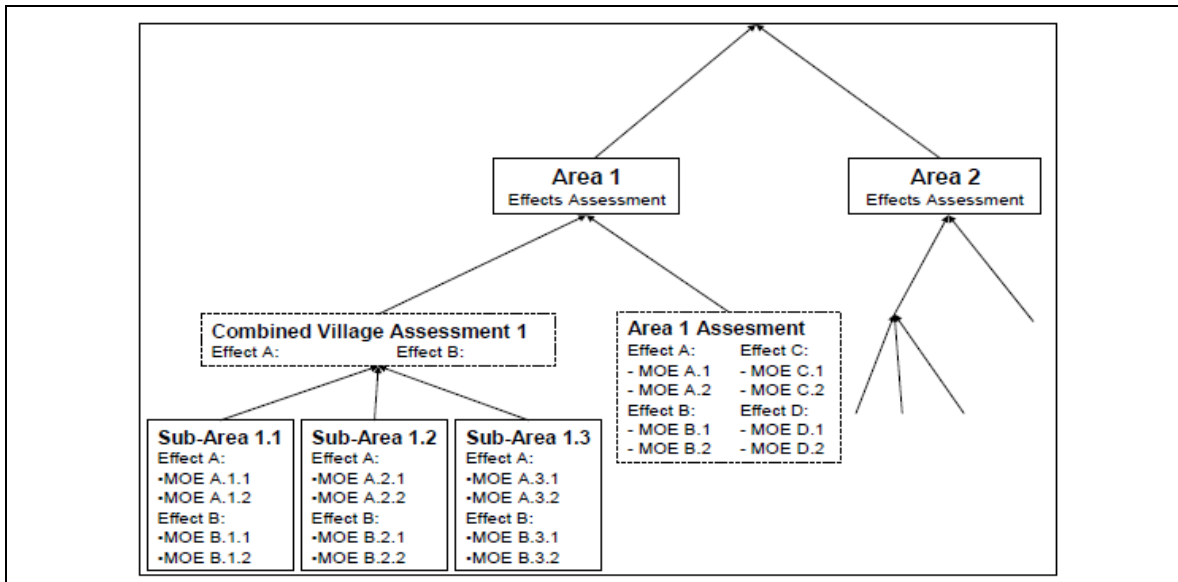
1.1 BUILDING AN EFFECTS-BASED ASSESSMENT FRAMEWORK

The operational plan developed by TF 5-09 was for an *effects*-based operation, effects being currently defined in Canadian doctrine as:

“...the cumulative consequences of one or more actions across the engagement space that leads to a change in situation in one or more domains. Aggregation of intended effects leads to the achievement of objectives.”

(NATO AAP – 6, 2010, Chapter 3, para 20c)

In essence, effects are the desired changes that a commander wishes to achieve across all aspects (security, political, development, etc.) of the operation. Effects that the operational-level task force were expected to achieve were derived directly, or logically, from national-level metrics such as the Canadian Government Benchmarks for Afghanistan (<http://www.afghanistan.gc.ca/canada-afghanistan/menu.aspx?404%3B/canada-afghanistan/progress-progres/benchmarks-reperes/index.aspx>), other strategic guidance, or were implied effects that would be needed to achieve strategic guidance.



**Figure 9 - 1: Multi-level Assessment of Effects across Organizational and Geographical levels
(Dobias Eles, Sprague, & Leahy 2010)**

Effects were seen by TF 5-09 planners as an operational-level construct and the framework from which specific metrics were derived and against which metrics utility was ultimately considered. Each effect had a number of metrics or measures designed in an attempt to map out the construct of that particular effect, though it must be said that this approach of linking specific metrics to effects was driven more by logic and analyst experience than any specific literatures or guiding scientific framework.

In an attempt to better look at the range of possible metrics, planners drew upon two distinct forms of measurement (see Figure 1): *Measures of Performance* (MoP) and *Measures of Effectiveness* (MoE).

Measures of Performance were seen by planners as providing quantitative, concrete measurement of an output (what has been done) and generally involved an increase or decrease of an asset, capability, resource, or a number of events (e.g., violent incidents). Examples from Afghanistan include increasing direct assets, such as new schools, and increasing concrete capacities, such as the number of teachers and size of the Afghan security forces.

Conversely, Measures of Effectiveness were seen as providing quantitative and qualitative insight into how well the desired effects were achieved (e.g., percentage of population trusting the local government, i.e., the Government of the Islamic Republic of Afghanistan, or GIRoA). While MoPs were often about counting exercises, MoEs attempted to capture the “how well are we doing” question through assessment of the population’s behaviours, attitudes, and perceptions.

This combination of metrics led to an eclectic mix of measures, which though it added variety and depth to the assessment framework, presented significant challenges in integration and synthesis of information.

1.2 BUILDING BEHAVIOURAL ASSESSMENT TOOLS FOR EFFECT ASSESSMENT

The initial effects framework laid the foundation for the development of specific methodologies and data collection approaches for addressing specific MoEs and MoPs. Measures of output already existing prior to 2009 and were kept generally intact. To augment these quantitative metrics the new assessment framework added a series of metrics focused on the population’s attitudes, beliefs and, most importantly, behaviour.

Though building metrics on the human dimensions was fairly novel in terms of an integrated approach, the need for such a focus flowed from counter-insurgency literature (i.e., Galula, 2006; Mackay & Tatham, 2009) and from Canadian doctrine that describes the need to make the population’s perceptions, attitudes, and ultimately behaviour the “strategic centre of gravity” (Canadian Counter-Insurgency Operations Manual, p. 4-3, 2008).

Yet, theoretical grounding aside, the practical inclusion of human-centric metrics was problematic. In 2008 there were no known valid or reliable data collection and analysis frameworks in place for qualitative human-centric data; nor was it known which key measures would be valid and practical in assessing progress on human-centric effects. To address the gap, planners considered a number of effects as psychological constructs that could only be assessed through the measurement of a number of different factors, with a corresponding range of data collection and assessment methods, since each facet, or Measure of Effectiveness, represented a complex set of behaviours and was in turn linked to a number of data sources that would, it was believed, provide a usable data source (see Table 1).

Most Measures of Performance included in the effects framework did not require any specialised design, and were collected using existing reporting mechanisms. Conversely, Measures of Effectiveness, which oftentimes relied on qualitative assessments of behaviour, attitudes, and perceptions, required careful development in order to attempt to develop metrics that would meet standards of reliability, validity, and utility while being integrated into the overall assessment framework.

Table 9 - 1: Examples of Effects and Linked Measures of Effectiveness

| Desired Effect | Measure of Effectiveness | Data Sources |
|---|---|---|
| Freedom of movement is achieved | <ul style="list-style-type: none"> • Ability to attend work and school • Ability to transport goods • Ability to visit family • Perception of how safe travel is | <ul style="list-style-type: none"> • Focus groups • Target surveys • GIRoA & non-governmental organisation (NGO) data • Semi-structured interviews |
| Insurgency is separated (morally) from population | <ul style="list-style-type: none"> ▪ School attendance ▪ Level of support for local research & development (R&D) ▪ Voting behaviour ▪ Engagement with government institutions | <ul style="list-style-type: none"> • Direct observation • Observation & provincial reconstruction team (PRT) data • Focus groups • NGO data • Semi-structured interviews |

Yet, as we noted earlier, planners were unable to draw qualitative metrics for various *effects* from the social science literatures. While effects were viewed as psychological constructs from a planning perspective, they in fact represented a complex series or clusters of human behaviour that were not adequately explained by a specific theory or assessment methodology that planners were able to identify. This gap between science and operational needs was challenging and oftentimes not “fixable” within the timeframe of an operation since most theories and assessment methodologies examined by planners lacked either validity or proven relevance to the operational context. A simple example of an apparently simple construct that became difficult to define and measures was a community’s “freedom of movement”:

- 1) Is freedom of movement defined by such quantitative metrics as number of attacks on highways or changes in the pattern of travel?
- 2) Is freedom of movement a psychological construct that reflects people’s perception of safety?
- 3) Do sociological or economic definitions and predictive models provide deeper insight?

How one measures “freedom of movement,” and by extension which data collection methods are employed in the real world, depends on how one defines the construct in the first place. Determining the “best” definition for these constructs was an unsolvable problem for the planners since no literatures or common best practices appeared to exist.

The other major limitation perhaps was the complexity of the environment, and the impact of that complexity on the validity, reliability, and utility of data collected. This limitation was perhaps not fully understood during the pre-deployment planning process. Staying with the freedom of movement example, it could be argued that macro-level analysis of such things as understanding movement trends along major routes and market activity, or the amount of goods travelling from rural communities to cities, would provide valid indicators of freedom of movement, at least on a regional scale. But how does one collect such data? In conflict situations such as Afghanistan there are no toll booths or government structures that would track such large-scale population movements or trends, while contracted solutions in the author’s experience were expensive, unverifiable, and in some cases fraudulent.

By the end of the pre-deployment period, the looming macro challenges of data collection and analysis remained blissfully unknown. The integrated effects assessment framework developed during the pre-deployment planning laid at least a starting foundation for incorporating human behaviour and attitudes into operational-level planning and assessment. The next challenge for the author was to determine whether or not this objective could actually be accomplished.

2.0 PRACTICAL METHODOLOGIES – COLLECTING BEHAVIOURAL DATA

The problem of data collection was addressed in two counter-productive streams: (1) use of contractors to collect macro-level data, and (2) the use of Canadian assets to collect reliable but limited qualitative and narrative data. Both approaches had serious limitations that impacted the ability to craft an integrated assessment framework.

2.1 COLLECTION OF MACRO-LEVEL DATA

During the effects development process, planners had an underlying assumption that polls would be a primary tool for answering a range of MoP- and MoE-related questions. Indeed, drawing from a western perspective, this is logical, since governments, companies, and other groups routinely contract out the extensive and often tedious job of data collection with reliable and valid results.

Based on this western tradition, planners assumed that polling data would be available, or easily generated, for macro-level effects assessment on a range of human-centric effects and related metrics. In fact, a number of very well designed polls in use by Canada and other groups in Afghanistan suggested that macro-level reliable and valid data was already being collected and could easily be incorporated into the integrated assessment framework that was being developed.

However, the author’s actual experience in theatre suggested that little reliable and valid data were being provided by the various polls. Common limitations included the fabrication of data by contractors, limited evidence of question reliability, low power from sample size distributions, and issues of bias in how data were collected. Indeed it was unclear from an evidence-based view whether any contractor, even assuming the best of intentions, could meet the standards of reliability and validity given several overarching limitations:

- 1) Profit vs. Cultural Pressures. An assumption in polling is that the contractor's, and at the tactical level the various supervisors' and data collectors', desire for profit will cancel out any secondary pressures based on tribal, political, or other relationships. Yet it is unclear why this would be the case and indeed given most contractors' reluctance or inability to allow outside observation of their data collection, there is no evidence that such biases are controlled.
- 2) Access to Populations. It is unclear, given the restrictions of such conflict zones, how contractors can gain access to a wide ranging population, including often-times claimed access to diverse tribal groups and women. Though not impossible, most contractors, the author observed, were reluctant to provide examples or details of how such access was obtained.
- 3) Local Variance. It is usually unclear what biases exist in communities. Local security conditions, power struggles, tribal affiliations, or other social factors logically must bias respondents and data collectors, who are often from that region or community in order to gain access. These local fluctuations, which may "wash out" at the macro level analysis, have a usually unknown impact on analysis and interpretation of polling data at the tactical or operational levels.
- 4) Threat & Respondent Bias. The issue of risk and bias should not be understated. Simple questions that probe respondent perceptions of the conflict by definition include threat. Ultimately, to the author's knowledge, no survey was able to control for the issue of respondent bias, nor is it even clear how one would do so, given that bias associated with threat would fluctuate based on variables such as demographics, tribal affiliation, location, and economic situation, among others.

Overall the initial assumption made by planners was that survey data, which were available, would be of sufficient fidelity to allow for linkages to be made between Canadian operations and changes in the overall population's attitudes, perceptions, or behaviours. This was simply not the case. While survey data did inform and support operational thinking of the battlespace, it was not possible to use surveys as an assessment tool of operations, which left a considerable gap in the integrated assessment framework.

The difficulty in using polling data for operational effects assessment can best be explained with a simple example. During a series of community-level operations in Afghanistan, Canadian Forces cleared and secured several key communities to allow for focused and sustained development. After several months of the project there was considerable concern that the survey results reported no change in the population's attitudes or perceptions in the impacted area – despite a massive infusion of funds and resources by Canada. Analysis eventually showed that the sample size of respondents from the area where the operations were taking place was minimal – in fact, most of the respondents from the general region came from the large urban centre rather than from the smaller communities where the operations were occurring.

The end result was that survey data were generally not effective in assisting operational-level effects assessment, which created a considerable gap in understanding effects at the provincial level.

2.2 COLLECTION OF MICRO-LEVEL DATA

The challenges of data collected at the micro or community level stood in sharp contrast to the macro-level data stream. Micro-level data collection drew from social science field research methodologies and used focus groups, behavioural observation, and semi-structured interviews to gather information from a select few communities.

The advantages of this direct field work were notable. It allowed uniformed social scientists to target data collection on specific communities and groups in order to address specific questions and metrics needed for effects assessment. Reliability and validity were directly managed by the data collectors in the community and questions or approaches that were ineffective could be modified in the field.

As an aside the decision to use primarily uniformed social scientists was made initially from availability and national caveats rather than any specific analysis. Defence scientists are not traditionally allowed to operate in among communities and populations where the risk of violence is high nor were social scientists always part of the assigned scientific team. Generally it was found that uniformed social scientists with military backgrounds and graduate-level training performed well in terms of conducting data collection and ensuring that data collection was linked into the operational planning process. What the uniformed social scientists notably lacked however was the deeper scientific expertise in designing a wide ranging data collection framework, identification of key variables, and data analysis. These are Lessons Learned, which are being built into the next version of the Canadian capability.

Taking the previous example of evaluating a community’s freedom of movement, it was possible to develop specific questions that when addressed in focus groups or interviews revealed general perceptions in a community about the danger associated with travel to the closest major city (see Table 2).

Most of the community-level data focused on behaviour rather than attitudes or perceptions since the intent of the integrated assessment framework was to understand how Canadian operations were creating changes in the behaviour of the local population. While an understanding of changing attitudes and perceptions is important as proxy measures of behaviour, it was felt that, given the limited amount of data that could be collected by the sole Canadian data collector in the field, a focus on behaviour would produce the strongest data. Similarly it was also believed that attitudinal change would be more easily and reliably captured in pre-existing macro-level surveys.

Table 2: Examples of Community-Level Data from Semi-structured Interviews

| | Community 1 | Community 2 | Community 3 |
|---|--------------------|--------------------|--------------------|
| Know of people robbed on roads | 5 ¹ | 4 | 0 |
| Know of people who paid bribes on roads | 7 | 3 | 0 |
| Travel regularly to major city | 26 | 23 | 16 |
| Reduce or limit trips to city due to fear of violence | 0 | 6 | 0 |
| Reported examples of intimidation | 4 | 8 | 2 |
| TOTAL | 26 | 23 | 16 |

The micro-level approach used in a few key communities appeared to show strong validity and reliability, though perhaps stretching the reliability and validity terms as they are often-defined in quantitative literatures.

Reliability of behavioural data in this context involved maintaining consistent observational context over a number of weeks. For example, interactions between local children and members of the Afghan National Police were observed in similar conditions (on Canadian led patrols) by the same researcher who had the same position on patrol over a period of weeks in one community.

Similarly validity, if we view it as an assessment of whether we are in fact measuring what we desire to measure or construct validity, was aided by the focus on specific behaviours. For example, traffic patterns used by the community for economic reasons were compared against recent efforts on route clearance to assess how the community perceived its freedom of movement over the district. In this sense data, which were collected through observation, semi-structured interviews and group discussion, provided a valid

¹ Those who reported being aware of robbery all reported that only people with obvious money were targeted; often robbers would track them to the city and then rob them on the way out.

assessment of which roads and approaches into the main city the community felt were safe. Interestingly, planners were never able to advance to the next logical state, predictive validity, which would have seen targeted changes in route clearance then impact the ebbs and flows of community economic activity.

Finally, the narrative data gathered through interviews provided a depth of understanding that allowed for Canadian leaders to understand the “why” behind community behaviour. For example, narrative data revealed that the community valued education but were unwilling to send children to the school for safety reasons. Once Canadian leadership addressed the security issues raised by the community, children started attending school, with the numbers increasing as the community came to see the security as stable.

Despite the advantages of micro-level data collection, the data stream had very limited impact on informing and answering key effects outlined in the integrated effects framework. The limited amount of data that could be collected through this approach, combined with the difficulty of transferring results to the wider province, and the difficulty of interweaving qualitative data with traditional assessment methods, limited the utility of community-level data collection.

Combining qualitative data with more traditional assessments was especially problematic. Put simply not enough time had been spent working with analysts and operational planners to develop approaches to viewing and interpreting qualitative data so that it could inform operational planning. In some cases this resulted in qualitative data providing insight into a community or area – but planners were left without the tools for how to use that data and insight to inform and shape the next series of operations.

However, the most critical weakness, from an operational perspective, was the resource cost. The qualitative approach revealed deep and useful knowledge about specific communities, but there were not sufficient resources to cover all of the communities in an operational area of interest; this created situations where the depth of analysis varied significantly across a series of communities.

The third challenge with micro-level data was the ability to incorporate this form of information into an integrated assessment framework. Put simply, deep qualitative data about specific communities did not fit within traditional assessment and reporting practices. No social scientist had worked at the community level before and staff at all levels across the WoG team were unfamiliar with how such qualitative data should be integrated and considered along with the more traditional and familiar quantitative data. Nor was sufficient attention paid during the development stage prior to deployment about how such data would be incorporated into planning processes. Addressing these concerns through additional review panels and research into the process of interweaving qualitative or deep narrative data into formal assessment and analysis processes would be beneficial.

The lack of familiarity with qualitative data by assessment staff, and of the author in building qualitative data into a formalised assessment framework, was confounded by the inability to develop any form of convergent validity over the deployment. Since survey data and qualitative data in the communities tended to assess two distinct populations (oftentimes surveyors did not even go to communities where the author worked), there were no instances where different data collectors assessed the same population in a rigorous way.

Overall, the gathering of qualitative data in key communities was both a resounding success and a notable failure. It was a success in that it provided proof of concept that rigorous data collection aimed at answering key questions of operational relevance is in fact possible in counter-insurgency operations. However, it was a failure in that the data were not able to inform a wider assessment framework or to be integrated with quantitative methods successfully.

3.0 SUMMARY

A successful integrated approach to effects assessment could be described as being able to address different questions for different groups across the organisation, or set of organisations/entities, as is the case within a comprehensive/WoG context. The integrated design was seen as a method of answering not only “what” questions through quantitative metrics but also attempting to address the “why” question through qualitative and narrative analysis.

Survey data can be a crucial data stream to address macro-level effects in an integrated framework. Despite a host of technical and ethical challenges embedded with using survey data in war zones, it is theoretically possible that surveys conducted by contractors can be reliable and valid and can offer utility for the assessment framework (though the evidence of surveys to meet those benchmarks remain unclear, despite the financial and resource costs that they consume).

The challenges of qualitative data collection are different in nature. Field data can be reliable and valid and may provide the critical “why” that leadership requires for understanding the linkage between Canadian operations and the impacted communities. In particular, qualitative data may be most useful for informing leaders at the tactical level. It was fairly easy to use field data to inform small-scale tactical problems that could be engaged by military leaders living and working in those communities. What qualitative data could not do was inform patterns of behaviour across the wider region. Nor could qualitative data be effectively linked to other data sources to allow for some form of convergent validity.

Overall, an integrated effects assessment is possible. More work in developing robust survey methods would allow for a wider range of reliable and valid survey data to inform the macro level. Development of a deep analytical capability, often-times referred to as “red teams,” would aid in interpreting macro-level trends and data. On the qualitative side, better training and support for data collectors working in the field would help to ensure that enough qualitative data are collected to allow for transferability of information, as appropriate. Further, analyst training on the use of qualitative methods would help incorporate narrative and qualitative data into planning and assessment processes.

4.0 REFERENCES

- [1] AAP-6(2010) - NATO Glossary of Terms and Definitions, 2010.
- [2] Canadian Counter-insurgency doctrine, B-GL-323-004/FP-003, 2007.
- [3] Dobias, P., Q. Innis, J. Caruana, (2009). Effects-Based Approach to Operations in Task Force Kandahar: Incorporating Effects in Planning and Assessment Cycle (S), *DRDC CORA TM 2009-049*, November 2009.
- [4] Galula, D. (2006). *Counterinsurgency Warfare: Theory and Practice*. Praeger Security International.
- [5] MacKay, A., & Tatham, S. (2009). Behavioural Conflict from General to Strategic Corporal: Complexity, Adaptation and Influence. *The Shrivenham Papers*, 9.
- [6] Rittel, Horst, and Melvin Webber; "Dilemmas in a General Theory of Planning," pp. 155–169, *Policy Sciences*, Vol. 4, Elsevier Scientific Publishing Company, Inc., Amsterdam, 1973. [Reprinted in N. Cross (ed.), *Developments in Design Methodology*, J. Wiley & Sons, Chichester, 1984, pp. 135–144.], http://www.uctc.net/mwebber/Rittel+Webber+Dilemmas+General_Theory_of_Planning.pdf

