

## Assessment of Effects-Based Operations Based on a Whole of Government Approach

Ms. E. Kalantzis  
P. Dobias, Ph.D.  
Mr. D. Connell

Defence Research and Development Canada  
Centre for Operational Research and Analysis  
National Defence Headquarters  
Ottawa, Ontario, Canada.  
e-mail: Kalantzis.E@forces.gc.ca

*Ms Eugenia Kalantzis is a defence scientist with the Centre for Operational Research and Analysis of Defence Research and Development Canada. Ms Kalantzis currently occupies the role of Canadian Expeditionary Forces Command Operational Research and Analysis Team Leader. The mandate of this team is to support operations through the provision of practical analytical advise to decision-makers. Ms Kalantzis also coordinates technical reachback to deployed analysts supporting Canadian Forces in theatre. She holds both a B.Eng (1994) and M.Eng (1997) from the Department of Civil Engineering and Applied Mechanics of McGill University in Montreal, as well as an MBA (2003) with a concentration in Operations Management. Her areas of interest include capability-based planning, performance measurement, business intelligence, and statistical analysis. Most recently, Ms Kalantzis has been focused on campaign plan assessment and the development of methodologies for the assessment of effects based operations.*

*Dr. Dobias received his Master's of Natural Sciences in theoretical and mathematical physics from Comenius University in Slovakia in 1998. He was awarded his Ph.D. in physics from the University of Alberta in 2002, where he also continued for an additional three years as a postdoctoral research fellow. His research area was primarily the investigation of the stability of dynamical systems. He joined the Centre for Operational Research and Analysis of Defence Research and Development Canada in 2005 as part of the Land Force Operational Research Team where his research focused on combat modeling. Other areas of research included the development of tactical measures of effectiveness/performance for combat models in operations with limited attrition. In 2006, Dr. Dobias was seconded to the CEFCOM Operational Research Team where his work focused on the Effects Dashboard. Dr Dobias also provides operational research support to CEFCOM J5 Effects and to the CEFCOM Effects Assessment Workgroup, and has been a participant in the Whole of Government assessment process.*

*Mr. Connell is a retired, senior air force officer (Air Navigator) with a background in operational flying, test and evaluation of sensors and systems, and advanced technical instruction and education. In 2001 he joined Defence Research and Development Canada as a defence scientist within the Centre for Operational Research and Analysis. During the past six years he has worked primarily as a concept developer and strategic analyst with the Canadian Forces Experimentation Centre, the Land Force Operational Research Team, and most recently within the Directorate of Strategic Operations Research, Strategic Joint Staff.*

*His experience with Effects Based Operations includes concept development, multinational experimentation with Joint Forces Command; and, operational experience with NATO ISAF as an operational analyst deployed to Kabul. This latter experience included conducting effects based assessments, campaign plan assessments, and periodic mission reviews.*

## **ABSTRACT**

An Effects-Based Operation (EBO) is defined as “*an operation planned and conducted to obtain a desired effect through the synergistic and cumulative application of the full range of joint, interagency, multinational and public capabilities on the physical, moral and cyber planes*”. When the operational focus involves the “*coordinated efforts and integrated defence, development and diplomatic strategies*”, it can be said that the mission is based on the Whole of Government approach. The analysis of such a mission presents several challenges to the analyst in that it requires the collaborative effort among all partners in the assessment of effects achieved across all objectives. Though it appears straightforward, this proposition is beset with difficulties associated with bringing together distinct organizational cultures.

To address the challenge of reporting on progress in the Afghanistan campaign, the Canadian Expeditionary Forces Command (CEFCOM) J5 Effects Cell developed a performance measurement methodology called the CEFCOM Effects Dashboard. Though a good initial product, certain limitations existed and the Operational Research Team was asked to strengthen the Effects Dashboard by reviewing the existing process and tools, and proposing and implementing changes that would lend additional rigour to the assessment process. This paper presents the recommendations put forward to strengthen the Effects Dashboard. At the heart of this new framework is a set of effects linked from the operational to the more tactical level, along three lines of operation related to governance, development and security. In total, the assessment process involves the evaluation of 132 effects; a set of indicators illustrating the desired characteristics of each effect and an anchored rating scale are defined to help decrease variability of ratings across raters and rating periods. Additionally, the incorporation of an impact scale and an automatic assessment process for higher-level effects are also discussed. Finally, a process is presented to integrate and gain synergies from the individual assessments across the different partners; this process starts with the assignment of ownership of effects to key partners and ends with the cooperative assessment of overall performance along all three lines of operation.

## **INTRODUCTION**

The concept of an effects-based operation (EBO) is not new. An EBO is defined as

“an operation planned and conducted to obtain a desired effect through the synergistic and cumulative application of the full range of joint, interagency,

multinational and public capabilities on the physical, moral and cyber planes”<sup>1</sup>.

When the operational focus involves the “*coordinated efforts and integrated defence, development and diplomatic strategies,*” it can be said that the mission is based on the Whole of Government approach. Nowhere do these definitions apply more in recent Canadian history than in this nation’s involvement in Afghanistan. This mission brings together diverse partners working to achieve objectives that are in themselves quite distinct, but nonetheless linked: the success or failure of one is intimately tied to the success or failure of another. Therefore, the EBO approach necessarily requires an assessment mechanism which links actions, effects and the achievement, or lack thereof, of objectives through a formal feedback mechanism that enables decision makers at the strategic level to complete the EBO cycle.

This type of mission presents a challenge to the analyst because it is inherently difficult to assess how, or even if, the cumulative results of a set of actions are in fact contributing to the desired effects. Two questions naturally emerge in the analysis of an EBO. The first is a question of tracking our progress: *Are we achieving the desired effects?* This question focuses on assessing and tracking progress along predefined and categorized measures encompassing the Whole of Government objectives. The second question is one of causality: *What action or set of actions has contributed to our observed progress in any particular area?* This question attempts to define the causal relationships between actions and measures, and the correlations amongst observed changes in the predefined measures. A prerequisite to answering these questions is the collaborative effort among all partners in the assessment of effects achieved across all objectives. Though it appears simple, this proposition is beset with difficulties associated with bringing together distinct organizational cultures. These difficulties include developing a common language and a common understanding of the roles and responsibilities across organizations, as well as the more tangible difficulty of developing an assessment tool that will eliminate duplication and encourage a unified effort across groups.

To address the challenge of reporting on progress in the Afghanistan campaign, the Canadian Expeditionary Forces Command (CEFCOM) J5 Effects Cell developed a performance measurement methodology called the CEFCOM Effects Dashboard. Though a good initial product, certain limitations existed and, in July 2006, the CEFCOM J7 Operational Research Team was asked to review the existing process and tools, and to propose (and eventually implement) changes that would lend additional rigour to the assessment process.

More specifically, the objective of the CEFCOM J7 Operational Research Team was to recommend and implement changes to the Dashboard to:

- Strengthen the core (i.e. the set of measures and the relationship between measures),
- Strengthen the assessment process (i.e. to render it more robust, consistent, and simple),

---

<sup>1</sup> Termium is an online definition/translation/abbreviation information resource provided by the Translation Bureau, Public Works and Government Services Canada: TERMIUM Plus @ <http://termium.gc.ca>.

- Implement an efficient data management tool and a more user-friendly interface,
- Simplify the presentation of results and highlight areas of risk, and
- Enable the integration of the requirements of all partners.

This paper presents background information related to performance measurement in Afghanistan; discusses the common limitations and constraints observed across most measurement methodologies; describes the key characteristics of the new Effects Dashboard; and proposes a process by which this one overarching framework can be used by different groups to conduct individual campaign assessment while benefiting from the synergies derived from a cooperative assessment of the mission.

Finally, it should be emphasized that, at project start, the original Effects Dashboard was already in place and being used to conduct assessments and report on mission progress. In order to minimize disruption to the assessment process and to retain a consistent presentation of results, the sponsor requested that the recommendations proposed by the analysis team incorporate, whenever possible, the features of the original Dashboard. Whenever relevant, these constraints will be discussed herein.

## **BACKGROUND**

### **PERFORMANCE MEASUREMENT – MEASURING WHAT?**

Performance measurement in Afghanistan is a pervasive activity and a requirement of nations and their populations to rationalize the expenditure of military and monetary resources dedicated to the mission. Many groups, both in and out of theatre, are spending tremendous amounts of effort on the design of evaluation methodologies and the creation of tools to accomplish tracking and reporting tasks related to the state of security, governance and development.

But what exactly are all these groups measuring, tracking and reporting? For the most part, these groups are evaluating the progress being made towards the achievement of a desired end-state in Afghanistan. The desired end-state in question is articulated in many documents, but two key documents are of note. The first of these is the Afghan National Development Strategy (ANDS). The ANDS outlines the strategic vision of the Afghan Government, as well as the mechanisms and priorities for achieving this vision along three pillars. These three pillars are:

1. Security
2. Governance and Rule of Law and Human Rights, and
3. Economic and Social Development.

In addition, counter-narcotics and gender equality have been singled out and identified as cross-cutting and vital pillar issues.

The second key document used in the development of performance measurement methodologies is the Afghanistan Compact. The Compact was the international community's follow-on to the Bonn Agreement for the Government of Afghanistan, and it outlines specific objectives (or benchmarks) to be met in the path towards the achievement of the strategic vision described in the ANDS. The benchmarks of the Compact include both qualitative and quantitative measures, depending on the nature of the objective. As an example, the benchmark for the number of police to be trained is articulated as follows: "A fully constituted, professional, functional and ethnically balanced Afghan National Police and Afghan Border Police with a combined force of up to 62,000 will be able to meet the security need of the country effectively and will be increasingly fiscally sustainable." This example clearly identifies a broad mix of qualitative characteristics with a quantitative limit; however, the language is too vague and imprecise to constitute an effective measure of performance.

Following a strong push for the Compact to be linked to the ANDS, the two documents were eventually aligned with the same pillars, and the Compact was officially annexed to the ANDS. These two documents, and in particular the set of benchmarks outlined in the Compact, served as a baseline for the development of many of the performance measurement tools currently being used to assess progress in Afghanistan.

#### PERFORMANCE MEASUREMENT – MEASURED BY WHOM?

Many different groups are interested in the evaluation of progress in Afghanistan. Among them are the following:

- The Joint Coordination and Monitoring Board (JCMB). The JCMB was created to assess and track progress along the benchmarks listed in the Compact. However, given the limited resources available for data collection and analysis, this group conducts a general assessment only and produces bi-annual reports for distribution to the international community.
- The NATO International Security Assistance Force (ISAF), the Combined Forces Command - Afghanistan (CFC-A, and now replaced by Combined Security Transition Command - Afghanistan (CSTC-A)), the UN Assistance Mission to Afghanistan (UNAMA) and other participating organizations. Despite the overall NATO ISAF structure that currently unites the various international partners, each organization conducts an independent assessment at the national, regional or Provincial Reconstruction Team (PRT) levels. Though the ISAF construct brings these assessments together, there is little to no cooperation between the various communities to combine assessments across different areas of interest and to benefit from both the synergies in data collection and the multi-dimensional view of progress.
- Individual countries. Members of the international community conducting operations in Afghanistan (within NATO, the coalition, the lead nations, or through bi-lateral arrangements) are also measuring progress, particularly within their area of operations. The United States, the United Kingdom, the Netherlands, and Canada have all put in place measurement methodologies to assess and report on progress. Additionally, even within each country,

different organizations (DFID, USAID, DFAIT, etc.) are conducting assessments to suit their own reporting objectives.

As one can see, there is a tendency to develop a reporting structure and performance measurement methodology to satisfy one's own needs in the short term. However, commonality of tools *is possible* and needs to be investigated further. Although they are packaged differently, most performance measurement methodologies are evaluating progress along the same three pillars (or *lines of operation*) using a similar set of measures. Despite this commonality of purpose, as well as the serious lack of resources for data collection, reporting and assessment on the part of most of these groups, few (if any) are working together to share assessments, to coordinate data collection and analysis, and to benefit from potential synergies either in theatre or outside.

### COMMON LIMITATIONS AND CONSTRAINTS ACROSS MEASUREMENT METHODOLOGIES

Certain key limitations and constraints have been observed across many of the performance measurement methodologies. They are highlighted here because, to some extent, some of these apply to the first versions of the CEFKOM Effects Dashboard - the tool that is the focus of this paper. However, more important to note is that the presence of these problems may help to explain why individual organizations resort to creating methodologies tailored to meet their particular needs, instead of adopting existing ones. Of course, there are many other reasons that further explain this phenomenon; nevertheless, these limitations and constraints were key considerations when determining what changes/improvements should be made to the CEFKOM Effects Dashboard to create the highest impact. Though not an exclusive list, some of the key limitations and lessons observed by the authors in their investigation of similar methodologies are as follows:

- There is generally an inadequate selection and definition of measures in the lines of operation (LOOs) not related to security, i.e. development and governance. This is primarily due to the fact that these methodologies were developed by military stakeholders that are not necessarily subject matter experts in economic development or governance objectives, or the measures of effectiveness associated with them. As a result, gaps in areas of interest are often observed.
- Even within the LOOs related to security, there is a general weakness in the selection and definition of measures related to less traditional or more specialized functions, such as information operations.
- At the more tactical level, there is a recognized weakness in defining effects and measures that are relevant to and measurable by the tactical level commanders, i.e. those that are tasked with conducting the assessment. As a result, measures are often not evaluated, or evaluated arbitrarily or inconsistently. Furthermore, the lack of relevance of measures makes the assessment process tedious and results in the disengagement of these key stakeholders.

- 
- The set of measures and the relationship between these measures (i.e. the overall tree structures) are not balanced. As a result, when automatic assessments of higher-level effects are conducted (e.g. using weights and 'roll-up' functions), there is generally an unintended and inconsistent weight associated with some measures. This is a consequence of how the overall tree structure is constructed, and often leads to unexpected levels of influence for particular measures (either too high or too low) and an inaccurate portrayal of overall performance.
  - Many measures are hard to evaluate quantitatively (due to the lack of exact numerical data). Often, the assessors are not given any alternative (more qualitative) means of assessing the measures. As a result, the assessment team does not provide any impression at all as to progress in a given area, though they may have been able to provide a more qualitative assessment had the appropriate mechanisms been in place.
  - Generally, benchmarks of performance (e.g. ratings of bad, good, best) are vague and not clearly articulated. As a result, the interpretation of the rating scale may be different across raters, and across rating periods. This leads to inconsistent results, meaningless trends and a questionable interpretation of assessments.
  - There are too many effects and measures to assess. Many measurement methodologies were originally designed to include a reasonable number of measures. However, many were overtaken by the addition of measures with each subsequent iteration, to the point where the number of measures rendered the management of the assessment impractical. First, the analysis of the measures becomes overly burdensome as the number of measures grows. Second, the information provided by the more important measures is diluted or masked by the presence of unimportant measures. Finally, measures included are often irrelevant to the majority of stakeholders, leading to a misuse of the limited resources available to manage data and compile assessments.
  - The assessment process further suffers from problems associated with an unreliable and/or un-validated measurement process. A lack of ownership regarding the development of the measures as well as the tendency towards poor documentation contributes to unreliable and inconsistent reporting. The data measurement and assessment process is further eroded by a lack of baselines, inadequate rating schemes, and the universal application of scales, measures or assessments for national, regional or central purposes.
  - Staff involved in the assessment process represent another source of error which contributes to unreliable and inconsistent assessments. There is a tendency to have separate staffs involved in the design and development of the process and measures versus those charged with their implementation. Further, there is insufficient training and education at all levels, and indeed a lack of validated doctrine regarding the EBO approach and assessment processes. Finally, there are insufficient structures, capacity and expertise available to deploy, implement and maintain the majority of assessment methodologies.

- Finally, in the development of any performance measurement methodology, issues related to data quality must be kept in mind. There is generally a poor level of availability of data, poor reliability, and no commonality of data across most organizations that feed the assessment process at the various levels. This limitation must be accommodated in the design of a performance measurement methodology.

Taking the above limitations and lessons into account, there are a number of key insights that can be drawn related to the development of an EBO assessment methodology. First, there is a requirement to put in place formal processes and organizational structures to take ownership of the assessment methodology – its development, use and management. Second, within these organizations, staff must be trained and assigned roles directly supporting planning and operations. This would ensure that the assessment mechanisms put in place would positively influence the development of future plans and would guide the conduct of current operations. Third, the assessment methodology must be relevant and capable of informing decision makers across the Whole of Government, and at all levels (strategic to tactical). Finally, the methodology must be manageable, as well as adaptive and flexible. To remain relevant, it must be flexible enough to accommodate change. However, growth must be strictly controlled to ensure that the process remains as simple as possible and that unnecessary complexity is avoided.

## THE SPONSOR AND THE EFFECTS DASHBOARD

The sponsor of this project is the CEFCOM J5 Branch. CEFCOM is one of four operational level command headquarters in the Canadian Forces (CF), and is responsible for the planning and conduct of all CF international operations, with the exception of operations conducted solely by the Canadian Special Operations Forces Command. At the present time, a primary operational focus of CEFCOM activities is the Afghanistan mission. As part of its mandate, CEFCOM is responsible for reporting on progress in Afghanistan to the strategic headquarters, as well as to the Government. To serve this requirement, the CEFCOM J5 Effects Cell developed a performance measurement methodology called the CEFCOM Effects Dashboard, or Dashboard.

The original version of the Dashboard consisted of a catalogue of measures of performance (MOPs) and measures of effectiveness (MOEs) along three LOOs related to security, governance and development. MOPs were typically assessed in theatre, and were quantitative in nature. These assessments were reviewed and combined with other information by a team of subject matter experts (SMEs) at CEFCOM headquarters; this information was then used to evaluate a set of MOEs that were more qualitative in nature. This cycle was repeated every six weeks. For the most part, the process resulted in a subjective assessment, achieved mostly by consensus among the assessment team members. The tool used to record these assessments was a specially designed EXCEL spreadsheet (one per assessment period) with a color-coded assessment of the current status of the various measures.

Figure 1 shows a partial representation of the Dashboard (all data in Figure 1 are fictitious). Though a good initial product, certain limitations existed. Among others, there was no efficient way to view past results and to analyse trends; the assessment scales were

not anchored which made cross-period rating consistency questionable; the definition of MOPs and MOEs did not lend itself easily to interpretation; and the tool was difficult to manage (over 20 pages of printed material), therefore difficult for decision makers to utilize effectively.

Strategic Objective	JCMB Benchmark	Benchmark MOE	Benchmark Comment	Op Effects	MOE	MOE status	MOE Comment	Task (Action)	MOP	MOP Status	MOP Comment
Stabilize region	International Forces			Secure & stable AO	Number of attacks		Less than last cycle	Occupy AO	Number of patrols		More pts than last cycle
					Trends in attacks		Still decreasing cycle to cycle		Coverage of patrols		Focus on X area
					Initiative		Coalition has initiative		Engagn't initiated by CF		95%
	National Forces			Efficient and effective Security Forces	Ops with CF		Ok, need supervision	Sp Security Forces	Number of joint patrols		55%
					Independ't Ops		Lack of planning capability		Can plan Ops		Not yet.
					Ops with Police		Many incidents of fratricide		Can work with Police		12 fratricide this cycle

Figure 1: Partial Representation of the Original CEFCOM Effects Dashboard (all data are fictitious).

Despite these limitations, one of the main strengths of the process was that it was a first attempt to include the participation of other government departments, particularly in the assessment of LOOs for which these other organizations were both SMEs and custodians. As a result, the assessment process brought together an inter-departmental working group, which included elements of the Canadian International Development Agency (CIDA) and the Department of Foreign Affairs and International Trade (DFAIT) to conduct the assessment of status along each of these measures and LOOs.

In light of the general endorsement of the Effects Dashboard, and to build on the momentum generated, it was felt that the process would benefit from an unbiased review. The intent was to build on the strengths of the methodology and to correct, to the extent possible, any limitations observed. The recommendations resulting from this evaluation are presented below.

### STRENGTHENING THE EFFECTS DASHBOARD

In July 2006, the CEFCOM J7 Operational Research Team was asked to review the existing process and tools, and to propose (and eventually implement) changes that would lend additional rigour to the assessment process. More specifically, the objective was to review the set of measures and fill in any gaps, to strengthen the relationships between these measures and the LOOs at the tactical, operational and strategic levels, to make the

evaluation process less variable and more robust, and to improve our ability to track changes over time and extract possible links between our actions and the effects observed.

### KEY CHARACTERISTICS OF THE NEW EFFECTS DASHBOARD

Following the evaluation of the Effects Dashboard, recommendations were made to address the objectives previously presented. To some extent, these recommendations also address all but the last two items listed in the section entitled *Common Limitations and Constraints Across Measurement Methodologies*. The following proposed recommendations are discussed herein<sup>2</sup>:

1. Strengthened core: new tree structure.
2. Indicators for each L3 effect.
3. Anchored performance scale.
4. Anchored impact scale.
5. Automatic assessment of higher level effects.
6. New interface and data management tool.

### STRENGTHENED CORE: NEW TREE STRUCTURE

The first recommendation proposed was to strengthen the core, i.e. to adopt a new tree structure. The term *tree structure* refers to the set of LOOs, effects and measures (e.g. MOEs/MOPs) that are collected, tracked and used to assess progress within the effects and along the three LOOs, as well as the relationship between these measures.

The proposed tree structure was constructed following a period of review of existing measurement methodologies, as well as surveys and reports highlighting the priorities and needs of the Afghanistan people (see References). The latter research contributed greatly to ‘filling in the gaps’ often observed in the measurement of performance along the LOOs of governance and development. For example, the readings highlighted the high importance of the formal recognition of land ownership by the Afghanistan people – a measure not included in most tools but added to this proposed structure. Therefore, strengthening of the tree structure relied on an analysis of the environment upon which the desired effects were meant to operate which partitioned or categorized it within the framework of a campaign plan for action and response.

Following this period of review, a set of measures was proposed to capture the full range of desired effects along the three campaign plan LOOs (as outlined in key documents such as the ANDS), along with a proposed set of relationships between these measures. A general

---

<sup>2</sup> At the time of publication, most recommendations have been implemented, though some remain in the final development phase.

representation of the resulting tree structure is presented in Figure 2. The key characteristics are as follows:

- As in the original Dashboard, there are three LOOs referred to as Governance, Development and Security. These are comparable to all of the various campaign plans developed by organizations and nations involved in the Afghan mission.
- Each of the three LOOs is broken down progressively from Level 1 (L1) effects, to Level 2 (L2) effects, and finally to Level 3 (L3) effects, with each level representing more granularity than the previous. In the new Effects Dashboard, L1 effects represent operational level effects, while L3 effects represent effects that are more closely related to the tactical level. In all cases, these are the *desired* effects. Three levels were selected because the resulting construct provided a natural and intuitive separation of effects and an appropriate level of granularity amongst the L3 effects. Furthermore, the three levels were originally selected to correspond to a breakdown roughly equivalent to the strategic, operational and tactical levels.

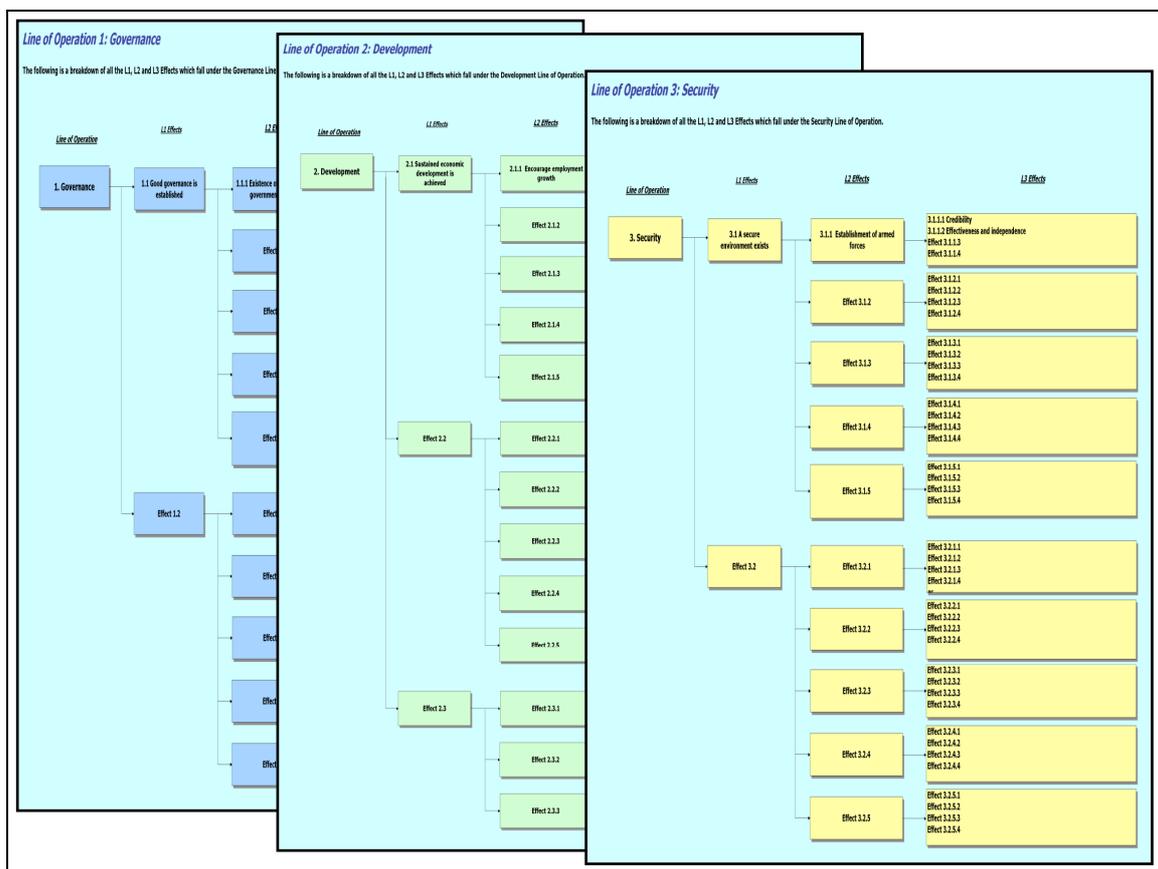


Figure 2: Representation of the New Effects Dashboard Tree Structure.

- Taking into account the lessons identified earlier, a balanced, rigid set of guidelines was applied in the development of the tree structure. Within each LOO, the tree structure is balanced; each node breaks up further into a specified and fixed number of branches (each L1 effect breaks up into either

three or five L2 effects; each L2 effect breaks up into four L3 effects). This ensures that any effect from any given level is given equal weight (at least initially) and that any change will not influence performance at the higher levels by a disproportionate amount (i.e. when the performance at higher levels is calculated from the performance of lower level effects).

- In order to prevent problems associated with previous methodologies, and as a means to focus the assessment mechanisms (collection, analysis, information distribution and management) on measuring key effects, the structure and the number of effects will be fixed, i.e. will not change or grow. When the number of effects becomes too large (over 150), it becomes very difficult to analyse the data set and extract relevant underlying relationships with the currently available resources and short reporting cycles. Controlling the number of effects will enable the analysis team (including plans and operations staff) to manage the assessment process, and will allow them to become conversant and expert in the interpretation of the results. With experience this number may be reduced further.
- Overall, there are fewer effects to assess. The structure is broken down as follows: in LOO 1 Governance there are 2 L1, 10 L2, and 40 L3 effects; in LOO 2 Development there are 3 L1, 13 L2, and 52 L3 effects; and in LOO 3 Security there are 2 L1, 10 L2, and 40 L3 effects. It should be noted that, in the new assessment process proposed herein, only the L3 effects are assessed directly. The performance ratings of L2 and L1 effects may be determined automatically from the results recorded for the L3 effects (to be discussed later). In total, 132 L3 effects are evaluated at each reporting cycle (versus approximately 300 in the original Effects Dashboard).

### INDICATORS FOR EACH L3 EFFECT

The second recommendation to strengthen the Effects Dashboard is the introduction of a set of indicators for each L3 effect. Rather than creating a set of dedicated MOEs and MOPs that would be difficult to define and apply in the Afghan theatre (for reasons previously discussed, such as the lack of exact numerical data and staff training), it was decided that a set of indicators or attributes would be more appropriate to guide the assessment process. Indicators are provided for L3 effects only because, as mentioned earlier, assessors are only required to provide an assessment of the L3 effects.

The set of indicators for each of the L3 effects consists of approximately five attributes or conditions that describe the characteristics that would be present if the desired effect was fully achieved. The main objective in including these indicators was to decrease variability and make assessment across effects, assessors and rating periods more consistent and robust. These indicators serve as a guide to the assessors when evaluating progress along a given effect and ensure that assessors consider the same general attributes when making their assessment from one period to another. This was the first step in introducing more rigour to what is essentially a subjective assessment. Note once again that, at the present time, the environment in theatre is not conducive to a rigorous objective measurement system.

### ANCHORED PERFORMANCE SCALE

Another method introduced to decrease variability and make the assessment more robust is the use of an anchored performance scale, as seen in Figure 3. The anchored performance scale is a simple five-point scale (from 1 to 5) describing the different levels of progress towards the achievement of each of the L3 effects.

5	<b>Fully achieved:</b> The effect can be considered achieved; all of the indicators are present to a significant degree, only some minor improvements can still be made.
4	<b>Acceptable standard:</b> The effect is almost fully achieved; though most of the indicators are present to a significant degree, some improvement can still be made.
3	<b>Below acceptable standard:</b> The effect has not been fully achieved, but a noticeable improvement is observed. Most of the indicators are present, albeit to a minor degree, though improvement is still required in this area.
2	<b>Minor:</b> The effect has not been achieved; however, a few indicators are present to a minor degree, but significant improvement is still required in this area.
1	<b>Negligible:</b> The effect has not been achieved; almost none of the indicators are present and significant improvement is required in this area.

*Figure 3: Anchored Performance Evaluation Scale.*

The benefits of anchoring the scale are that it makes the measurements less prone to variability as a function of who is conducting the assessment, and allows for the assignment of a ‘score’ that will allow us to develop the tools to track and display performance in time. This will provide a sense of movement along the scale (‘trend’ from one reporting period to the next).

The five-point scale was implemented to accommodate a reasonable range of scores while limiting the potential for variability by clearly delineating the performance associated to each of the levels. By doing so, an assessor with limited training can readily and confidently apply the scale in the assessment of effects. Furthermore, the universal application of this evaluation scale across all measures contributes to easing the assessment of L3 effects and the interpretation of results.

This anchored performance scale is used to assess each of the L3 effects, using the set of indicators for that particular L3 as a guide in the assessment. In practice, the combination of these tools (indicators and anchored performance scales) has shown to contribute to both decreasing the variability in the assessment, and making the assessment process easier and quicker by ensuring that all members of the assessment team have a common understanding of the effect they are evaluating and the score they are assigning.

## ANCHORED IMPACT SCALE

In addition to the assessment of performance, a second scale is currently being considered to capture the degree of impact of a given effect on the overall mission.

A word of caution is required at this point. The initial intent of the methodology proposed in this paper is to ensure the visibility and equal consideration of all the effects included in the tree structure. It may therefore appear inconsistent to implement an impact scale. However, it has been generally observed that, although all effects in the new Dashboard are essential, some may exact a higher level of influence or require special attention. This observation has prompted an investigation into the implementation of an impact scale. An example of an impact scale that could be used is shown in Figure 4. Prior to implementing any impact scale, some key points must be considered:

- Impact categories such as those presented in Figure 4 are quite subjective and potentially controversial,
- Assigned levels of criticality (or impact ratings) could be transitive, i.e. not fixed in time,
- Potential exists for the implementation of impact ratings to unduly complicate the assessment process, thus negating past efforts to reduce the ambiguity and subjectivity of the process, and finally
- Over time, the trends in the assessment of performance may in fact provide more pertinent evidence of impact than would a subjective categorization such as this one.

<b>IV</b>	<b>Critical:</b> The effect ranks among the <i>highest priority</i> effects identified by the Canadian Government and is also considered an <i>essential or key enabler</i> to our ability to achieve other priority effects.
<b>III</b>	<b>Important:</b> The effect ranks among the <i>highest priority</i> effects identified by the Canadian Government and is an <i>enabler</i> to our ability to achieve other priority effects.
<b>II</b>	<b>Necessary:</b> The effect is considered an <i>essential or key enabler</i> to our ability to achieve other priority effects.
<b>I</b>	<b>Desirable:</b> The effect is considered an enabler to our ability to achieve other priority effects, and is considered a desired outcome of the mission.

*Figure 4: Anchored Impact Evaluation Scale.*

Nonetheless, keeping the above in mind, a preliminary impact scale has been developed for consideration. It consists of a four-point scale (from I to IV) describing the degree to which a given effect is an enabler to the achievement of the mission and of other high priority effects. Additionally, if each of the L2 and L3 effects is assigned an impact rating,

then these could be used to conduct an automatic assessment of the performance of higher-level effects (discussed below). However, unlike the rating of performance, which is performed every six weeks, the assignment of the level of impact would be done only once and the same values used at each assessment cycle (though adjustments/corrections could be made, as required).

### AUTOMATIC ASSESSMENT OF HIGHER-LEVEL EFFECTS

As discussed previously, the assessment of L3 effects is performed directly by the SMEs in the assessment team. These assessments consider performance at the most granular level. Since the overall assessments at the L1 and L2 levels are dependent on these L3 assessments, it is proposed that L2 and L1 performance ratings be determined automatically by rolling up or combining the ratings of performance and impact for each of the lower level effects. As illustrated in Figure 5, each L2 effect would be determined from the rating of the L3 effects associated to it through an automatic roll-up process; similarly, each L1 effect would be determined from the rating of the L2 effects associated to it.

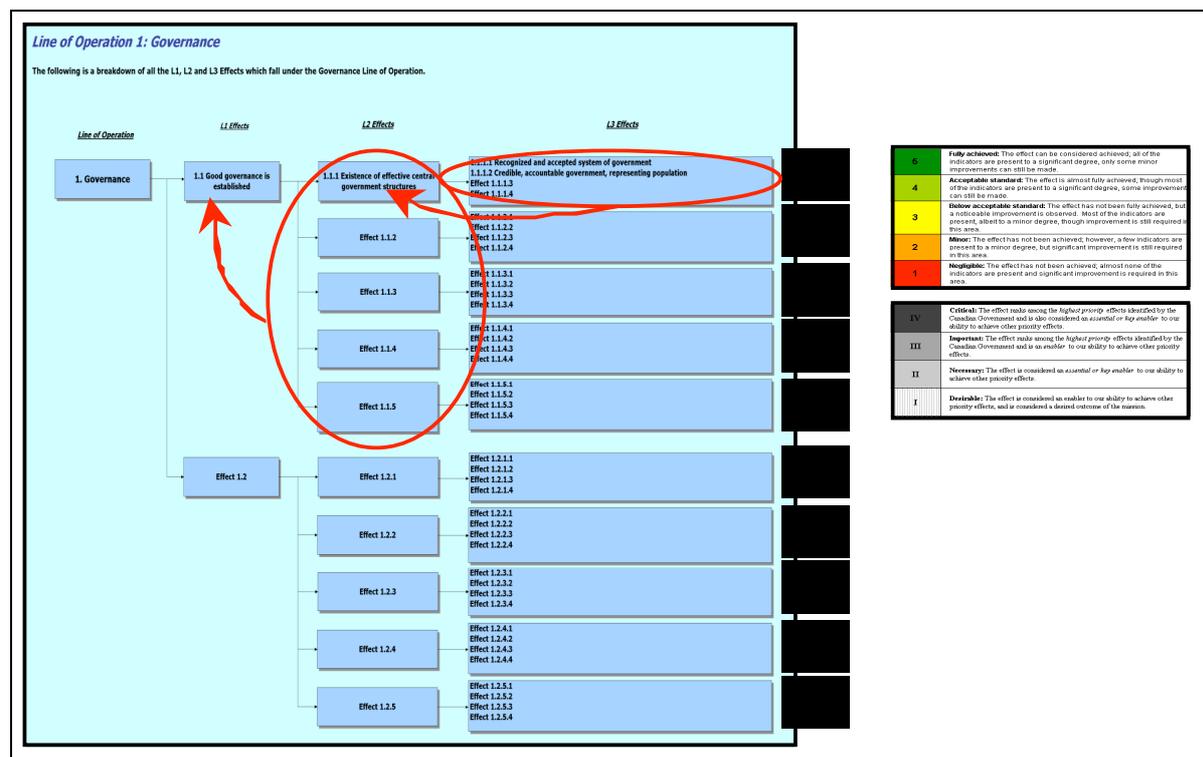


Figure 5: Automatic Assessment of Higher-Level Effects.

One possible set of ‘rules’ governing the roll-up is illustrated in Figure 6. This represents a very straightforward assessment of higher-level effects that corresponds to the natural interpretation and intuitive weighting of the effects by the SMEs. In the roll-up, effects rated as Critical and Important would exert a higher influence than the others, and the result of the roll-up would yield a performance rating along the original five-point scale. The roll-up would automatically be done within the data management tool and all results would be tracked over time.

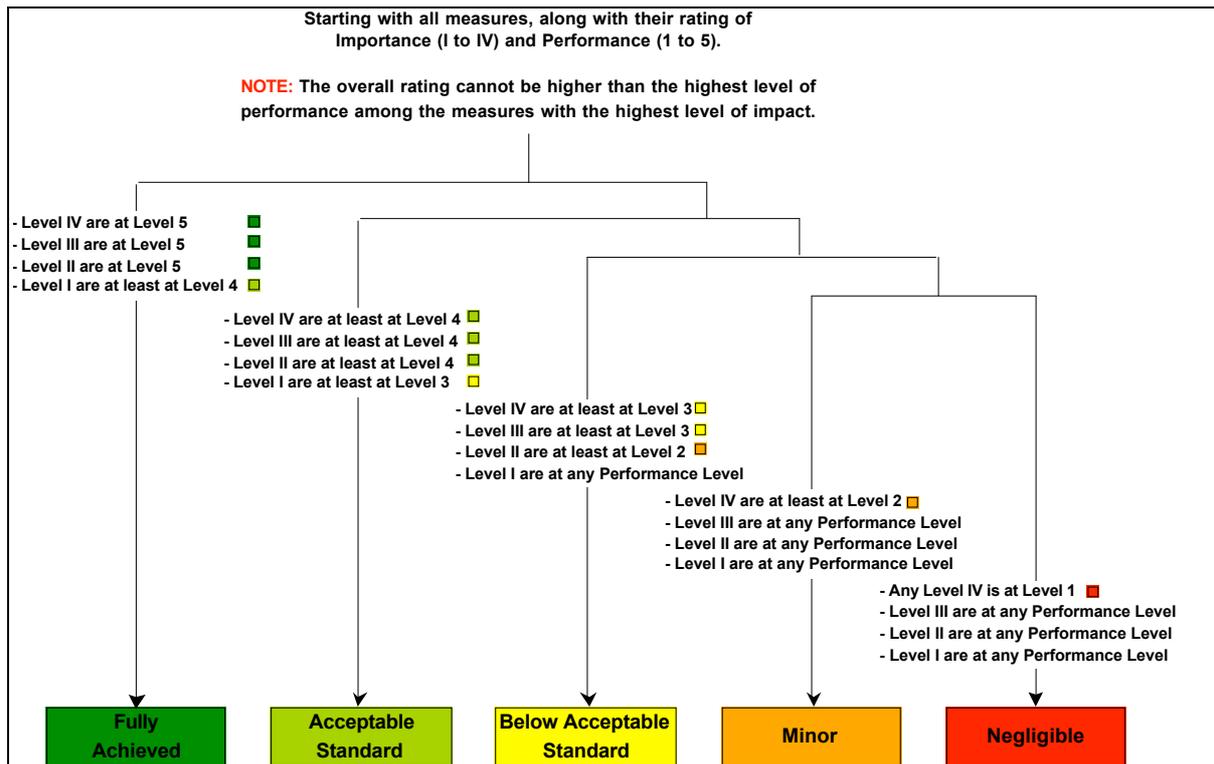


Figure 6: Possible Set of Rules for Automatic Assessment of Higher-Level Effects.

A benefit of the automatic assessment is that it would allow for a consistent assessment across all effects at a given level based on the application of a common set of rules on the lower-level effects. Furthermore, though the assessment of the higher-level effects provides less visibility into the reason a given effect is rated as it is (which can be best seen by looking at the lower level effects), it nonetheless provides an indication of the performance of a broader category of effects, which can be beneficial for flagging areas of risk and prompting the audience to investigate further. Finally, the roll-up mechanism combined with the data management tool provides the ability to drill down to successive levels to explain the scores observed, as well as to support analysis by investigating links, relationships, and temporal variations.

### NEW INTERFACE AND DATA MANAGEMENT TOOL

The final characteristic of the new Effects Dashboard to be discussed herein is the new interface and data management tool. The tool has been streamlined and consists of a more compact and user-friendly interface that can be used both in hardcopy and softcopy forms. The tool was developed in EXCEL. EXCEL was chosen because it was found to have the required functionality (particularly with the ability to program using Visual Basic for Application (VBA)), users were familiar with the software, and it was available to all participants in and out of theatre.

The main components of the tool are the three main worksheets, one for each of the three LOOs. A snapshot of these worksheets is shown in Figure 7. For each LOO, the entire tree structure (all L1, L2, and L3 effects) is presented in one page, along with the performance

rating scale and the results of all past assessments (including assessor comments). Although not represented in the Figure, the worksheet also includes a section for the recording of comments related to the assessment of L3 effects, as well as dedicated cells to store the rating of impact assigned to each of the effects. Finally, it combines all performance and impact ratings and performs the automatic assessment of higher-level effects.

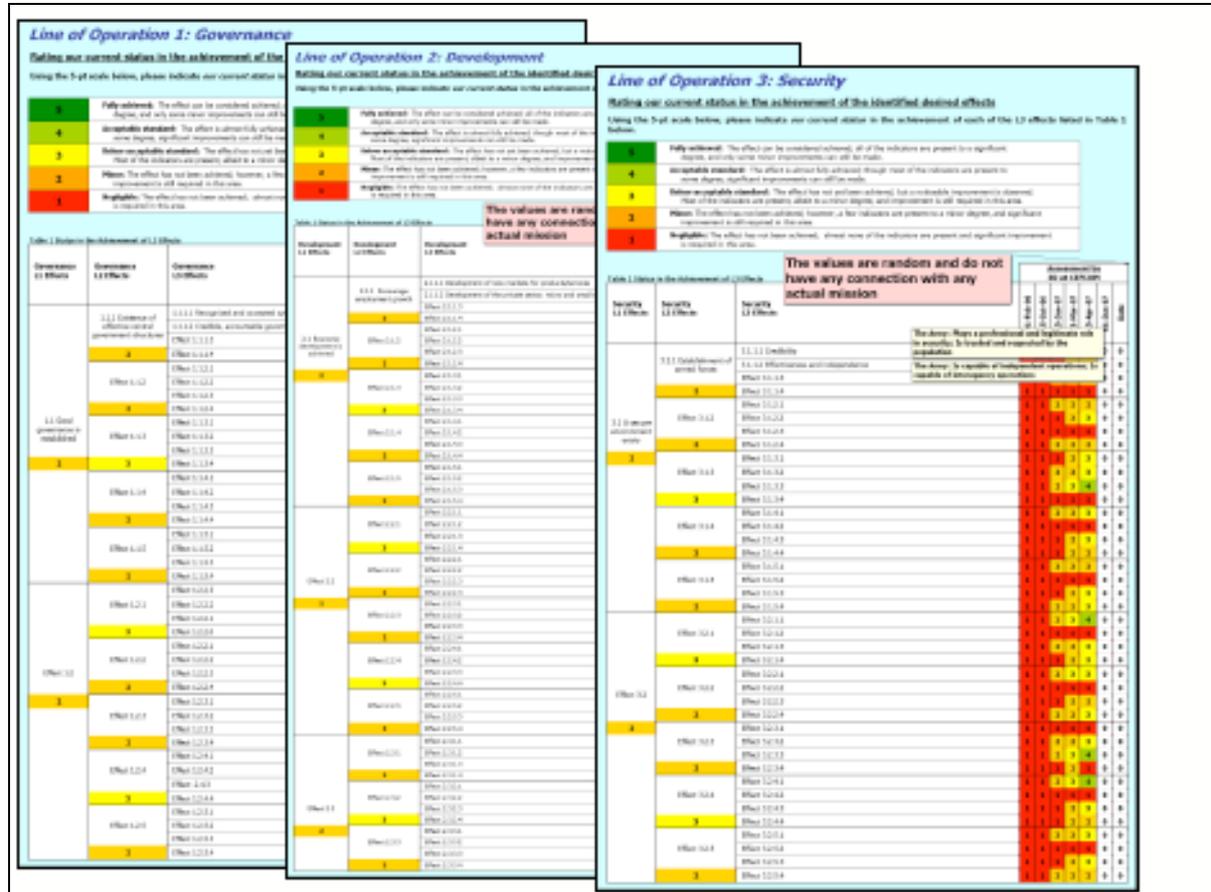


Figure 7: New Interface and Data Management Tool. Note: Diagrams contain no real data; they are for illustrative purposes only.

Overall, the new interface is more intuitive and user-friendly, the relationships between effects are clearly represented, and the tool renders the assessment exercise easier and quicker to conduct.

### THE WHOLE OF GOVERNMENT APPROACH TO ASSESSMENT

To date, there has been much scepticism as to how a single process will allow all departments to have their individual values and interests adequately reflected, while still uniting the different points of view. It is envisaged that this tool will allow each group to retain their independence while bringing the groups together to benefit from possible synergies, and to correct any differences in opinion.

But how can different groups use this one overarching tree structure to do campaign assessment, each with their own objectives? For example, how can it be used by CEFCOM,

by DFAIT, by the tactical level commander, and by the Government of Canada? The idea is as follows: The modular nature of the effects in the new Effects Dashboard, from L1 down to L3, should allow for the natural assignment (or ownership) of particular effects to the most appropriate stakeholder. Though only one of the many possible approaches, the following is an example of how the new Effects Dashboard may be used to enable a Whole of Government approach to assessment:

1. First, each group would identify the effects within the overall tree structure that it is interested in tracking as part of its own campaign assessment (these will likely be the effects related to the areas on which their efforts will be focused). For example, the effects that CEFCOM identifies may reside primarily within the LOOs related to security and governance, and perhaps mostly at the operational level (though they may have an interest in tracking other levels as well). DFAIT may select effects predominantly in the LOOs related to governance and, potentially, development.
2. Next, when each of the effects is assigned to a primary stakeholder, the maintenance and assessment of that effect would then rest officially with that stakeholder. Other stakeholders would be free to assess all effects, if they so chose, or to simply assess the effects that they have selected that are of interest to them.
3. Stakeholders would then identify the source of the information that would be used to assess their subset of effects. Information sources would be presented for approval to the rest of the assessment team. This activity would serve to validate the quality of the information to the other stakeholders, as well as to the outside community; it would ensure both transparency of the process and accountability of the assessors.
4. Prior to each assessment cycle, all stakeholders would provide their assessments (for their subset of effects) to a central agency (most likely within CEFCOM) that would combine all results into one consolidated Effects Dashboard. This Dashboard would then be fed back to all participants at the assessment working group meeting for further discussion and validation. Finally, as per the current process, the output of this working group meeting would be the final approved Dashboard for that given assessment cycle.

In addition to the approach outlined above, each of the stakeholders would also be responsible for providing a rating of *impact* for each of the effects in the new Effects Dashboard. A comparison of these ratings across stakeholders would provide very valuable information as to possible areas of agreement, or areas of friction that may arise from differences in opinions as to the impact of particular effects. In cases where the impact of an effect is rated differently across stakeholders, this would signal possible areas of disagreements that could be discussed and resolved immediately. Similarly, where there is agreement, this would highlight areas that are unilaterally assessed as having high or low impact and this may help justify the assignment of resources to these areas.

In the process proposed above, the new Effects Dashboard is simply a starting point, i.e. an enabler for the coordination of effort across the different stakeholders. Work remains to be done to confirm the effects included in the Dashboard, particularly within the Governance

and Development LOOs. However, it is felt that the tool is flexible enough to accommodate the individual values and priorities of the multiple stakeholders - provided of course that the resources and scope to cooperate are present.

## CONCLUSION

Though there is a tendency to develop reporting structures and performance measurement methodologies to suit one's direct needs, in some cases, commonality of tools *is necessary* and needs to be pursued to support unity of effort within the theatre of operations. One such case is in the assessment of progress in an operation in which more than one group plays an important role. Despite the difficulties associated with bringing together stakeholders with different objectives and, more importantly, with different organizational cultures, the benefits in doing so are immeasurable. In fact, it can be argued that *not* bringing all players to the table will guarantee failure of the overall mission or, at a minimum, will guarantee that the limited resources available to each stakeholder will be used inefficiently or ineffectively. The work presented herein is one possible mechanism to enable the alignment of objectives and the coordination of activities across stakeholders by creating a common methodology and assessment mechanism.

In strengthening the CEFCOM Effects Dashboard, the initial intent was to review the set of measures and fill in gaps, to strengthen the relationship between the measures and the LOOs at the tactical, operational and strategic levels, to make the evaluation process less variable and more robust, and to improve our ability to track changes over time and extract possible links between our actions and the effects observed. Using this process cooperatively and comparing ratings across the different stakeholders would provide very valuable information as to possible areas of cooperation or areas of friction between the different groups. At a minimum, the process would prompt discussion and promote a better understanding of the objectives of each group. Though still in its nascent stage, and with a lot of work yet to be done, the new Effects Dashboard has the potential to become an effective methodology for the assessment of effects-based operations based on a Whole of Government approach.

## REFERENCES

1. Anderson, J., 2000. *Military Operational Measures of Effectiveness for Peacekeeping Operations*, Naval War College, Newport RI, Joint Military Operations Dept, ADA381728.
2. Bowman, C.W., 2002. *Operational Assessment – The Achilles Heel of Effect-Based Operations*, Naval War College, Newport RI.
3. Burkett, W.H., 2003. *Assessing the Results of Effects-Based Operations (EBO): The Relationship Between Effects-Based Operations and the Psychological Dimension of Warfare*, Army War College, Carlisle Barracks PA.
4. Cares, J.R., 2005. *New Metrics and MOEs for Unmanned, Distributed Forces*, Alidade Inc, Newport: ADA448211.
5. Chiarelli, P.W. (MBen), Michaelis P.R. (Maj), 2005. Winning the Peace, The Requirement for Full-Spectrum Operations, *Military Review*, Issue July-August.

6. Clark, H.W., Wallfesh, S.K., 1995. *Measuring Effectiveness of Theatre IW/C2W Campaigns*, Dynamics Research Corp, Wilmington MA, ADA330423.
7. Combs, I.I., Ray, A., 2006. *Assessing Strategic Effectiveness in the War on Terrorism*, Army War College, Carlisle Barracks PA, ADA448536.
8. Cordesman A.H., 2004. *The Post-Conflict Lessons of Iraq and Afghanistan – Testimony to the Senate Foreign Relations Committee*, CSIS Report.
9. Courtney, M., Riddell, H., Ewers, J., Linder, R., Cohen, C., 2005. *In the Balance - Measuring Progress in Afghanistan*, Center for Strategic and International Studies.
10. Davis, P.K., 2001. *Effect-Based Operations: A Grand Challenge for the Analytical Community*, RAND Report.
11. Denhard, D., 2005. *Assessing Effects-Based Operations*, Air Force Institute of Technology, Wright-Patterson AFB OH, ADA448066.
12. Goodwin, W.P., 1995. *Measures of Effectiveness. The Transition from Peace-Enforcement to Peacekeeping*, Naval War College, Newport RI, Joint Military Operations Dept, ADA297892.
13. Green, J.M., 2001. *Establishing System Measures of Effectiveness*, Raytheon Systems Co, San Diego Naval and Maritime Systems, ADA405408.
14. Horowitz, M.D., 2004. *ROE - Are We Falling Victim to the Enemy's Effects-Based Operations?*, Naval War College, Newport RI, Joint Military Operations, ADA422829.
15. Jones, D. D., 2006. *Understanding Measures of Effectiveness in Counterinsurgency Operations*, Army Command and General Staff College, Fort Leavenworth KS, School of Advanced Military Studies, ADA450589.
16. Jones, R.D., 2006. *Assessing Resource Value and Relationships Between Objectives in Effects-Based Operations*, Air Force Institute of Technology, Wright-Patterson AFB OH, School of Engineering and Management/Department of Engineering Physics, ADA449610.
17. Kellogg, R.H., 2004. Evaluating Psychological Operations: Planning Measures of Effectiveness, *Special Warfare: The Professional Bulletin of the John F. Kennedy Special Warfare Center & School*, Vol. 16 Issue 4, p32-37.
18. Lopez, R., 2006. *On Learning: Metrics Based Systems for Countering Asymmetric Threats*, Army Command and General Staff College, Fort Leavenworth KS, School of Advanced Military Studies, ADA450454.
19. McCrabb, M., 2005. *Effects-Based Assessment*, DMM Ventures Inc, Yorktown VA, ADA445849.
20. McKeever, Jr., William E., Gilmour, D.A., Hillman, R. G., 2006. *An Approach to Effects Based Modeling for Wargaming*, Air Force Research Lab, Rome NY, Information Directorate, ADA451887.
21. McKeever, Jr., William E., Walter, M.J., Gilmour, D. A., Hanna, J.P., 2006. *Simulating Effects Based Operations*, Air Force Research Lab, Rome NY, Information Directorate, ADA451888.
22. Morell, T.O., Wong, E.Y., Goerger, S.R., Kwinn, C.M., Dodge, R.C., Goerger, C.S., McGinnis, M.L., 2006. *Effect-Based Assessment Support System (EBASS)*, OP RES Center of Excellence, Tech Report DSE-TR-0539, ADA448132.
23. Nelson, J.J., Newett, S.L., Dworken, J.T., McGrady, K.A., LaMon, K., 1996. *Measures of Effectiveness for Humanitarian Assistance Operations*, Center for Naval Analyses, Alexandria VA, ADA362468.
24. Patel, S., Ross, S., 2007. *Breaking Point - Measuring Progress in Afghanistan*, Center for Strategic and International Studies.

25. Perl, R., 2005. *Combating Terrorism: The Challenge of Measuring Effectiveness*, Library of Congress, Washington DC, Congressional Research Service, ADA444930.
26. Pusateri, A.E., Berg, T.S., Donlin, M.W., Thompson, D.P., Harrison, W.K., Kehrer, W.M., Guthrie, D.J., Pusey, C., Taylor, A., Schnur, M.T., Dittrich, S.M., 2004. *Metrics to Monitor Governance and Reconstruction in Afghanistan: Development of Measures of Effectiveness for Civil-Military Operations and Standardized Tool to Monitor Governance*, US Civ Affairs and PSYOP Command, Fort Bragg NC, USACAPOC Tech Rep. 04-01.
27. Schneider, D.A., 2002. *Measuring The Intangible: How Can An Operational Commander Know If His Information-Based Operations Are Effective?*, Naval War College, Newport RI, Joint Military Operations, ADA406201.
28. Sproles, N., 2001. *Establishing Measures of Effectiveness for Command and Control: A Systems Engineering Perspective*, Defence Science and Technology Organisation, Salisbury Australia, Electronics and Surveillance Research, ADA392037.
29. Stapleton, B., 2003. *A British Agencies Afghanistan Group Briefing Paper on the Development of Joint Regional Teams in Afghanistan*, British Agencies Afghanistan Group.
30. Thoele, B.A., 2006. *A Methodology for Performing Effects-Based Assessments*, Air Force Institute of Technology, Wright-Patterson AFB OH, School of Engineering and Management, ADA446340.
31. Umstead, R.K., 2005. *Effects-Based Decision Making in the War on Terror*, Air Force Institute of Technology, Wright-Patterson AFB OH, School of Engineering and Management, ADA435165.
32. Umstead, R., Denhard, D., 2005. Viewing the Center of Gravity through the Prism of EBO, *Military Review*, Issue Sept.
33. Wadsworth, M.P., 2006. *Winning the War of Ideas: Assessing the Effectiveness of Public Diplomacy*, Army War College, Carlisle Barracks PA, ADA448271.
34. Wagenhals, L.W., Levis, A., Haider, S., 2006. *Planning, Execution, and Assessment of Effects-Based Operations (EBO)*, George Mason University, Fairfax VA, Center of Excellence in Command Control Communications and Intelligence.
35. Wagner, R., 2004. "Virtual Victory": *The Influence of "Post" Modern Warfare on Post Conflict Operations: Is Defeat Necessary to Start "Phase IV"*, Army Command and General Staff College, Fort Leavenworth KS, School of Advanced Military Studies, ADA435839.
36. Wall, A.N., 1997. *Situational Awareness in Military Operations Other Than War: A Look at Measures of Effectiveness in Humanitarian Assistance Operations*, Naval War College, Newport RI, ADA325055.
37. Young Jr., W.E., 2005. Discovering the Effect-Endstate Linkage: Using Soft Systems Methodology to Perform EBO Mission Analysis, *Proceedings of 10<sup>th</sup> International Command and Control Research and Technology Symposium*, ADA464121.
38. 2004. *Metrics to Monitor Governance and Reconstruction in Afghanistan: Development of Measures of Effectiveness for Civil-Military Operations and a Standardized Tool to Monitor Governance Quality*, United States Combined Forces Command / ROK Operational Analysis Group, Sand Francisco, ADA422444.
39. 2005. *Post-Conflict Reconstruction Essential Tasks*, Office of the Coordinator for Reconstruction and Stabilization, US Department of State.
40. 2006. *Provincial Reconstruction Teams in Afghanistan, An Interagency Assessment*, USAID Report.
41. 2006. *Afghanistan National Development Strategy - An Interim Strategy for Security, Governance, Economic Growth & Poverty Reduction*.
42. 2006. *The Afghanistan Compact - Building on Success*, The London Conference on Afghanistan.

43. 2006. *Afghanistan in 2006 - A Survey of the Afghan People*, The Asia Foundation.
44. 2006. *Measuring Stability and Security in Iraq*, Report to Congress in Accordance with the DoD Appropriations Act 2006 (Section 9010).