

Measuring and Evaluation

Paul Clarke and Thomas Davies

Defense institution building (DIB) seeks to produce relevant institutional change with partners by addressing complex problems in dynamic environments. In order to determine their impacts, these changes must be continuously monitored and their effects, positive and negative, evaluated. However, the methods and techniques commonly used for the monitoring and evaluation of security cooperation activities often prove inadequate to produce the information required for thoughtful DIB decisions. Further, information requirements differ between those necessary for the development and justification of DIB authorities, policies, resources, guidance, and programs (the “DIB Enterprise”), and the management toward DIB outcomes by programs and other implementers (“DIB Activities”). Evidence from measures at the activity level will heavily influence decisions at the DIB Enterprise level. In addition, measures to support monitoring and evaluation must enhance the effectiveness of DIB without over-burdening the limited capacity of the small footprint, high impact teams that carry out the work. This chapter will first address the role that measures play in DIB decision-making and the unique complexities of measuring in the DIB context. The importance of integrating measures into every aspect and phase of DIB will then be discussed, as well as what appropriate measures should be for DIB activities.

Terminology

The term “measure” in this chapter refers to the quantitative or qualitative description of an input, output, or outcome of DIB policy, programs, or activities. Measures are often further defined as measures of performance, which seek to describe progress toward an objective, or measures of effectiveness, which seek to describe the degree to which an objective was met. The term “metric,” which is not used in this chapter, is often used interchangeably with “measures,” but metrics can be understood to refer to the combining of multiple measures to provide information (examples include rates such as “cost per hour” and changes from a baseline such as “increases in staff capacity”). For simplicity, the use of the term “measure” in this chapter is intended to capture the wide array of techniques available to process, analyze, and present measures, inclusive of the use of metrics. “Accountability” and “learning” are the primary purposes of DIB measures. “Accountability” refers to measures that seek to compare DIB inputs, outputs, or outcomes with expectations for these. “Learning” refers to measures that seek to improve the effectiveness or efficiency

of DIB policies and practices. “Monitoring” and “evaluation” are processes that use measures to inform the decisions required to manage DIB policy, programs, and activities. “Monitoring” occurs on an ongoing basis for DIB programs and activities, and is generally accomplished through internal resources (or in conjunction with those of partners). “Evaluations” seek to answer specific questions related to DIB policy and programs, occur in conjunction with specific events (such as the end of an activity), and often include some degree of independence from the programs and activities they target.

What Should Measures do for DIB?

DIB measures should enable evidence-based decisions. Evidence-based decisions about policy, programs, and practices are those that are grounded in research and informed by experiential evidence from the field.¹ These decisions can support DIB policy, programming, priorities, and resources, as well as the management of DIB activities, and improve their outcomes. The production of experiential evidence requires deliberate and considered monitoring and evaluation of the progress and effects of DIB activities. As such, the most critical aspects of integrating effective measures for DIB is to anticipate the decisions that need to be made, who will make the decisions, when the decisions will be made, the information required to make the decisions, the level of fidelity required from the information, and how the information needs to be presented. Further, the information gathered needs to be linked to decision makers at the right time and in the right way to have the intended influence on outcomes. While these decisions may not be apparent or even known at the onset of DIB activities, the more insight gained and guidance provided as early as possible regarding requirements for experiential evidence, the better information can be provided for DIB decision makers.

The inclusion and implementation of measures supports both accountability and learning. *Figure 1* shows examples of information requirements for accountability and learning that can inform decision-making for the conduct of DIB activities and management of the DIB Enterprise. Accountability in the DIB context should explain what has happened, particularly to confirm or challenge assumptions and to determine to what extent objectives have been met. This has been critical in the problem-driven, iterative, and adaptive approaches that have proven successful in DIB in practice—where “failing fast” can be a virtuous component of the process.² The ability to rapidly determine what works and what does not, allows teams and partners (as well as DIB policy makers and planners) to focus on efforts that contribute to positive outcomes while divesting themselves from, or avoiding, efforts of limited relevance or effect. Learning in the DIB context describes a process of gaining knowledge or skill by practicing or experiencing something; measures facilitate learning by demonstrating patterns and trends both within and across DIB activities, programs, and contexts. In the dynamic environments where DIB takes place, learning occurs across many different cycles, from insight into how the “political will” or “absorptive capacity” of a partner will affect the design of an activity, to the long-term effects of institutional capacity on Foreign Military Sales decisions. Varying cycles

of learning require measures with feedback loops that are appropriate for the decisions and outcomes they are meant to support. Learning at the project level may require the immediate judgment of a team on the ground. While learning at the program level, such as for the refinement of practices, may need to be widely and deliberately vetted to ensure consideration across various contexts and appropriate application. Learning at the policy level may require information across multiple years to inform and justify change(s), such as the need for new DIB authorities or programs.

Figure 1: Typical DIB Information Requirements

DIB	Accountability	Learning
Enterprise	<ul style="list-style-type: none"> ▪ DIB's Relevance/Impact ▪ Prioritization of DIB Activities ▪ DIB Resource Planning and Management ▪ DIB Program Effectiveness 	<ul style="list-style-type: none"> ▪ What Works in What Contexts ▪ New DIB Programs/Authorities ▪ Integration of DIB and Security Cooperation/ Building Partner Capacity ▪ Engaging with the DIB Community of Practice/Interest
Activities	<ul style="list-style-type: none"> ▪ SMART DIB Objectives ▪ Leading Institutional Change: Performance - are we doing things right? Effects - are we doing the right things? ▪ Efficiency, "Bang for the Buck" 	<ul style="list-style-type: none"> ▪ Scoping New or Existing DIB Lines of Effort ▪ Testing Theories of Change ▪ Adapting to Dynamic Environments ▪ Disseminating Lessons ▪ Refining Practices

At its best, DIB is a facilitated process of partner-led institutional change toward mutually beneficial outcomes, and the role of the partner is therefore an indispensable aspect of DIB measures and evaluations. That said, the practice of partner ownership presents challenges to measures. There are some elements of performance and effectiveness that a DIB team can be directly responsible for, but much of it they cannot. To the extent we hold DIB teams accountable, perverse incentives can emerge for implementers to do the work for the partner—undermining the enduring and institutionalized change that comes from partner ownership. The challenge is therefore to have multiple tiers of measures and show their overlap: so an outcome for the DIB team is really only an output from the partner's perspective. Further, the partner's active involvement in monitoring and evaluation may relieve some of the burden of information gathering for DIB practitioners, though it should be recognized that the partner's capability and capacity for monitoring and evaluation might need to be developed over time.

Integrating partners into the process of monitoring and evaluation can be a means of institutionalizing good governance and management. A partner's self examination of the decisions necessary to implement and institutionalize change, analysis of the information requirements to support these decisions, development of appropriate measures to provide the information, and management of the information as it is collected, will facilitate

implementation of the DIB activity. At the same time, this process enhances the partner’s ability to ensure changes endure and adapt with the partner’s defense establishment. Further, a practical implementation of measures can demonstrate the means of establishing transparency, accountability, and evidence-based decision-making more broadly for partners. As these practices are adopted, they become tools of good governance that address potentially sensitive issues, such as corruption, without directly taking them on.

Box 1: Hierarchy of Evaluation³

In a piece published in the Small Wars Journal, RAND Senior Scientist Christopher Paul describes the importance of explicit theories of change and a hierarchy for the evaluation of planning and assessments. He explains that “having an explicit theory of change helps the assessor identify what to measure and supports an assessment structure that tests assumptions as hypotheses, which leads to improved operations, outcomes, and assessments.” He also describes a hierarchy of evaluation with five nested levels that “helps the assessor match assessments to stakeholder needs and identify where improvements are needed when operations do not produce desired outcomes”:

Level 1 - Assessment of Need for Effort
<ul style="list-style-type: none"> ▪ Focuses on the problem to be solved or goal to be met, the population to be served, and the kinds of services that might contribute to a solution. ▪ Establishes the problem to pursue, as well as the intended objectives.
Level 2 - Assessment of Design and Theory
<ul style="list-style-type: none"> ▪ Focuses on the design of a policy or program, and is where an explicit theory of change should be articulated and assessed. ▪ If program design is based on poor theory or mistaken assumptions, then perfect execution may still not bring desired results. ▪ If the theory does not actually connect the activities with the objectives, efforts may generate effects other than those intended.
Level 3 - Assessment of Process and Implementation
<ul style="list-style-type: none"> ▪ Focuses on program operations and the execution of the elements prescribed by the theory and design. ▪ Poor execution can foil the most brilliant design. ▪ Traditional measurements at Level 3 are measures of performance (MOP).
Level 4 - Assessment of Outcomes
<ul style="list-style-type: none"> ▪ Focus is on how outputs are translated into outcomes or achievements. ▪ Outputs are the products of activities, and outcomes are the changes resulting from these efforts. ▪ Measures at this level are often referred to as measures of effect (MOE).
Level 5 - Assessment of Cost-Effectiveness
<ul style="list-style-type: none"> ▪ Only when desired outcomes are at least partially observed can efforts be made to assess their cost-effectiveness - before you can measure “bang for buck,” you have to be able to measure “bang.”

What is Unique about Measuring in DIB?

The unique nature of the cooperation between the United States Government (USG) and partner nations poses challenges for measurement across the DIB effort such that several factors must be considered in the tailored development of DIB measures, including: institutional perspective, various actors, motivations, political will, cultural context, the low-capability of some partner countries, and the fact that teams will be dealing with a mix of defense governance systems. At the DIB Enterprise level, the stakeholders are many, including senior policy makers, legislators, and multiple Department of Defense (DOD) agencies, each of which has a different perspective on success, requiring measures that accommodate these perspectives. In the case of the Security Governance Initiative—a DIB-related program in Africa—the work is inherently interagency, and the lead is with the State Department rather than the DOD. At the program level, measures have to be agreed upon by numerous USG actors overseeing the projects, while the U.S. expert teams that actually execute the projects may draw expertise from multiple think tanks, private contractors, and other providers. The result is often a tenuous relationship between the measures envisioned in the project's design, execution, and assessment.

The perspectives and motivations of the multiple players in a DIB project complicate the conceptualization of measures, although all seek to achieve success via institutionalized improvement. The United States has its own internal measures related to budgets, programs, competing security interests, and methodologies, while the partner has its own internal measures, which may stem from its motivation to secure the ongoing engagement of the United States and maintain the flow of U.S. training and equipment. Another challenge is determining whose work is to be measured. DIB work is produced by three broad categories of workers: the U.S. experts and the partner nation each working separately, and the U.S.-partner teams working together. As mentioned before, measures for each of these need to overlap to create the appropriate incentives for partner ownership and enduring change.

Typifying a DIB Engagement is also difficult, since the roster of partner nations has grown rapidly. The result is a diverse body of engagements in different security environments and socio-economic realities that runs from Bangladesh to Kosovo, and from Colombia to Guinea. For the DIB Enterprise, this manifests itself in finding measures that have legitimacy in many different contexts and an understanding that rolling up or aggregating measures without these contexts may result in a loss of significance. This diversity is less of a challenge at the program level, since it is to be expected that each partner will have its own context. Still, DIB partners are chosen for a reason—to cure a particular challenge, to take advantage of an opportunity, or to shore up a partner in a tough security environment.

Every partner has its own context, including some level of internecine conflict from the banal (interagency) to the truly divisive (insurgency or sectarian conflict). And, of course, the motivations of the partner, including the perspectives of different elements of a partner's government, remain an important factor. In the world of political insight, U.S. DIB practitioners may be able to discern basic truths and speak the unspoken, but they remain novices in the partner's political setting. It can be tough to discern: what is the

unstated priority of the leadership? Which accommodations should not be tampered with? Which graft grinds the system down and which is lubricating the gears? In an environment where appropriate institutional fixes must be married with political reality, powerful measurements of progress can be difficult to create.

U.S. experts face several imperatives when designing projects and the measures to evaluate their effectiveness. The “do no harm” imperative is always in play, but for countries that face an active threat there is a corollary of “fix the system without breaking it.” The work in such cases should consider how to determine when the work impedes the running of operations. The U.S. experts may not perceive when they cross such a threshold, but the partner may also not be fully aware of the trade-offs that are being made. For instance, how much the partner staff’s work on DIB soaks up energy from other efforts, how support systems might be stressed by changes, or how adjusting budgets may affect operations.

Measuring progress and evaluating success can also be challenging in countries where a multitude of outside actors are advocating for some form of political or security sector reform. External initiatives might come from interested states, regional bodies, or the United Nations. In Western Africa for example, we see French, European Union, and United Nations’ efforts, together with regional (African Union) and subregional (Economic Community of West African States) efforts to produce defense reform. DIB efforts can aid in these reforms, and the synergy between those initiatives and U.S. efforts can be powerful. Yet, when it comes to measurements, the cause of progress might be hard to discern since many countries may have set the stage for reform through long years of engagement, and the payoff may be manifested in part through the U.S. DIB effort.

Designing measurements in the early phases of a DIB project is similarly problematic, since discovering and incorporating into a baseline the existence of related efforts by other actors can be very time intensive. The partner might also struggle to meet tasks and deadlines because they are engaged in similar but different international efforts. One nation may have started with the French system of defense forces, then embraced the Soviet system for decades, turned back to the French and the United States and now find itself with Chinese equipment and training. Mixing these many defense systems poses a challenge to measurement design and collection, since different systems track and evaluate equipment using dissimilar tools, or have different approaches to human resource management. In the example of logistics, the partner may have received instruction in tracking equipment readiness from various donors, and these different systems may exist side-by-side in the same institution, hindering understanding of their own state of readiness. For outside experts, the design of measures can be complicated by these multiple standards, the complexity of which can only be revealed by in-depth study.

Inevitably, DIB efforts are increasingly taking place with low-capability countries in grave crisis, which have often become the recipients of large amounts of material assistance. In the past, countries where the U.S. was engaged in DIB-related efforts had more capacity to absorb the concepts and the workload involved in implementing DIB; take for instance Colombia, Chile, and former Warsaw Pact countries in Eastern Europe. While there are

advantages to working with low-capability countries—for example, there is low-hanging fruit to pick, and simple solutions can have big impact—the very concepts of DIB can be difficult to translate into such institutions, and devising useful measures for partners' success can therefore be challenging. For such low-capability countries, the U.S. DIB tools and methodologies are underdeveloped and untested. In time, the methodology will advance, and further studies may find that DIB work in this category of countries has both great risk and great payoff. If absorptive capability is a challenge for developed ministries of defense, it is doubly so with low-capability partners, who are already receiving so much material assistance that it poses a risk of choking the support system. Measures should seek to reveal the absorptive capacity of a partner to inform the levels of engagement that are effective for DIB, the capacity constraints most burdened by material assistance to focus effort, and the limits the partner faces in integrating and sustaining material assistance to inform future material assistance.

Low-capacity countries are often dealing with significant societal challenges. Tolstoy begins *Anna Karenina* by noting, “All happy families are alike; each unhappy family is unhappy in its own way.”⁴ DIB practitioners need to find the unique qualities that make a security establishment “unhappy,” and this involves understanding the nature of the defense institutions, of course, but also more broadly the society and political values and processes in the partner. On the societal side, when formulating measures, DIB planners need to consider how sectarian, social, and regional issues influence decision-making and resource allocation. The ongoing effects of these on DIB activities should be monitored to allow activities to adapt in order to remain relevant in dynamic environments. The work itself and the measures developed must take into account the subtle nuances that keep an institution together. Many motivations may come into play, and corruption (hidden rice bowls, ghost soldiers, etc.) is a reality one must consider when asking for institutions to measure their work or conduct other self-assessments. A bit of investigation will often find a lack of materiel, missing parts, inadequate training, faulty communications, insufficient logistics support, the wrong mix of capital assets and so on, but rarely are the cultural foibles in the institution evident upfront. In designing measures, DIB experts should consider where the power and authority reside, who has oversight, and how those qualities interact. Measures should be designed with foreknowledge of how they might produce winners and losers.

Measures are most useful when there is support for the DIB effort by the political leadership, rather than just the career defense officials. Political leaders can provide the authority and resources to ensure DIB success, since they represent the will of the people. Without such buy-in, measures may not become institutionalized or the information provided may lack key insights. Political buy-in aids the overall effort, ensuring that timelines are met, that replacements are found for DIB counterparts, and that the program's efforts remain consistent with the evolving priorities of the partner. Political clout can also overcome divergence between the political level and the operational level, and between different agencies and services, helping to ensure that reliable measures are developed.

Without that support, different processes between different ministries (Interior and Defense) may impede collection of information used for measurements or create confused data. Finally, lower level officials lack the authority to collect useful and sensitive data, so the political class can give these officials more power to perform that function. Political leaders should be brought into the process early on by commissioning the work, setting the terms of reference, reviewing progress, and approving and directing the implementation of any recommendations.

Integrating Measures into DIB

The effectiveness of DIB will be greatly increased if measures are integrated from the start into every aspect of DIB policy, program, and project methodologies—monitoring and evaluation should not be an afterthought. In fact, some information may be lost if baselines are not established and deliberate measures put into effect. Additionally, DIB measures should themselves be actively managed throughout all phases of projects to account for the complex problems and dynamic environments they address.

DIB outcomes center on change and measuring change requires the establishment of baselines as a point of reference for comparison. A baseline is a description of a condition at a point in time, ideally before the effects of DIB take place. Baselines often focus on those things that pose the greatest challenges to achieving desired outcomes. An example of this for DIB could be the “as is” process map for a partner’s annual budget, which includes information on how its inputs, outputs, actors, and decisions affect current outcomes. However, DIB baselines should also include the conditions that may indicate opportunities for action such as engaged and empowered senior sponsors, relevant dynamics, such as absorptive capacity, that may be favorable or unfavorable for change, and the people and politics that influence the DIB environment. Baselines are necessary for effective action, so they should be defined upfront, and fleshed out or clarified as the activity proceeds, as they can be difficult and costly to create after the fact.

All measures have a cost, and the cost of producing measures can have a tremendous impact on DIB operations. In the context of DIB, cost refers less to the financial impacts of monitoring and evaluation (though these should be considered) and more to the opportunity cost of producing the information. If the burden of collection and analysis of information for measures falls on the typically small DIB teams carrying out the work, then there must be consideration of the effect on the project before levying information requirements. The importance of the decisions, the level of detail needed to make the decision, the need for establishment of a baseline, and the frequency of updates should all be factored in and adjudicated before stakeholders request information about DIB activities. What DIB information is then produced should be widely available to satisfy the situational awareness and decision-making requirements of as many stakeholders as possible. Another potential cost is that measures can create perverse incentives for DIB practitioners, such as interfering with partner ownership, stifling learning and adaptation,

setting safe goals when meaningful change requires more, and standardizing practices when tailored approaches are appropriate. The potential perverse incentives formed by any measure should be considered and controlled for as the measure is developed and implemented. Finally, the type of information and the timing of the requirements can have an effect on the relationship between the DIB implementers and the partner's interlocutors. Building trust takes time, and it can be lost quickly or never achieved at all if the partner perceives the DIB team as "collectors." Information is not useful if it negatively affects the overall ability of the DIB activity to progress toward its ultimate outcomes.

Integrating Measures for the DIB Enterprise

Measures should support decisions regarding DIB authorities, policies, priorities, resources, and programs. Foundational to this is the establishment of enterprise-level DIB objectives that are nested in broader security cooperation objectives. The Specific, Measurable, Achievable/Agreed, Relevant and Time Bound (SMART) approach advocated for security cooperation planning, which links broad end states to concrete tasks, should carry through for DIB and its measures.⁵ DIB Enterprise-level objectives should provide sufficient direction for programs and activities to develop measures that provide the relevant information necessary for greater understanding and more effective management of DIB's array of complex challenges and dynamic environments. In addition to these objectives, policy makers and planners should articulate information requirements related to the performance of DIB programs, priorities, and resource allocations. The DIB Enterprise may also pool monetary and manpower resources to ask and answer specific evaluation questions through means more external than a program's own monitoring. Though the evaluation may be external to the DIB program, the collection of information should be integrated within DIB activities as much, and as early, as possible to increase the value of the information collected. The information gathered to support enterprise-level DIB objectives should drive ongoing and thoughtful dialogues among the decision makers who must link strategies to DIB programming and resources, and who must justify the results.

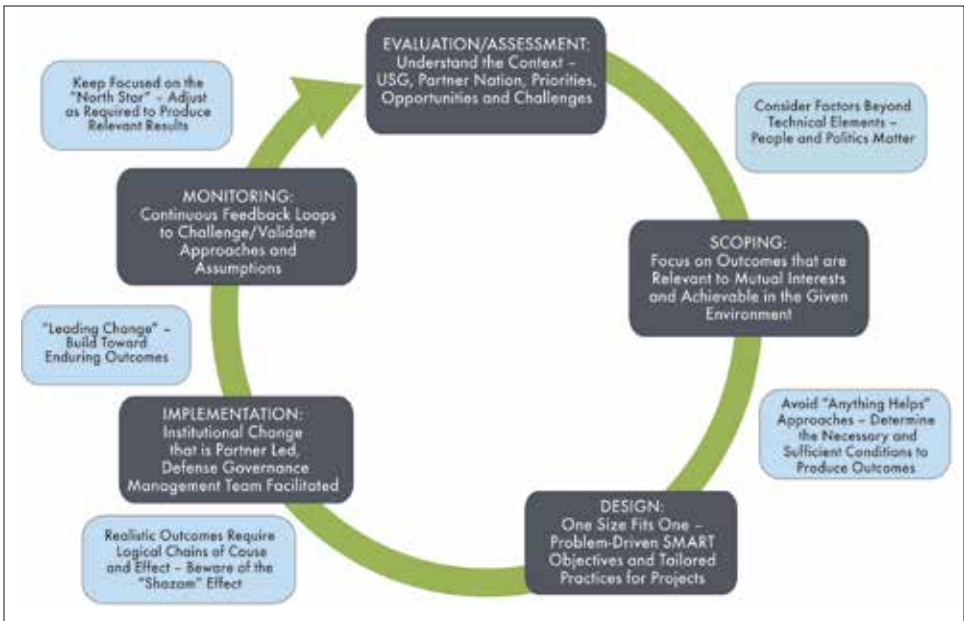
Measures can also be used to educate and inform the DIB community of interest.⁶ The DIB community of interest can include legislators, policy makers, planners, and implementers involved in related security sector development and security cooperation activities. This community can also include independent contributors such as academics, think tanks, and the media. Across this broad spectrum of stakeholders, it is important to establish what DIB can and cannot do, DIB's impact and contributions, and how DIB affects other activities. To this community, measures can be a powerful means of communicating the appropriate application and potential impact of DIB activities, as well as the effectiveness of practices for partners, environments, and outcomes sought.

Integrating Measures for DIB Activities

Measures to support DIB activities need to be integrated throughout the cycle of building partner capacity (see *Figure 2*). Though monitoring and evaluation are explicit components

of this cycle, their effectiveness will depend on deliberate actions taken to plan for and integrate measures throughout the earlier scoping, design, and implementation phases. Further, the DIB environment is dynamic and facilitating partner-led change is a process that highlights different aspects as it proceeds, so measures should be adapted with the environment and through implementation to support problem-driven outcomes.⁷ Security cooperation and development programs frequently have separate evaluation teams that come in one to two years after the end of the project to evaluate effectiveness. However, given the newness of the field and its characteristics (volatile, uncertain, complex, and adaptive), DIB activities have focused heavily on effectiveness in earlier phases—especially monitoring with immediate feedback loops into country projects and for program-level practices (such as for DIB’s overall methodology).

Figure 2: DIB Model for Building Partner Capacity⁸

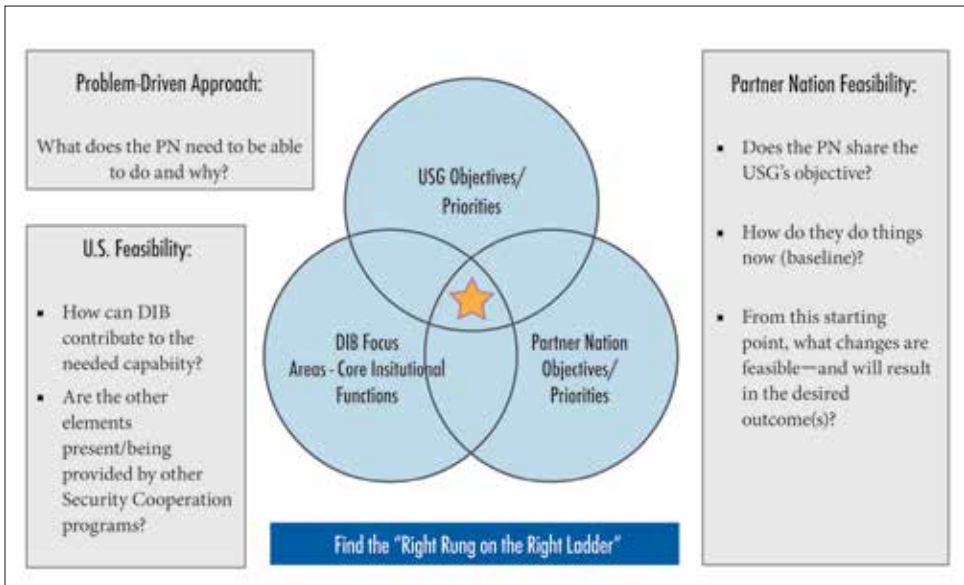


Scoping: Scoping is the process of transforming broad guidance to engage with a partner into a logical set of targeted and feasible recommendations that communicate the intent, and shape the design and implementation of DIB activities. As such, scoping establishes the foundation upon which the measures of DIB activities are built. Specifically, scoping begins to define the outcomes upon which to devise measures of progress and effect, and establishes the baselines that set the starting points for measuring change and defining what is feasible.

Scoping should, at a high level, define the intended and mutually agreed outcomes of a project. These should seek the “sweet spot” (see Figure 3) among U.S. government priorities, the objectives and priorities of our partners, and how institutional improvements

of defense governance and management can contribute to both. If realistic, these become the ultimate outcomes or “North Stars” that allow DIB projects over time to develop SMART objectives as part of the project design, implementation, and monitoring phases.⁹ From these project level outcomes, teams can begin to anticipate the information required to manage progress and determine the effectiveness of projects and their activities.

Figure 3: The “Sweet Spot” for DIB Objectives¹⁰



Scoping should also begin to establish the relevant baselines for a project. Baselines are a description of the relevant “as is” state of the partner’s institutions, opportunities for and challenges to mutual interests, and current level of performance relative to outcomes sought. Baselines should not only describe the current state of the partner’s institutional capabilities but also describe other factors related to the partner’s motivations for and limitations to change such as leadership, priorities, will, capacity, absorption, and spoilers. Baselines are the foundation from which any measures of change are built and are therefore paramount to managing progress and measuring the success of DIB activities.

Design: DIB projects are designed to address the unique institutional environment and outcomes sought with partners. In other words, when it comes to DIB projects, one size fits one. As such, measures must also be custom tailored to suit the needs of the project (See Box 2). Specifically, measures should be integrated into and support the logic (referred to as theories of change) for how teams intend to work with partners to achieve objectives.

Theories of change strive to incorporate and relate all of the necessary and sufficient elements required to achieve a desired outcome into a logical framework. They are generally expressed as a series of cause and effect or “if, then” statements that start from a baseline

condition and proceed to the outcome sought with no unrealistic assumptions or leaps of logic. Since these logic frameworks incorporate all of the elements to produce outcomes, they will not only include the technical elements of institutional development but also the human and political aspects of the institutional environment that accept, support, produce, and sustain the changes. An essential element of theories of change is that they are testable. If a portion of the logic proves wrong (either through a false assumption or a changed condition), then the theory must be adjusted to accommodate the new information. As such, measures must be developed to test the theories that underlie the design of DIB projects.

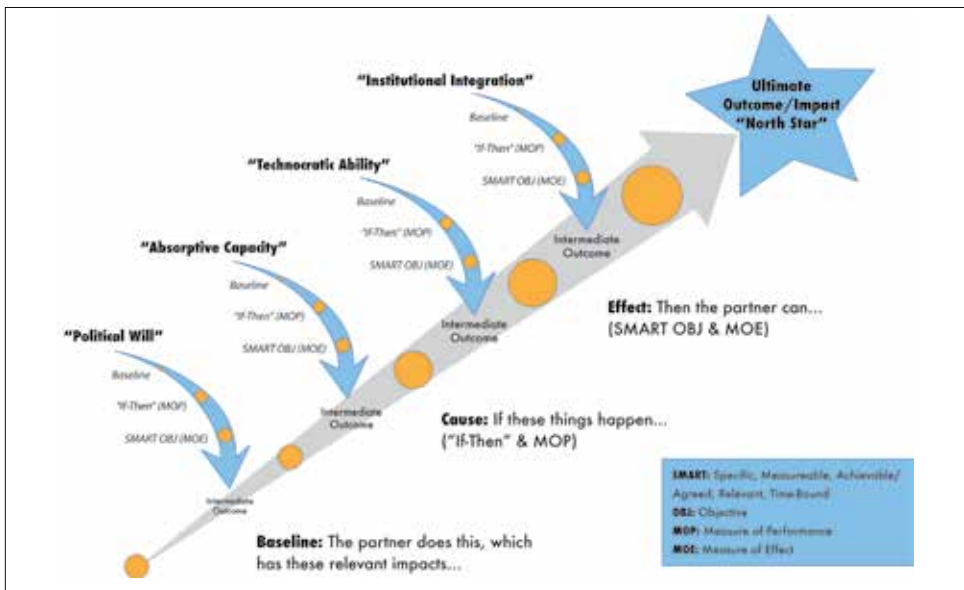
Box 2: Evaluating Complexity¹¹

In their work, Preskill and Gopal respond to a growing realization that, “systemic change is not linear, predictable, or controllable...problems are more resilient than previously thought and that traditional means of tackling them often fall short.” In response, they described characteristics of the complex systems that contribute to this realization (an accurate portrayal of the dynamic environments and complex problems DIB typically faces) and offered propositions for evaluation (measures) under such circumstances.

Characteristics of Complex Systems	Propositions for Evaluation
A complex system is always changing, often in unpredictable ways; it is never static	1. Design and implement evaluations to be adaptive, flexible, and iterative
Everything is connected; events in one part of the system affect all other parts	2. Seek to understand and describe the whole system, including components and connections
Information is the fuel that drives learning and helps the system thrive	3. Support the learning capacity of the system by strengthening feedback loops and improving access to information
Context matters; it can often make or break an initiative	4. Pay particular attention to context and be responsive to changes as they occur
Each situation is unique; best principles are more likely to be seen than best practices	5. Look for effective principles of practice in action, rather than assessing adherence to a predetermined set of activities
Different sources of energy and convergence can be observed at different times	6. Identify points of energy and influence, as well as ways in which momentum and power flow within the system
Relationships between entities are equally if not more important than the entities themselves	7. Focus on the nature of relationships and interdependencies within the system
Cause and effect is not a linear, predictable, or one-directional process; it is much more iterative	8. Explain the non-linear and multi-directional relationships between the initiative and its intended and unintended outcomes
Patterns emerge from several semi-independent and diverse agents who are free to act in autonomous ways	9. Watch for patterns, both one-offs and repeating, at different levels of the system

From theories of change, objectives can be developed that provide the “connective tissue” from baselines to outcomes, which guide activities to get from “here” to “there.” Measures must be developed to determine if and to what extent these objectives are met. This includes both intermediate outcomes that act as milestones toward the ultimate outcome, as well as measures of the ultimate outcome(s). All objectives, including intermediate ones, should strive to be SMART. Beyond the obvious “measurable” component, “achievability” and “relevancy” are critical to a project’s theory of change and may be subject to assumptions made or to the DIB environment. Time is also a measurable component (by either chronology or events) that can have significant impacts in terms of relevancy and relativity to other activities that may define a critical path for a project. For example, the sophistication of technical requirements may have to be balanced against the need to meet certain political deadlines.

Figure 4: Graphic Example of Objectives and Measures in a Logic Framework¹²



Implementation: The essence of DIB implementation is the facilitation of a partner-led change management process (see Box 3). As such, if implementation is when change happens, then implementation is when measures need to take place. Further, because change is a dynamic process, collection of information through implementation should not be static or risk not being relevant through the entire course of a project. For example, critical measures at the beginning of a DIB project may include partner leadership: assigning the correct stakeholders to the working group, committing to ensure continuity of personnel, taking briefings from the working group, and contributing to and making project design decisions. Do working group members show up, engage meaningfully, and undertake the agreed-upon work on the project between engagement visits? Measures midway through the

implementation of a project may focus on understanding and applying new concepts and approaches. The institutionalization of the changes to a partner’s supply chain management processes may be a more critical measure at the end of a DIB project. Measures through implementation should, to the greatest possible extent, be integrated into the conduct of a project so as to complement and enhance the work of DIB implementers, while providing the information needed for DIB decision makers. Therefore, frameworks and tools for the collection and management of a project’s measures should complement and be adaptive to the project’s needs (and other higher level DIB decision requirements) and highlight the information that is most relevant given where a project is along the change process.

Box 3: Leading Change¹³

Kotter’s well-known process of “Leading Change,”—recently updated to accommodate the context and objectives of a faster moving world (and more accurately reflect the complex problems and environments of DIB)—is useful for demonstrating the dynamic informational requirements to support the implementation of DIB activities. It is important to note that change for DIB is rarely linear as Kotter’s “steps” may suggest, and Kotter’s update describes running the steps “concurrently and continuously” in such cases:

Step 1: Create a Sense of Urgency	Craft and use a significant opportunity as a means for exciting people to sign up to change their organization.
Step 2: Build a Guiding Coalition	Assemble a group with the power and energy to lead and support a collaborative change effort.
Step 3: Form a Strategic Vision and Initiatives	Shape a vision to help steer the change effort and develop strategic initiatives to achieve that vision.
Step 4: Enlist a Volunteer “Army”	Raise a large force of people who are ready, willing, and urgent to drive change.
Step 5: Enable Action by Removing Barriers	Remove obstacles to change, change systems or structures that pose threats to the achievement of the vision.
Step 6: Generate Short-Term Wins	Produce, track, evaluate, and celebrate volumes of small and large accomplishments and correlate them to results.
Step 7: Sustain Acceleration	Use increasing credibility to change systems, structures, and policies that do not align with the vision; reinvigorate the process with new projects, themes, and volunteers.
Step 8: Institute Change	Articulate the connections between the new behaviors and organizational success, and develop the means to ensure leadership development and succession.

In order to manage the information from a DIB project, it is useful to have a systematic way of collecting and interpreting the data. It is also useful to have a means to assess the totality of what is being measured to determine if it meets the informational requirements of the DIB activity without exceeding the reasonable capacity of the implementers. There are various methods for this kind of data management and the methods and tools used should fit the needs and requirements of a particular project. For programs that use

theories of change in the development community, for example, this is often done through a “results framework” that provides a tabular format for succinct statements of data such as baselines, objectives, intended results, observations, and analyses.

Monitoring and Evaluation: Monitoring and evaluation of DIB activities involves gathering, processing, analyzing, presenting, and managing information to support evidence-based decisions. Though the groundwork for effective monitoring and evaluation is established in the scoping and design of DIB activities, the groundwork means nothing if it does not get the right information, in the right form, to the right people, and at the right time to make evidence-based decisions. Therefore, deliberate planning and management should be applied to the effective utilization of the information gained through the monitoring and evaluation of DIB activities (see *Box 4*).

Box 4: Evaluation in Fragile and Conflict-Affected Environments¹⁴

The UK’s Department for International Development (DFID) has “established a partnership with a consortium of leading organizations in the fields of conflict, security and justice to develop more effective approaches to the use of data in the design, implementation and evaluation of programs that contribute to reducing conflict, crime and violence.” This has yielded a significant body of knowledge of its own regarding fragile and conflict-affected environments in general and for the integration of measures in these environments specifically.

The DFID consortium advocates that integrating theories of change and measures provides planners with a basis to determine whether a project, program, or strategy is on track to accomplish the desired change, and if the environment is evolving as anticipated in the project or program design. DFID establishes that theories of change enable evaluators to ask hard questions about why certain changes are expected, the assumptions of how the change process unfolds, and which outcomes are being selected to focus on and why.

The work of the DFID consortium also proposes that the process of monitoring assumptions and theories of change involves an iterative cycle of regular data collection, analysis, reflection, feedback, and action. They hold that theory-based evaluation helps assess whether underlying theories of change or assumptions of a program are correct by identifying the causal linkages between different variables and is particularly useful for learning and accountability—whether the success, failure, or mixed results of the intervention were due to program theories and assumptions, or implementation. Finally, the DFID consortium asserts theories of change need to be as reflective of the actual environment as possible without overly complicating the situation—clearly defining the boundaries of the theory and its assumptions is critical.

The challenges that arise during DIB projects are often multi-causal and highly political, and may involve many people and organizations. Therefore it is important to

identify measures that indicate progress and adequate resolution. Testing theories of change and making adjustments to DIB activities, including the measures used to support them should be highlighted as features of the monitoring process (and captured for learning within the DIB activity and across DIB programs). Making adjustments to the plan should not be taken as indications of failure—failure would be to rigidly follow a plan and miss opportunities to achieve outcomes in the face of new information or changing circumstances. Therefore, evaluation must also take into account that DIB activities intentionally evolve beyond scoping and design, and throughout implementation, in order to achieve intended outcomes. Measures should be able to “tell the story” that justifies the evolution of a DIB activity by providing the evidence that led to the decisions to adjust course. Without this evidence, a DIB activity may be evaluated on objectives that were only relevant at the time they were established, decisions that may be scrutinized as arbitrary, and activities that are criticized for not having discipline or direction (when the discipline was in the deliberate exercise of agility in the face of changing circumstances or new information).

Appropriate Measures for DIB

Public policy deals with tough questions, such as: is NATO a success? One would answer this question differently at different times in history. Early on, NATO demonstrated the resolve of the United States in Europe, but during the Cold War it sometimes demonstrated the lack of Western unity and at times it was outflanked by Soviet initiatives. At the fall of the Berlin Wall and the dissolution of the Soviet Union, it appeared a great success. But soon after, its utility was in dispute again, while its value rose after the Russian Federation became more provocative. One could ask the same thing about DIB. Is DIB as an enterprise a success or is a specific DIB activity a success? What are the measures as a whole and by case?

One measure of the success of a DIB effort is to query the ultimate stakeholder of DIB work, the citizen of the partner nation. This construct of a measure of success could be evaluated through surveys, and that science has come a long way with accurate public opinion surveys being conducted in war-torn areas, such as Iraq and Afghanistan.¹⁵ One can even find data on how the defense institutions we work with are viewed by their consumers, their citizens.¹⁶ On the other hand, such surveys are certainly far removed by time and effort from any direct DIB projects, and the ability to disaggregate causes from effect can also be challenging.

Program evaluation is essential in public policy. DIB is a relatively new concept with much still to be learned or developed, but there is a growing basis of understanding about the challenges and some useful tools to deal with them. Among them are tools that rely upon timely and accurate qualitative and quantitative information, and a process that aims to monitor and evaluate intermediate objectives as the primary means to assess program success. It should be noted that, while there is great power in the analytics associated with quantitative measures, there is no inherent superiority of quantitative over qualitative

measures—each has its time and place. This is important since the nature of DIB issues and environments will often necessitate the use of qualitative information to support evidence-based decisions.

Quantitative Data Collection and Assessment

Though DIB environments often necessitate their use, DIB should not develop its own bias for qualitative measures when quantitative measures may be available and appropriate to the situation. At the enterprise level, programs can be evaluated long after the activities are over, changes in methodologies can be weighed, and categorization of various programs in different contexts can be developed. And while the data itself may be labor intensive to produce, the data analysis tools at the activity level tend to be rather simple, requiring basic trend tracking and past and future consumption and spend rates. This also applies to measures that DIB teams may help partners integrate into their management systems where a simple tool can require a good deal of labor to develop and support it. For example, creating a unit readiness system (a factor of personnel and equipment, and the training and maintenance of the same) requires a great deal of work by the partner, evaluations of the processes of the system by experts, and a judgment of success by the DIB team and partner-nation leadership.

Quantifiable data can be categorized by the amount of information (too little or too much), the reliability of the information, and timeliness. Having too much information is not often a problem, but discerning what is important and what is not is an issue. Quantitative information should be informed by contextual insight. Experts are best served when they have the key bit of data or insight that informs the data overall: what is missing? How much information is enough? Answers to these questions can be hard to divine. The more common situation is a paucity of information—an undefined force structure, a partial insight into the status of vehicle and equipment maintenance, a lack of records on how the force performs under stress situations (exercises and operations), etc. When developing measures for DIB projects, practitioners will have to take into account the amount of information available and determine if it is adequate to effectively manage, and, if not, what could be done to gain the necessary data to produce relevant outcomes.

One of the greatest obstacles to overcome in the quantitative realm is that usable data often arrives late in a program. Partners may not have the tools to understand their own capabilities, and in fact, often those skills, programs, and processes are an intermediate outcome of the program. If the partner is in crisis, then collection of quantifiable data may be more difficult, yet becomes even more crucial. The solution is clear. If there is no baseline early in the design of the program, other than the realization that there is an unknown, then a line of effort of the program should be to create that baseline.

In order to be confident in the use of data to drive decisions and make evaluations, we must ascribe to that data a level of reliability. As we noted earlier, measures can be thwarted by any manner of political, cultural, and administrative challenges. For instance, in the process of the program, the DIB team may have urged the partner to duplicate measuring

systems that do not capture a useful baseline. If the authority is not there to collect the data, or if there is graft or a need to save face, then the DIB expert faces a dilemma. Fortunately, DIB experts can rely upon others for evidence. The U.S. Embassy country team may have data related to the partner's practices, as well other donor countries. In the realm of resource management, there is usually a wealth of World Bank and other reports to provide a basis of insight. Some data will be more accurate than others, and the DIB professional has to discern how that might reflect upon the institution as a whole: Is the information accurate across the institution or do we know the status of the Special Operations Forces unit because the United States spent years working with them? Is the data simply a snapshot reflecting a moment in time, a recent bit of support to a partner? Will reliable data continue to inform the institution once we have gone?

The challenge of collecting hard, reliable data leads to satisficing, eschewing optimal results for practical outcomes. As Herbert A. Simon posited, "decision makers can satisfice either by finding optimum solutions for a simplified world, or by finding satisfactory solutions for a more realistic world. Neither approach, in general, dominates the other, and both have continued to co-exist in the world of management science."¹⁷ DIB practitioners strive to make a defense institution better not perfect, and the goals they devise with partners reflect an appraisal of what the defense institution can produce throughout the program. In some cases, rough order magnitudes of hard data are what will satisfice to move the institution forward.

Qualitative Data Collection and Assessment

Qualitative approaches can be used throughout the program to test questions that hard data cannot address. Qualitative assessment answers the core questions about the work of the program: "are we doing all the right things to achieve objectives?" which is essentially about progress, and "are we doing things the right way?," which is about the effectiveness of the program. Qualitative practices can be particularly useful for tying hard data to the relevance question: is the program contributing to improving the security capabilities of the partner? Are all the consequences (good, bad, and indifferent) of the program being evaluated? Does the partner have the capacity and political will to follow through on DIB efforts, to make adjustments during the program? Judgments on institutional learning also rely mostly on qualitative evaluations.

One of the most common qualitative approaches is a form of process mapping, which seeks to understand the *how* and *why* of institutional decision-making, often with a goal of increasing organizational efficiency and effectiveness. As applied to DIB, the *who* and *what* are also appropriate questions, providing an appraisal of what an institution is doing well and when it is not. Of course, such assessments are best made with an appreciation of data, but a simple comparison of processes against adapted international practices is itself useful for DIB purposes to measure progress for both U.S. experts and partners' institutions.

Even when quantifiable data is used throughout a program, it is often aided by qualitative measures. For example, a logistics review (a key DIB assessment) starts with a

baseline of data of the current status of the partner's system, followed by an analysis of this data (often tied to operational support).

Setting Intermediary Objectives

During program execution, those who provide oversight will ask if the program is going according to plan. This is can be addressed by those measures of performance (MOPs) that have been built into the program. A MOP weighs whether or not the inputs are producing outputs and outcomes in the project. Are the experts and partner-nation representatives meeting, is work being produced, is the timeline being met, etc.? MOPs are sometimes criticized, because they do not answer the existential question of whether the work will lead to the designed end-state in the institution. But inputs and activities are essential for creating impacts, and knowing their success or failure can help keep a program on track or suggest that a new tack is required. A program will also include measures of effectiveness (MOEs) that address the larger question of how "well" the program is, and if it is leading to desired outcomes and eventually program impacts.

Using both qualitative and quantitative tools, the practitioners weave together measures all along the course of the program. Some of these will serve the purpose of external evaluation for overseers and other stakeholders, while others are internal tools to ensure a program adjusts to the changing arrangements. In this process, intermediate objectives serve as the link between the performance measures and the North Star of a program.

Conclusion

Sigmund Freud posited that psychoanalysis had a limited mandate and, "much will be gained if we succeed in transforming your hysterical misery into common unhappiness."¹⁸ As he saw it, psychoanalysis brought merely "common unhappiness," and DIB practitioners share the same sort of bounded goal, bringing defense institutions into the unsatisfying realm of imperfect public policy and out of the world of unacceptable risk. Within that world are the many challenges unique to working on complex problems with such partners—the political realities, the cultural and social characteristics, and a multitude of actors with a variety of perspectives and motivations. These are intensified as DIB activities increasingly reach a diversity of partners across a spectrum of capability. Despite these dynamic environments, DIB efforts strive to integrate both quantitative and qualitative tools, as appropriate, to measure the progress and effectiveness of programs and activities. In that context, these measures serve to track progress toward intermediate outcomes that compose the causal logic and framework of action toward ultimate outcomes for a partner. When these measures are effectively integrated into DIB practices, a base of experiential evidence sufficient to support accountability and learning is created. This base should be thoughtfully managed to support the decisions, and enable the disciplined agility, to adapt for better outcomes within DIB activities and across the DIB Enterprise. The way ahead is to build on practices for measures and collate knowledge of what is effective and what is not for wider dissemination. As DIB becomes institutionalized within security cooperation, it

should move toward more formal evaluation, since the focus of measures so far for DIB has been on the earlier phases of building partner capacity through monitoring.

Notes

- 1 Sophie Sutcliffe and Julius Court, "Evidence-Based Policymaking: What is it? How does it work? What relevance for developing countries?," Overseas Development Institute, November 2005, available at <<https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3683.pdf>>.
- 2 Matt Andrews, Lant Pritchett, and Michael Woolcock, "Escaping Capability Traps through Problem Driven Iterative Adaptation (PDIA)," Center for International Development at Harvard University, Working Paper No. 240, June 2012, available at <https://www.hks.harvard.edu/content/download/70633/1255342/version/2/file/240_Andrews%2C+Pritchett%2C+Woolcock_BeyondCapabilityTraps_PDIA_FINAL.pdf>.
- 3 Christopher Paul, "Foundations for Assessment: The Hierarchy of Evaluation and the Importance of Articulating a Theory of Change," *Small Wars Journal*, July 30, 2013, available at <<http://smallwarsjournal.com/jrnl/art/foundations-for-assessment-the-hierarchy-of-evaluation-and-the-importance-of-articulating-a>>.
- 4 Leo Tolstoy, *Anna Karenina* (Moscow: The Russian Messenger, 1878), 1.
- 5 Michael McNerney, Jefferson Marquis, Rebecca Zimmerman, and Ariel Klein, *SMART Security Cooperation Objectives: Improving DOD Planning and Guidance* (Santa Monica, CA: RAND Corporation, 2016) available at <http://www.rand.org/pubs/research_reports/RR1430.html>.
- 6 The DIB community of interest includes, but is not limited to: the international community (bi-lateral and multi-lateral, governmental and non-governmental actors engaged in the same "space" as DIB); the interagency; U.S. Office of the Secretary of Defense, the Joint Staff, Services, and agencies; U.S. regional Combatant Commands and their components; country teams and security cooperation practitioners.
- 7 John Kotter, "8 Steps to Accelerate Change in 2015," Kotter International, 2015, available at <<http://www.kotterinternational.com/ebook/eBook-Final-Copyright-2015.pdf>>.
- 8 Graphic used in Defense Governance and Management Team (DGMT) briefings to illustrate the process and considerations used for DIB engagement with partners.
- 9 McNerney et al., *op. cit.*
- 10 Graphic used widely in Defense Governance and Management Team briefings to illustrate the confluence of factors that define where DIB can be most effective.
- 11 Hallie Preskill et al., "Evaluating Complexity: Propositions for Improving Practice," November 2014, available at <<http://www.fsg.org/publications/evaluating-complexity>>, 3.
- 12 Graphic used in Defense Governance and Management Team briefings to illustrate how theories of change and SMART objectives contribute to intermediate and ultimate outcomes.
- 13 Kotter, *op. cit.*
- 14 Vanessa Corlazzoli and Jonathan White, "Back to Basics: A Compilation of Best Practices in Design, Monitoring & Evaluation in Fragile and Conflict-affected Environments," Department for International Development, March 2013, available at <<http://cdacollaborative.org/wordpress/wp-content/uploads/2016/02/Back-to-Basics-a-Compilation-of-Best-Practices.pdf>>.
- 15 Arturo Muñoz, *U.S. Military Information Operations in Afghanistan: Effectiveness of Psychological Operations, 2001–2010* (Santa Monica, CA: RAND Corporation, 2012), available at <www.rand.org/pubs/monographs/MG1060.html>.
- 16 David Chuter and Florence Gaub, "Understanding African Armies," *European Union Institute for Security Studies* Report No. 27, April 2016, available at <<http://www.iss.europa.eu>>.
- 17 Herbert A. Simon, "Rational decision making in business organizations," *The American Economic Review* 69, No. 4 (1979), 493–513.
- 18 Sigmund Freud, *Studies in Hysteria*, translated and edited by James Strachey (New York, NY: Basic Books, Inc., 2000).