Recognizing Systems in Afghanistan

Lessons Learned and New Approaches to Operational Assessments

BY WILLIAM P. UPSHUR, JONATHAN W. ROGINSKI, AND DAVID J. KILCULLEN

Until it was overhauled in 2011, the assessments process in Afghanistan's Regional Command South was mired in 240 metrics and indicators—some of which were uncollectable while others were entirely irrelevant. It lacked focus, failed to define the problem, and was divorced from decisionmaking cycles. That is to say, it was representative of how operational assessments are usually conducted. There was a general understanding that measuring the conflict environment was vital to the mission and to operational success. But what that was supposed to look like and how it was supposed to be accomplished were never articulated. What resulted was a frenetic approach that tried to measure the universe—attempting to analyze everything and accomplishing little.

The years 2009 and 2010 brought a sense that the soon-to-be decade-long war in South Asia needed a new and better defined focus. The campaign in Afghanistan had evolved to a universal, all-encompassing mission, a set of tasks for which the term *mission creep* is euphemistic. These tasks included counterinsurgency with all its associated complexities, counterterrorism, stability operations, developing rural and urban economies, improving governance, countering corruption, improving the rule of law, promoting female empowerment, building government institutions as well as Afghan military and police organizations, and countering the growth and

William P. Upshur is a Senior Associate with Caerus Associates, LLC. Major Jonathan W. Roginski, USA, is an Operations Research and Systems Analyst who served as Deputy Chief of the Afghanistan Assessments Group of 10th Mountain Division during its deployment to Regional Command South. Dr. David J. Kilcullen is Founder and Chief Executive Officer of Caerus Associates, LLC.

movement of narcotics—to name but a few. In Afghanistan, there is nothing we were not doing because everything could be justified as necessary to accomplish what was in reality a vague notion of success. There can be little wonder that operational assessments processes reflected the ambiguity of the mission—it is hard for any metrics system to be more precise than the goals it is designed to measure progress against.

Back in Washington, graduate schools, think tanks, and policy circles had been consumed by the debate about how to apply new focus—whether to shift U.S. presence in Afghanistan to a light footprint and focus on counterterrorism operations, or surge forces forward to replicate what was by then starting to be seen as victory in Iraq. In December 2009, the United States decided to surge its troops by 30,000, bringing the total to 100,000—with many of those troops headed for the south of Afghanistan. For the first time in the nearly decade-long conflict, the President directed that the United States would begin its drawdown in July 2011.

> it is hard for any metrics system to be more precise than the goals it is designed to measure progress against

After the North Atlantic Treaty Organization Lisbon Conference agreed on a coalition withdrawal date of 2014, the headquarters element of 10th Mountain Division deployed to Kandahar Province to take command of international forces in Regional Command South. Kandahar and its environs were, and continue to be, some of the most violent territory in the country.

The war had now acquired a new focus and urgency. The United States had pegged itself to a timeline, even in an environment as violent as southern Afghanistan. Transitioning security responsibility to Afghans became an overarching imperative.

Despite the widespread intellectual understanding of such realities, bureaucracies are ships that do not easily turn course. Organizations (and individuals) at war are fixated on what they know. Like mountain climbers halfway up a difficult rock face, people in war zones respond negatively to new and untried ideas, preferring for safety's sake to stick to what they know. Missions, projects, and endeavors develop staunch political and emotional constituencies. Sunk costs are difficult to rationalize when the ground becomes hallowed by blood already shed.

These dynamics play out on the battlefield as much as in the operational and strategic commands that develop campaign plans and then seek to measure progress in an intensely complex environment. Civilian agencies and nongovernmental organizations are subject to the same conditions.

Theoretical and Practical Problems of Assessments

Within this context, the Assessments Group of 10th Mountain Division, based in Kandahar, engaged throughout its tour in a constant struggle to make sense of the environment, understand changes in it, and communicate judgments about it, clearly and usefully, to the division's command group under Major General James Terry, USA.

The division had leveraged a wide array of expertise in an attempt to synthesize the nuance and complexity of the environment. A stroll around the headquarters (like any operational headquarters in Afghanistan sometimes even down to brigade level) would

find anthropologists, mathematicians, political scientists, area specialists, and a variety of academic experts and analysts gathering, interpreting, and publishing information about the battlespace. This was, of course, in addition to the vast number of military and civilian intelligence personnel deployed throughout the theater at all levels and the vast array of technical collection and analytical means at their disposal.

Within the broader analysis effort, the operational assessments cell was charged with synthesizing, analyzing, and interpreting information and presenting conclusions about the state of operations in a manner that could aid the commander's understanding and decisionmaking. The cell was also called upon to give the commander evidentiary ammunition justifying his decisions to higher headquarters and to policymakers miles away.

There were other teams in the headquarters whose analysis was also relevant to the decision cycle, but operational assessment was the organizational mechanism by which the division monitored the progress of its plans, evaluated their execution, and recommended required changes. As such, the operational assessments cell and its process represented a vital clearinghouse for information that described the counterinsurgency environment. Its task was to establish itself as a set of information receptors attuned to feedback that would allow 10th Mountain Division to understand how its actions affected its environment, and vice versa.

Critiques of the Assessments Process. The difficulty in understanding this environment, and linking that understanding to operational plans, have been justly scrutinized. Indeed, the operational assessments process has been much written about, and much maligned over recent years—with good reason.

One persistent criticism is that operational assessments teams have overreached in the pursuit of perfection. Some have tried to measure the universe, attempting to aggregate all the disparate information in the battlespace. Others, at the other end of the spectrum, have thrown up their hands and accepted the constraints of statistical reporting, merely counting events rather than interpreting them. Another criticism is that assessments often proceed from flawed assumptions with little real-world evidence. The varied cast of agencies performing assessments can at once be criticized for being too complex in their methodology and too simplistic in their analysis. This has resulted in understandable disenchantment with the assessments process.[1]

> one persistent criticism is that operational assessments teams have overreached in the pursuit of perfection

Some analytical products have also been criticized as "coloring book assessments" that hamper understanding of the nuanced counterinsurgency environment. They use a familiar grey-red-orange-yellow-green rating scale to create operational planning maps, color-coding areas from very unstable to very stable. Critics correctly argue that this does not give sufficient information for commanders to make operational decisions and that it is difficult to understand what, if any, analytic processes or data are behind the colors. Others have suggested that some field commanders have developed an unhealthy obsession with changing the color of boxes (colloquially known as "shade-shifting"), rather than looking beyond the five-level color scale to the complexity it

seeks to represent. Yet commanders can only work with the tools their assessment teams offer them. If indeed the role of campaign analysis is to enable better decisionmaking, then it behooves analysts to develop useful and informative tools.

Perhaps in reaction to these criticisms, and perhaps because of the quantitative and technical education of the Operations Research and Systems Analysis branch—the designation for officers who generally lead operational assessments—there has been a great infusion of science, or at least a façade of quantitative rigor, into assessments. Unfortunately, this often culminates in processes built on junk arithmetic and junk logic.

Precision versus Insight. Part of this stems from the experience of the past decade, in which the United States and its allies have conducted an expeditionary counterinsurgency. Expeditionary counterinsurgents, like all expeditionary organizations, bring foreign perspectives to the environments in which they operate. Incidents in both Iraq and Afghanistan contrast the institutional preference for rigor that seems to be inherent in expeditionary force assessments with the qualitative and impressionistic assessments of host nation partners. This is no doubt a matter of perspective. It is also an indicator of understanding—the outsider does not have the emic context and shared memory necessary to make sense of the environment in ways that are meaningful to local partners, while local partners often lack the etic language required to convince an external ally of their perspective.

Expeditionary organizations must also answer to homeland constituencies. The axiom that "some numbers beat no numbers every time" plays out in reports to the U.S. legislative and executive branches. Consumers demand precise analysis in order to justify the ongoing expenditure of blood and treasure and to show that operations are having the intended effects. As one recent study of intelligence in expeditionary counterinsurgency points out:

> commanders on the ground have to justify their actions and judgments to decision makers who may be thousands of miles away and thoroughly out of touch, with little "fingertip feel" for the environment—making quantifiable data a key commodity in the tricky process of handling distant superiors' interventions, and convincing home governments to support on-scene commanders' judgments. Intelligence staffs are therefore pushed to find quantifiable, verifiable, and replicable indicators to support assessments, as ammunition in the discussion with higher headquarters. This is especially so in cultures like that of Western (especially US) intelligence communities, which already place significant weight on numerical data, even if [those] numbers are often used to express largely qualitative judgments.[2]

Some legislators and bureaucrats are likely befuddled by the military's reliance on quantitative reporting, believing that it lacks context. Others, more disposed to loosen funds and resources, may find such information comforting and supportive. Often this results in perfunctorily quantitative analysis and meaningless numeration. When a commander asks for an operational assessments update, he will hear that the "security rating in Kandahar Province is 3.24." To the uninitiated, this might seem impressively scientific until it is unpacked to expose the lack of precision beneath the spurious appearance of rigor.

Drowning in Data. This is not to say that there is insufficient information behind such seemingly precise ratings. In fact, the opposite is often the problem: many assessments cells endeavor to collect and report everything that is important to everyone at all times. As the operational clearinghouses for all of the information that innumerable stakeholders publish or expect to see analyzed, assessments cells are frequently guilty of subjecting commanders to information overload. Analysts are eager to highlight the vast data behind their assessments but are often oblivious to the fact that much of it adds no useful context. It is common to see operational assessments models with hundreds of metrics processed through complex formulae.

In the development of assessments models—those quantitative, *Excel*-powered, esoteric machines that spit out security ratings of 3.24—this overuse of information is referred to as "metric bloat." It is in the pursuit of perfect analytical precision that models become bloated—an attempt to include every possible piece of available data that might have even the tiniest effect on the assessment, without a clear conceptual model that allows analysts to prioritize important factors.

There are negative returns on the investment of adding minutiae to an assessments model. Detail may allow analysts to repose more comfortably on their enormous mounds of information—the sheer quantity of which should preempt any questions as to the veracity or accuracy of their conclusions—but it makes the process slow and unwieldy, diminishes the signal-to-noise ratio for analysts, and far exceeds what decisionmakers find useful. It also leaves gaping holes where data cannot be collected, which are easily hidden behind 3.24.

Even in gathering and analyzing all the data within reach, assessments cells generally put too little energy into information design. Operational assessments are usually presented on a linear scale with a marker to represent progression from left to right, or from "very bad" to "very good." Yet with near universal agreement on the complexity of counterinsurgency, and conflict environments in general, it would be difficult to find anyone who thinks that linear visualizations actually describe changes in the environment in an operationally useful way.

> analysts are eager to highlight the vast data behind their assessments but are often oblivious to the fact that much of it adds no useful context

All this has resulted in operational assessments being sidelined in many commanders' decision cycles. This is hardly surprising. What is described above is spurious decimal grading on a visual scale divorced from meaningful context, emerging from technical, esoteric, even occult quantitative processes understood by very few staff officers. At its worst, it represents an attempt to create an appearance of rigor through the use of quantitative language to express subjective judgment—an attempt that is easily seen through, undermining the credibility of those who engage in it.

Even so, and allowing for all of the criticisms levied against assessments—especially those from observers with relevant operational experience—these critiques disregard some aspects of operational decisionmaking. Assuming they contain rigorous analysis, colored maps (for example, so-called heat maps)

depicting area stability are more useful than overly calibrated, seemingly scientific models. The instincts that lead to metric bloat and information overload cloud the fundamental job of assessments: to give the commander sufficient and sufficiently clear information to decide the effective allocation of resources, priorities, timing, and objectives.

There are also important constraints imposed by the reality of counterinsurgency in Afghanistan, which also probably describe many conflict and postconflict environments. Afghanistan is an information-rich but data-poor environment. Data that do exist are generally of poor quality. Infuriating inconsistency is the norm; impressions, atmospherics,

> field research methods always impose an observer effect, where the act of collecting information changes the population's perception

rumors, and gut feelings abound. Most information comes in the form of anecdote. All this can be useful, and commanders will demand that it be taken into account, but it must be considered in a structured and self-aware way lest it distort decisionmaking.

The social sciences can enable rigorous analysis of qualitative data, but social science methods are hampered by the special circumstances of conflict environments. Field research in a war zone is dangerous to all involved. Attempts to leverage such expertise in Iraq and Afghanistan have had mixed results. Programs such as the U.S. Army's Human Terrain System have tried to deploy anthropological and area studies experts in the battlespace at the tactical and operational levels, but acquiring personnel with the proper

background has proven difficult. Even with the proper personnel, achieving sufficient unfettered interaction with the population and generating meaningful insights from that interaction have proven still more so. Though some commanders find such programs useful, overall results have been far less than hoped.

Field research methods always impose an observer effect, where the act of collecting information changes the population's perception. This dynamic is even more pronounced in the counterinsurgency environment, where researchers are not impartial but rather armed actors in a conflict; thus, there is a "combatant observer effect." The interviewer's obvious association with a combatant organization affects the openness and honesty of respondents, as does the power disparity between a member of an occupying military force and an unarmed local population.

To get beyond these limitations, remotely observable indicators are needed—data that can be collected without changing popular perceptions. In addition, expeditionary organizations have to learn how to make much more effective use of vetted, qualified indigenous researchers.

Thus not all assessments can come from pure, scientific rigor—though the social sciences can usefully inform the process. Analysts must know what is scientifically possible, operationally useful, and timely in the context of the commander's decision cycle.

But with so much analytical noise floating about the headquarters, and so many theoretical and practical problems associated with the assessments process, how would it be possible to parse out what is truly important and stave off metric bloat? If information was inherently unstructured and potentially unknowable, how could the assessments team build a usable and useful model?

Theoretically Grounded Assessment

Answering this question begins with the understanding that environments—especially conflict environments—have personalities. Effective plans must interact with those personalities if they are to have any chance of achieving their objectives.

Determining how 10th Mountain Division's plan was interacting with the environment of southern Afghanistan in 2010 required a general theory—a structured description of how things worked—in order to understand changes in the environment against a meaningful understanding of "normal" background conditions. This theory must be wholly divorced from the strategic theory of victory and from the current coalition operational plan. This general theory of the environment and an understanding of how the environment works —something referred to in other contexts as "territorial logic" or "systems logic"—would become a framework for understanding the logic of the environment and for mapping its dynamic systems, feedback loops, and causal links. By contrast, the campaign plan was a framework for how military operations will achieve specific goals. The former was a map, the latter a flight plan.

Beginning with a period of field observation and qualitative study, looking at a variety of districts and seeking to understand their logic at first hand, the assessments cell eventually posited a general theory. The data at hand suggested that there were dynamic cycles of stability in the provinces for which 10th Mountain Division was responsible.

These "double-loop" stability cycles were driven by the general public's perception of security, degree of government institutionalization (or lack thereof), popular confidence, willingness to invest in noninsurgent institutions, and community resiliency. Community resiliency described the ability of a given community to absorb shocks and the speed with which it returned to a steady state (albeit, perhaps, at a different level of violence than before). A positive catalyst, such as a strong-willed or charismatic political leader or an improvement in security, could lead to greater popular investment in noninsurgent institutions, prompting a virtuous cycle of improved stability and then improved resilience as people's expectations about the future changed.

The Double-loop Stability Model. In common with many field analyses—including, interestingly, the World Bank's *World Development Report 2011: Conflict, Security, and Development*—the assessments cell found evidence for a mutually reinforcing cyclical effect among improvements in confidence on the one hand, and improvements in local security, governance institutionalization, and community resiliency on the other (see figure 1).[3] A negative impetus could of course function in the opposite direction, prompting a vicious cycle of declining security.

This model of the environment, like all models, is a greatly simplified description of a complex and nuanced real-world dynamic system. Furthermore, it is important to note that these dynamics are, at least in theory, specific to a particular place and time. A theory that worked in southern Afghanistan in 2011 cannot necessarily be applied to another theater or elsewhere in Afghanistan. It is highly likely the theory that held in southern Afghanistan in 2011 will no longer hold there in 2015. Any theory needs constant reevaluation as ongoing observation and new information change the general understanding of the theater of operations.

Figure 1. Local Double-loop Stability Model
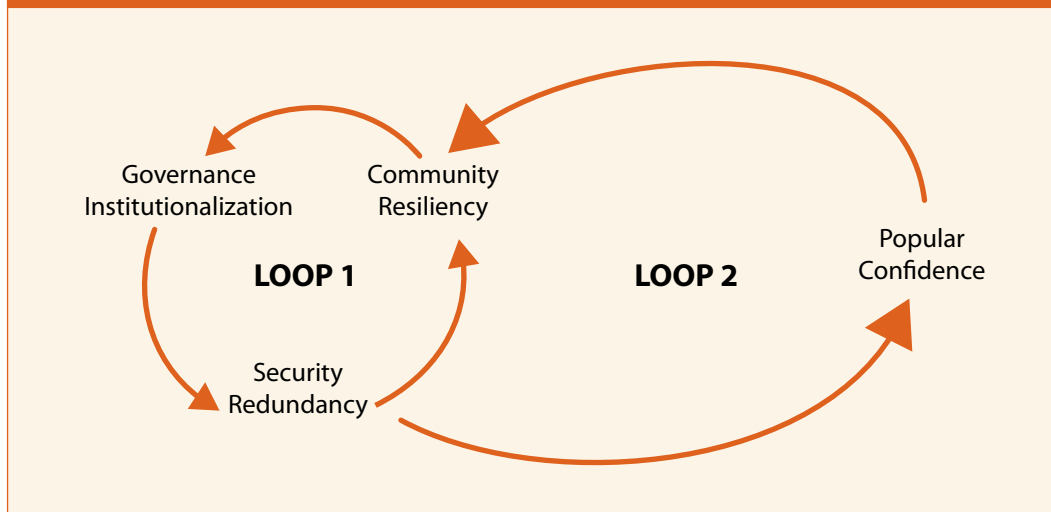
Still, this theory of the environment provided a sufficient basis to guide the development of a set of metrics against which the team was able to assess stability. Using the theory outlined above, operational assessments cells developed a concise list of 11 metrics described by 18 indicators—a far cry from the previous set of 84 metrics, or the 240 metrics used before that. These metrics, described in table 1, represented a dramatic reduction in the analytical burden on the headquarters and reduced the reporting burden for 10th Mountain Division units. Framing the assessment within the context of the real-time operational environment (rather than collecting information against a generic, universal set of indicators as had been previously done) allowed the team to develop a focused assessment process that was at once simpler and more useful.

This approach did not represent an exhaustive list of every piece of information that was important in the 10th Mountain Division battlespace at this time—indeed, it was expressly designed not to. Metrics that are usefully able to describe the environment are a small subset of measurable descriptors. Thus, to be useful, the stability model had to be both a structured and selective description of the environment. It focused on only those particular features that were assessed as important, during that specific time and in that place.

This process of systematic and targeted simplification was, of course, a qualitative one, and this provided the qualitative input needed to imbue subsequent quantitative analysis with meaning. Without the qualitative analysis involved in the triage process of selecting metrics, a purely quantitative analysis would have faced all the problems of rigor without the meaning described earlier. This assessments process thus involved an inductive, qualitative phase in which the team sought to make sense of the environment, and a deductive, quantitative phase in which the indicators (qualitatively designed based on field work) were deductively analyzed.

Metrics	Indicators
Table 1. Assessment Metrics and Indicators in Southern Afghanistan 2011	
Resiliency of government institutions	Length of tenure and quality of government officials
	Degree of government penetration
	Budgetary execution
Survivability of government officials and community leaders	Length of time local officials and community elders survive in office
Public confidence in government legitimacy and effectiveness	Popular perception of government
	Popular use of district center
Predictability and acceptability of corruption	Popular perception of corruption
Strength of Shura and councils	Shura and council assessment
Civilian freedom of movement	Freedom of movement assessment
Function and effectiveness of rule of law	Reliability and accessibility of rule of law
	Popular perception of rule of law
Degradation of the insurgency	Effectiveness of the insurgency
	Capacity of the insurgency
	Will of the insurgency
Function and effectiveness of government security institutions	Popular perception of government security forces and institutions
	Afghan National Security Forces ability to defeat the insurgency
Functioning community security	Popular perception of local security
Resilience of subsistence economy	Economy assessment

This approach also facilitated assessments under the conditions of remote observation often imposed by violent conflict.[4]

Conceivably, identifying the important nodes of a dynamic system and developing pertinent metrics would allow assessments to be conducted on almost any program designed to interact with that environment. In the case of counterinsurgency operations in southern Afghanistan, the team grouped each metric with others in order to describe the division's progress or regress against campaign objectives.

Yet as the forgoing criticisms of operational assessments show, that is not the difficult task. The greater problem is how to incorporate intransigently unscientific information into a rigorous model

to give commanders an accurate assessment and thus enable more informed decisions.

Enabling Better Decisions

To solve this problem, the team applied the widely used method of rating definition levels (RDLs), but with greater granularity than usual. RDLs are the tool with which operational assessments cells—especially in Afghanistan—create their grey-red-orange-yellow-green colored maps. An RDL is essentially a 1–5 Likert scale wherein each level is given a sufficiently specific definition so there can be little disagreement about what level may be assigned to a given area for a given line of operation. Until recently in Afghanistan, this method has been applied only to broadly defined indicators for security, governance, and development.

> the assessments team developed definitions by seeking input from representatives from all around the division

In this case, the 10th Mountain Division assessments team, with representation from across the military and civilian staff, developed an RDL for each of the indicators associated with the 11 metrics in the model. This meant that all the relevant information available in the battlespace could be organized—including information that could not be considered "scientific" on its face. Instead of tacking on a raw narrative, or even worse, forcing anecdotal evidence into specious numeration, information could be categorized along a well-defined 1–5 rating continuum. In this way the pretense of precision was traded for more reliable accuracy.[5]

The RDLs in table 2 were designed to facilitate an analyst's rating of the tenure and

quality of district-level government officials by describing how observations aligned with pre-determined definitions. The assessments team developed these definitions by interviewing and seeking input from representatives from all around the division. This included coalition strategic planning officers, operations officers, and intelligence staff, as well as social scientists, civilian analysts, and representatives from coalition civilian government organizations. For the operational assessment product to have weight and relevance in the command, it was crucial that all stakeholders agree on the definitions associated with each RDL.

It is important to note that these definitions included implicit normative assumptions about what was desirable, based on the theory of the environment outlined earlier. If the theory was flawed in any important way, the assessments would also be flawed. For this reason, as noted earlier, the stability theory and its associated indicators had to be reevaluated with every assessments cycle, refining the theory through the addition of new information and updating it as the environment itself changed.

The team had to walk a fine line in developing these definitions. An RDL must provide sufficient analytical guidance, but the more specific the definitions, the more exclusive each level would become. If it became too difficult to describe conditions on the ground using the definitions developed in the RDL, the definitions would need to be reworked to be less restrictive and more useful. Just as the theory required continuous reassessment and reevaluation, so too the indicators derived from it had to be continually updated.

It was also important to peg each RDL to an aspect of the plan. For instance, in Regional Command South in 2011, the planning process was focused on Afghan leadership, and

	Table 2. Indicator Rating Definition Levels for Government Official Tenure and Quality
5	Most government officials in role longer than 12 months; government provides essential services and protects the population without need for significant international community (IC) aid or security
4	A government official in position more than 6 months is the rule rather than exception; government provides essential services and protects the population but elements of incompetence limit effectiveness; some IC aid needed
3	Government best characterized by officials being appointed but of relatively short tenure (less than 6 months); government hindered from providing essential services and protecting the population due to incompetence; dependent upon external Afghan government or IC aid to perform many functions
2	Government best characterized by frequent, unpredictable turnover; government cannot provide essential services and protect the population due to incompetence; dependent upon significant external Afghan government or IC help to provide essential services to public; local government in danger of collapse absent external aid
1	Government officials frequently replaced; external Afghan government or IC forced to provide for most essential services

objectives were designed to accomplish this. A given area was assessed as ready for transition when its governance and security apparatus was "sufficient" and "sustainable." These terms were the anchors around which each RDL was developed. *Success conditions* were defined by transition-readiness, which in turn was defined by sufficiency and sustainability. This allowed headquarters staff to define what sufficiency and sustainability looked like for each indicator—which in turn helped determine what a sufficient and sustainable overall environment would look like.

Yet even the most precise or accurate results are not useful until they are communicated and understood by decisionmakers. The linear Likert scale is simple and clear but misleading. It fails to show the uneven velocity of progress against objectives. Moreover, in a counterinsurgency context, it applies an arbitrary linearity to the phases of the counterinsurgency continuum defined as "shape," "clear," "hold with expeditionary forces," "hold with indigenous forces," and "build." An example of such a scale is shown in figure 2, with 1 representing an initial assessment and 2 representing a subsequent assessment of progress.

The 10th Mountain Division team discovered by experience that tying the assessment to pre-determined counterinsurgency phases was both uninteresting and analytically unhelpful. There was little or no controversy about which theoretical phase of counterinsurgency the operational headquarters was engaged in at any given time—indeed, the planners had already established this

Figure 2. Example of Linear Counterinsurgency Assessment Scale

S	C	1	2	H/CF	H/ANSF	B

before the assessment process even began. Tactically, clearance operations were taking place in one area while indigenous forces were capable of holding other areas without the need of coalition partnering. The tactical, and by extension, the operational battlespace was a jumble of all stages at once.

Perhaps most egregiously, such linear representation could only present objectives in isolation. This implied that each objective was being pursued in a vacuum, divorced from other aspects of the campaign. Decisionmakers intuitively knew that such one-dimensional simplicity belied reality. Linear depictions of progress, because they failed to show dimensionality, represented a lost opportunity to facilitate discussion among key leaders about the efficacy of their plans and the allocation of scarce resources.

In reality, of course, counterinsurgencies are complex adaptive systems and, as such, are nonlinear by definition. They must be assessed and presented as such. Objectives are interrelated, as are the mostly fungible assets applied against them. The visual representation of the team's campaign assessment had to depict that interrelatedness and at its core be a useful tool with which the commander could make more informed decisions about the allocation of resources in the battlespace.

In communicating assessments, the 10th Mountain Division team found that a simple multiaxis radar diagram addressed these issues more effectively than a linear color scale. A radar diagram could concisely show multidimensionality. It could also clearly display assessments for the past and present, projections for the future, symmetry of progress, and interrelatedness and completion points for objectives, and it could capture nonlinearity in the environment.

The team used radar diagrams to display detailed assessments of holistic provincial or district stability, broad lines of operation such as "security" or "governance," or detailed assessments of specific objectives. In assessing an objective, each axis represented one of the metrics that affected or was related to the objective. The amalgamation of metrics was used to form the axis representing the assessment of an objective as it related to a larger line of operation. Lines of operation could be then applied to provincial- or regional-level assessments, depicting how each objective applied to the wider environment.

Figure 3 conveys an immense amount in information simply and succinctly, shows multidimensionality, and highlights the interrelatedness of metrics. The darker grey polygon in the middle of the diagram represents the previous assessment. The larger grey polygon, outlined in black, represents the current assessment. It is immediately clear that progress has been made in degrading the insurgency and in improving the popular perception of the predictability and acceptability of corruption. It is also immediately apparent that government security institutions are functioning and effective to the degree sufficient to accomplish the objective as marked on the outer perimeter. The dashed line

**Figure 3. Operational Assessment Radar Diagram of Notional Objective**
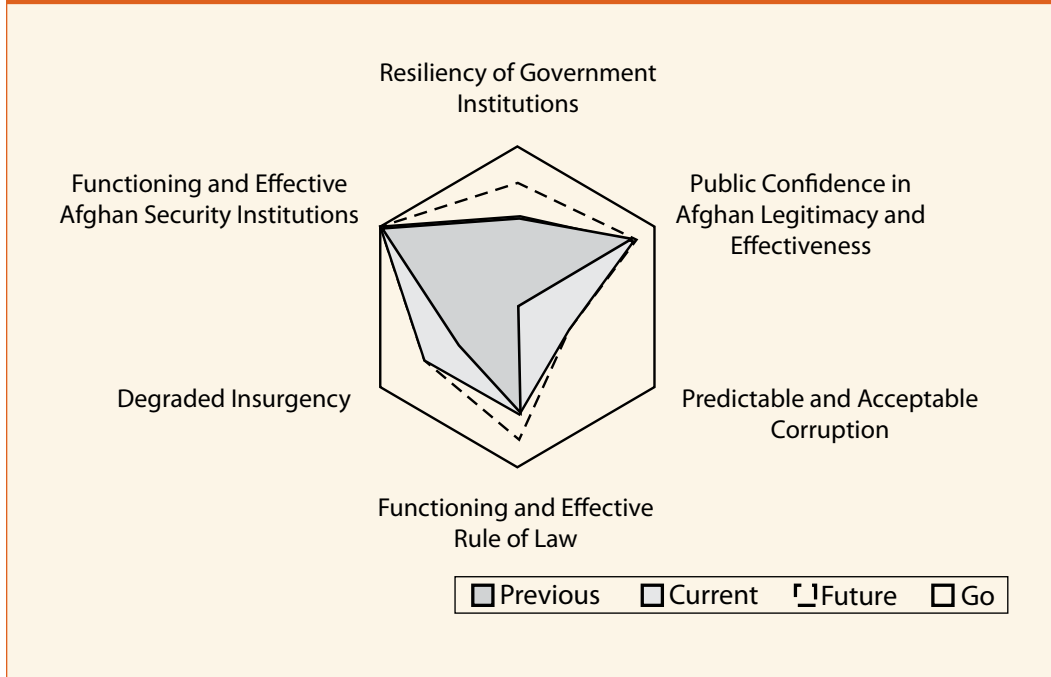
depicts a qualitative projection of the next assessment—progress is projected in the areas of rule of law and resiliency of government institutions. Public confidence in government legitimacy and effectiveness, however, seems to have stalled and is not projected to improve. Most strikingly, there is a significant gap between the current state of the acceptability of corruption and the minimum level required to achieve the desired objective.

As much information as there is on the face of the diagram, much of the aesthetic noise has been stripped away. What is not seen in figure 3 are the axes that mark each metric and the 1–5 RDL scaling. These can be seen in figure 4, which also shows why these diagrams are also referred to as "spider charts."

Stripping away superfluous design helped the team communicate its operational assessment to decisionmakers. It is extremely unhelpful for an information consumer to get hung up on why an assessment is 2 as opposed to 3—something forgotten by organizations that operate on ratings such as 3.24. The important messages to communicate were movements, projections, and gaps against a defined endstate, which spoke directly to the planning process. The scale, therefore, was a distraction.

Note also that what is seen to be the outer perimeter in figure 3 is actually one unit removed from the actual reach of the diagram—the outermost line in figure 4. As a tool to assess the allocation of resources, and changes to resource allocation in a plan, the design had to be able to depict excess or overachievement. In figure 5, for example, functioning and effective community

Figure 4. Radar Diagram "Invisibles"
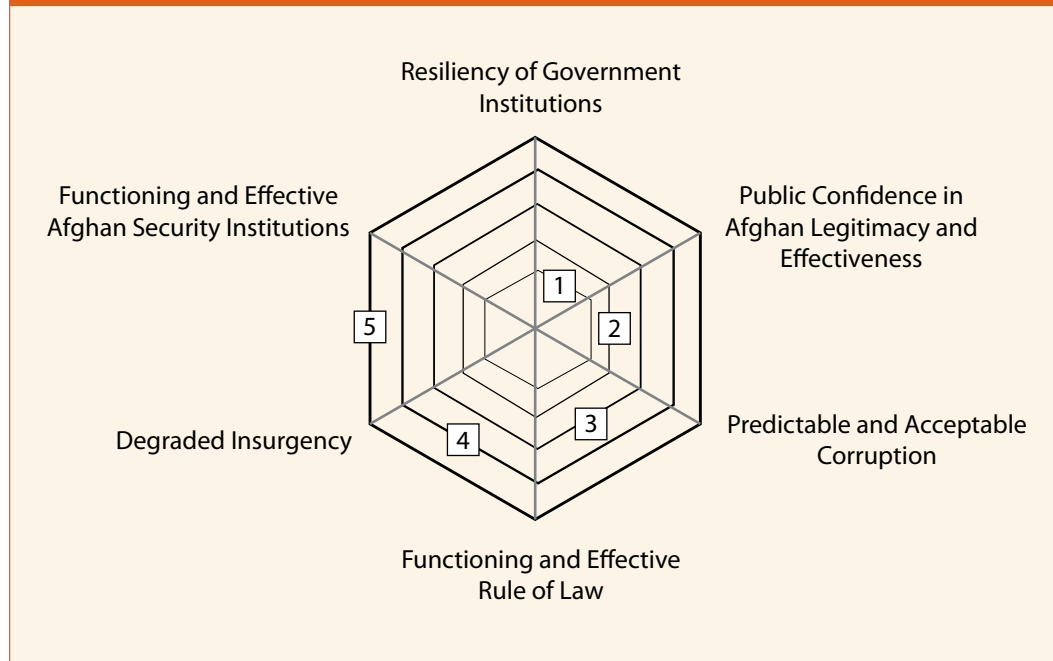
security has progressed past the point deemed necessary to accomplish the objective. In a plan constrained by scarce resources, a decisionmaker would thus have enough data to consider reallocating effort away from this aspect toward degrading the insurgency and improving government security institutions.

Undoubtedly, some would question the utility of information design in the first place. Why not just present information in written form and avoid the risk of being misunderstood? Good graphics might also provide a disincentive for stakeholders to read a deeper narrative and thus leave the wider organization with a shallow understanding of the assessment. The utility of information design has to do partly with an organization's personality. Many military decisionmakers prefer graphic presentation. Good design can allow information to be absorbed quickly and can show trends and projections more succinctly than prose ever could. What visual depictions may lack in detail they make up for in ease of consumption.

In the U.S. military especially, graphic presentation has become a feature of organizational culture—it would be surprising if an operational assessment were accepted in simple prose. Even if it were, it would make for an extraordinarily long and dull briefing to officers working 18-hour days on little sleep and would thus stand little chance of penetrating their thinking. Assessment products must be widely absorbed throughout an organization to be effective, and the ease of absorbing well-displayed information makes graphics a powerful medium.[6]

Figure 5. Radar Diagram Showing Progress Beyond Sufficiency in an Objective
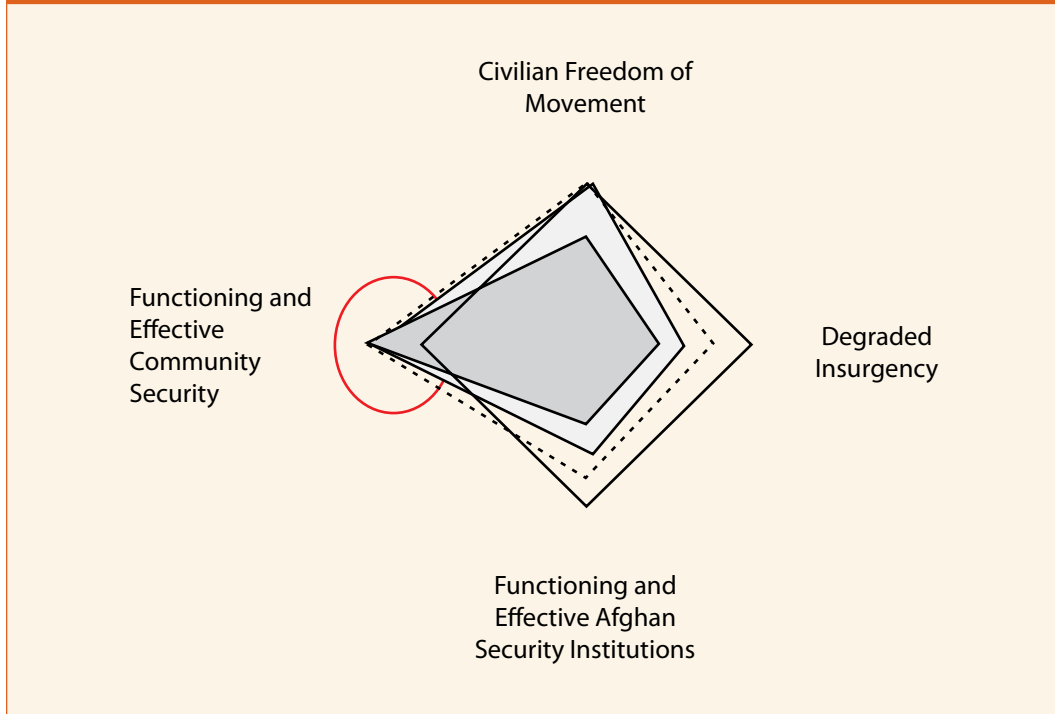
Information design also has to do with an organization's capacity. The myriad constant demands on staff officers and harsh working conditions in the deployed environment affect the quality of writing, content, and information consumption. Many organizations that do most of their work in the field respond better to visual depictions than to lengthy written assessments. The usefulness of graphical depiction of assessments comes from not only their wide applicability but also the ease with which they can be implemented and replicated by organizations with limited excess capacity.

There is also an important epistemological difference in how an organization decides to produce and consume information. A written narrative generally presents information as part of an argument in which the author has consciously or unconsciously staked out a position. It is a rare narrative that presents unbranded information without seeking to give the reader the answer. Wandering too far down this path in a document handed to a commanding general, or any powerful executive, might undermine the document's relevance. Most well-informed consumers do not want to feel steered to a predetermined conclusion.

Well-displayed graphical information, however, is different. It does not smell of predetermined conclusions. Appropriately explained and understood, it empowers thought and discussion. So long as the data are good, a visual depiction need not represent an argument that requires acceptance or dismissal, but can simply act as fuel for ideas.

In fact, it was exactly this discussion among the senior leaders of 10th Mountain Division, its general officers, and their subordinate staff that resulted in significant changes to the campaign in Regional Command South in 2010–2011. This medium also provided a forum for the commanding general of 10th Mountain Division to discuss the status of the campaign during his handover of Regional Command South to 82nd Airborne Division in October 2011.

Still, it can indeed be powerful to present operational assessments in detailed long-form analysis. For historical purposes, especially regarding warfare, well-reasoned prose and well-designed graphical information can be combined to great effect. The design described here would be an excellent complement to a pithy executive summary detailing the nuances of a given environment. Even if a consumer required narrative in lieu of graphic information—and some do—graphical information is a powerful analytic tool with which to construct a written product, even if it is never shown to anyone beyond the lead analyst.

Tentative Conclusions

Overall, the process developed by the 10th Mountain Division assessments team in Kandahar Province in 2010–2011 was simpler and more agile and could reasonably be expected to be more accurate than previously used assessments methods. The most time-consuming aspect was the inductive field research process and the need to acquire sufficiently grounded field experience to develop a cogent theory of the environment and to define pertinent metrics and indicators. Once this was completed, it then took little analytical effort to form the data into a coherent assessment.

The method's simplicity and usability allowed it to inform command decisions at the operational level more frequently than other methods. When it was put into practice in Regional Command South in summer 2011, it reduced the time needed to complete the assessment process from 6 to 2 weeks, and finally cut it to a matter of mere days. Eventually, a comprehensive campaign assessment could be produced virtually on demand.

Of course, assessments processes are still open to criticism. The RAND Corporation's Ben Connable argues that metrics and indicators, which by definition are static even when drawn from a coherent theory of the environment, must fail to account for the nuance and complexity in a conflict or postconflict environment. As has often been said, these environments are made up of highly localized and time-sensitive mosaics. Still, choosing the right metrics and developing a descriptive theory—one verified by observations in the field—was highly informative.

This highlights an undeniable weakness in applying a Likert scale to complex environments. As noted, environments have personalities. What matters in one area will not necessarily matter in another. The RDLs in this model had to be applied to broad regions (four Afghan provinces in the case of Regional Command South), and it was extraordinarily unusual for RDLs to apply perfectly in each area. While it would be possible to develop distinct RDLs for each local area, it would be incredibly taxing to most organizations. Even so, to paraphrase the renowned statistician George Box, this model, like all models, was wrong—but it was more useful than those previously tried.

The process laid out here may represent one element in a broader way forward for operational assessments methodology and assessment information design. It accounts for common

criticisms while simultaneously acknowledging that Afghanistan, like any conflict or postconflict environment, is not an academic problem set.

In the summer preceding this method's development, southern Afghanistan was reeling from its most violent period in recent history. The surge of U.S. forces into Taliban strongholds had resulted in spikes in violence. There was an average of nearly 200 discrete violent events per week in the hotly contested areas of central and western Kandahar. In many cases, Afghan security forces were weaker than insurgents operating in the area and often preyed on the people they were supposed to protect. There was little freedom for citizens to conduct routine business or for government officials to move among their constituencies.

Despite appearances, through its first holistic assessment process, the team recognized that security generally progressed faster than improvements in governance and development. This analysis of the friction surrounding the pace of security, governance, and development efforts led to a discussion of techniques for maintaining the security of an area after major combat operations concluded. Failing to explicitly recognize this fact had previously allowed coalition forces to be carried forward under their own military momentum ("taking the fight to the enemy"), leaving immature governance and development structures in their wake with too little mentorship to grow. With new thinking around the concept of a "sustained hold," gains became more entrenched and solidified. The historically contested Arghandab District of central Kandahar experienced a 90 percent reduction in violent activity between the summer of 2010 and summer of 2011.

In the end, the most valuable output of the assessment process is not a final briefing to the commanding general, a report submitted to a higher headquarters, or a cable sent to the Department of State. It is shared situational understanding among members of the operational staff, between the staff and its commander, and among commanders at different levels that contributes most effectively to leveraging resources against any problem or threat.

> **the team recognized that security generally progressed faster than improvements in governance and development**

In its most mature state, the assessment process becomes larger than any staff section. It becomes ingrained in the way each section, agency, or department operates, with a continual dialogue that includes appraisals of how organizational efforts drive toward common goals. With these methods deployed in their staffs, the leadership of 10th Mountain Division in Regional Command South was able to develop a more sophisticated understanding of progress and the interconnected system in which they designed and executed plans. The combined team benefited from the shared situational awareness derived from its process of assessment, adapting its plan to address the changing landscape of the counterinsurgency environment. **PRISM**

Notes

[1] Ben Connable at the RAND Corporation and Dr. Stephen Downes-Martin at the Naval War College have written some of the most useful and elucidatory work on the subject.

[2] David J. Kilcullen, "Intelligence," in *Understanding Counterinsurgency Warfare*, ed. Thomas Keaney and Thomas Ridd, 145 (London: Taylor and Francis, 2010).

[3] See *World Development Report 2011: Conflict, Security, and Development* (Washington, DC: World Bank Group, 2011), available at <http://wdr2011.worldbank.org/fulltext>.

[4] In the process of developing a radically simplified and structured description of the environment, the team found, with mixed feelings, that its efforts tracked closely with the processes of structured simplification and description discussed by James C. Scott in *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven: Yale University Press, 1999).

[5] It is important to note, as Downes-Martin points out, that a priori accuracy is impossible to determine. Accuracy can only be assessed after the objective has been empirically and verifiably accomplished. Only then can retrospection determine if the model was representative of reality. It is likely, however, that there may never be real proof of accuracy, in which case post hoc analysis can only assess the logical applicability of the process.

[6] The U.S. military is harshly criticized for being overly dependent on PowerPoint from within and without. It has been argued that this has wholly supplanted well-written staff work, but that is another subject. Moreover, it concerns the quality of analysis that underpins graphic presentation of information rather than the visual display of information as such.