

2011

Operations Assessment in Afghanistan Is Broken—What Is to Be Done?

Stephen Downes-Martin

Follow this and additional works at: <https://digital-commons.usnwc.edu/nwc-review>

Recommended Citation

Downes-Martin, Stephen (2011) "Operations Assessment in Afghanistan Is Broken—What Is to Be Done?," *Naval War College Review*: Vol. 64 : No. 4 , Article 9.
Available at: <https://digital-commons.usnwc.edu/nwc-review/vol64/iss4/9>

This Article is brought to you for free and open access by the Journals at U.S. Naval War College Digital Commons. It has been accepted for inclusion in Naval War College Review by an authorized editor of U.S. Naval War College Digital Commons. For more information, please contact repository.inquiries@usnwc.edu.

OPERATIONS ASSESSMENT IN AFGHANISTAN IS BROKEN

What Is to Be Done?

Stephen Downes-Martin

In the absence of a credible numbers-based theory of counterinsurgency there can be no objective, numbers-based assessment for Operation ENDURING FREEDOM. The U.S. military nonetheless has attempted to conduct a numbers-based assessment process. Thus, when a new commander and staff take over duties as a regional command in Afghanistan, they inherit an operations assessment process riddled with highly visible flaws that emanate from the improper use of numbers and flawed logic. While no assessment process can be perfect or free of any criticism, the flaws the author observed during a six-week stint in-country are sufficiently egregious that they seriously reduce the value those

Dr. Stephen Downes-Martin, research professor at the Naval War College, has over thirty years of experience in developing and applying war gaming, game theory, decision analysis, and systems thinking to tactical, operational, and strategic decision support. He has a PhD in physics from King's College, University of London, was a military intelligence officer in the British Army, and is now a U.S. citizen. The commander of I Marine Expeditionary Force (Forward) invited him to be his assessments adviser for the spring of 2010, during which time he worked with, observed, and interviewed officers, senior noncommissioned officers, and civilians at a variety of assessments-related staffs at regional commands, provincial reconstruction teams, the International Security Assistance Force Headquarters, and the International Security Assistance Force Joint Command Headquarters. This article is a product of that work and of discussions with officers and analysts during the year since his visit.

assessments provide to commanders' decision support. In addition, the visibility of these flaws means that military assessments, and by association the military commanders, are rightfully distrusted by higher civilian authority and by other organizations within the theater. It is therefore imperative that incoming commanders and staffs taking over responsibilities for regional commands address these flaws to improve decision making and to earn the trust of higher civilian authority and organizations with whom they have to work.

Staffs and commanders in Afghanistan created operations assessment processes under extraordinarily difficult circumstances while fighting, and it is extraordinary how well they have done given those

Naval War College Review, Autumn 2011, Vol. 64, No. 4

circumstances. Nevertheless, it is necessary to identify and fix the flaws in the present operations assessment process to strengthen decision support and the validity of the assessments, without apportioning blame or criticism. In this article I avoid identifying individuals or organizations as much as possible in discussing flaws in the assessments processes that I have observed. Instead, I address the proliferation of “junk arithmetic” and flawed logic within the currently used assessment processes and discuss why regional commanders and their staffs should care about these problems, by describing the damage to commanders’ credibility and decision support created by flawed processes. Finally, I propose an approach to operations assessment that regional commanders can immediately put into place. I do not discuss or comment on strategy, operations, or the broader arguments concerning counterinsurgency versus counterterrorism. I focus solely on the operations assessment process.

Dr. Jonathan Schroden convincingly argues (in an accompanying article) concerning operations assessment that “there is an entire failure cycle at work, and until its associated issues are rectified, the theoretical promises of operations assessment will continue to go unrealized.” Regional commanders do not have to wait for higher command to deal with these key issues, nor should they; they have the authority and capability to deal with them within their own regional commands. The assessment approach I propose in this article meshes with the requests for information from higher command and fits within the operational planning process of the regional command. It uses military and civilian professional judgment applied to the appropriate combination of objective and subjective data, backed up with valid arithmetic and sound logic.

WHAT IS OPERATIONS ASSESSMENT?

Joint doctrine describes assessment as “a process that measures progress of the joint force toward mission accomplishment.”¹ Joint doctrine also makes clear that simply measuring progress is insufficient, that the assessment process must “help commanders adjust operations and resources as required, determine when to execute branches and sequels, and make other critical decisions to ensure current and future operations remain aligned with the mission and military end state.”²

Implications of the Doctrinal Definition of Operations Assessment

How exactly does one measure progress toward accomplishing a mission? Unless one has already accomplished the mission, mission accomplishment or failure will occur in the future, but one does assessment in the *present*, using present and past information. Therefore, by definition, operations assessment is an attempt to forecast future success based on current and past experience.³ It attempts

to provide insight into how well the command is achieving its operational objectives (occurring in the future) using information from the environment (describing the present and the past). Assessment should include a measure of the gap between the current situation and the future desired end state, the rate of closure (or widening) of that gap and a forecast of its future rate of closure (or widening), and an assessment of the risk to the endurance of the end state after the achievement of objectives and termination of military action.

Therefore, in order to provide decision support to the commander within the guidelines laid down by joint doctrine, operations assessment must answer what I call “the assessment question,” which in general has the form: “What is the likelihood of, and what are the risks to, the conditions for the specified end states occurring or remaining stable if military operations are terminated on the specified date?”

In the context of Afghanistan, I propose that the assessment process at the regional command level must answer the more specific question: “What is the likelihood of, and what are the risks to, the conditions for the specified end states occurring or remaining stable if the region transitions from coalition forces to full Government of the Islamic Republic of Afghanistan (GIROA) control on the specified date?”⁴

If the operations assessment cell answers this question, the commander and staff will be in a better position to report progress toward the objective, identify risks to achieving the objective, and propose reallocation of resources to minimize or mitigate the risks.⁵ Assessing any individual line of operation or objective is a matter of answering the assessment question in the context of the likelihood that the end states specified for the objectives remain stable after transition and what the risks are to those end states should transition occur.⁶ In my opinion, lines of operation or objectives in Afghanistan are only of interest to the extent that they enable a transition of power to GIROA by the specified date with the specified end states remaining stable.

Forecasting has a long and dubious history, full of pseudoscience, junk arithmetic, and flawed logic, practiced by witches and listened to by kings.⁷ Forecasting should be done using a combination of subjective professional judgment, objective logic, (social) science, and mathematics.⁸ However, for many people the differences between pseudoscience and real science are hard to spot. Furthermore, approaches that are valid in one context can become invalid in others, even if the differences in the contexts are not obvious. So although most officers would subscribe to the notion of using valid logic, mathematics, and science (everyone believes that they themselves are rational and logical), it is difficult for those not explicitly educated and trained in science, analysis, and critical

thinking to identify whether an approach is logically or scientifically valid. This difficulty is the root cause of many of the flaws I have observed in operations assessment as practiced in Afghanistan.⁹

Requirements for Operations Assessment

I claim that four primary requirements must be met if an operations assessment process is to provide good decision support to a regional commander.

Assessments at the different levels of warfare (tactical, operational, strategic) and across the instruments of national power must be linked. Linkage “up” the levels of warfare means that the assessments for all districts in a region should play major parts in the regional assessment and that the assessments for all regions should play major parts in the national assessment. There will be emergent effects at each level that are not described by combining the assessments from lower levels, but any assessment should provide logical reasons for them, based on the inputs. Linkage “across” the instruments of national power provides an integrated assessment combining the separate diplomatic, information, military, and economics assessments from civilian and military branches of the involved coalition governments.

Metrics and operational end states must be logically connected. There must be a logical connection between the processed metrics data and a forecast of the endurance of the end states if transition were to occur—that is, credible and logical reasons *why* the (qualitative and quantitative) values of the metrics forecast the stability (or instability) of the end states, should transition occur. Satisfying this requirement provides a mechanism for addressing the risks to the end states, which is critical to supporting the commander’s decisions about the allocations of resources needed to deal with those risks. In addition, this requirement helps determine what metrics are required.

The appropriate metrics must be identified and the data to process them collected. There are three possible failure modes here: collecting irrelevant metrics, not collecting necessary metrics, and not knowing which of the previous two failures are present.

The metrics data must be processed using valid logic, arithmetic, and science. In the absence of an objective, numeric theory of counterinsurgency, it is especially necessary that assessment staffs creatively apply subjective, professional judgment to the objective and subjective metric data in order to answer the assessment question.¹⁰ It is critical that such creativity not violate established rules of logic, mathematics, or science, lest staffs generate unidentified errors when making an assessment and damage any decision that uses the assessment.

I argue that the military assessments I have observed in Afghanistan clearly do not satisfy these four requirements.

FLAWS IN ASSESSMENT AS CURRENTLY PRACTICED

I have identified by direct observation six major flaws to credible and quality operations assessments. These render assessments unfit for providing decision support to the commander, thereby casting doubt on the credibility of the assessments and upon the credibility of the commands and commanders that use them.

Overoptimism

An officer must be prepared and able to wear two very different hats. One of these is the “planner and analyst” hat worn when, among other occasions, doing assessments. The other is the “leader” hat, worn when leading subordinates in the execution of plans. The first requires a pessimist mind-set (“the glass is half-empty”) and focuses on critical thinking and the application of logic to identifying and overcoming what can go wrong, in order to identify and mitigate risk and, in turn, to ensure mission success. The second requires an optimist mind-set (“the glass is half-full”) and a focus on inspiring subordinates, politicians, and civilians to achieve objectives despite the risks.¹¹ The two mind-sets (pessimism and optimism) are very different, but the former is critical to assessment, for the following reasons.

Operations assessment requires analysis using logic and elements of the scientific method. A critical component of the scientific method is the concept of “falsifiability”—that a hypothesis must be capable of being disproved in order to be worthwhile. One does not prove some hypothesis to be true; instead the best one can do is fail to disprove it.¹² Similarly, in applying evidential reasoning to distinguish between alternative explanations using “analysis of competing hypotheses,” one does not compare the strengths of supporting evidence for various alternatives; instead one compares the weaknesses of evidence against them.¹³ Underlying these concepts is the fact that knowledge is contingent.¹⁴ These notions are counterintuitive, but they are well established, and they underlie the last three hundred years of successful Western science and the last 2,400 years of Western philosophical and logical thought.¹⁵

Officers who wear their leader hats when analyzing or assessing risk producing poor analyses or assessments, and officers who wear their analyst hats while leading place execution at risk. In addition an officer must be very careful if deliberately deciding to wear a “glass half-full” hat when reporting an assessment, whether up the chain of command (including to higher civilian authority) or to external organizations, such as the media. The risk here is inappropriate optimism.

In addition to the necessary critical attitude for assessment, an officer who is required to bring creative subjective assessment to bear must be capable of applying inductive logic, which requires a divergent mind-set capable of

recognizing patterns from partial information. The danger here is that humanity has evolved mostly to err on the side of false positives when looking for patterns (“better safe than sorry”), which in the case of a naturally positive attitude will lead to claiming optimistic patterns in the data.

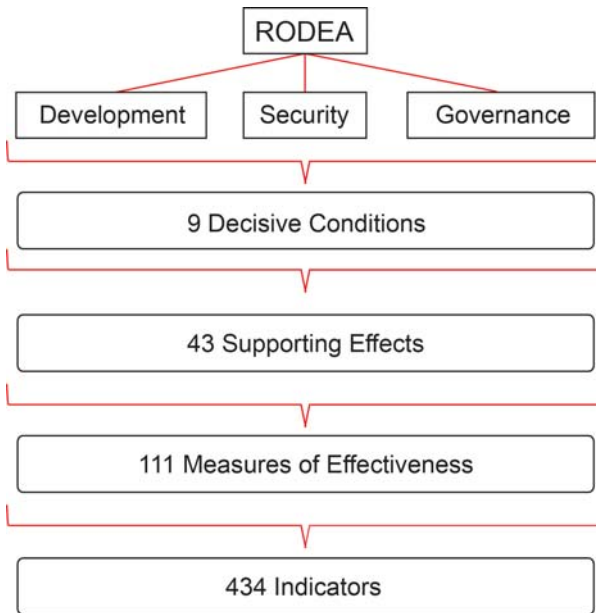
Although military commanders and their staffs work hard at avoiding over-optimism, they tend to bring their leadership (“we can do it”/“glass half-full”) characteristics to bear during analysis and assessment, and there is an institutional drive to produce “good-news stories.” This latter drive is partially in response to the “bad-news stories” reported in the press, and partially in response to the imperative to show progress in time to serve the ends of various political timetables. It is extremely difficult, to the point of impossibility, for an individual to achieve the correct positive and negative balance, but an organization can, and it best achieves this balance by deliberately setting up an adversarial process, using devil’s advocacy. The goal of this process is to identify and examine all the ways in which things can go wrong, in order to institutionalize the critical “glass half-empty” attitude and ensure that the natural desire for good news to pass on up the chain of command does not dominate assessment or reporting.

Metrics Collection

I have observed two major types of metrics collection problems. The first problem, promiscuous metrics collection, breaks down into two parts, self-inflicted and inflicted from above. Some assessments cells and teams told me they collect as much information and as many metrics as they can think of, “just in case.” In these cases, the stated goal was to be able to “change what we analyze as objectives or requests for information change without having to change what we are collecting.” This results in a high likelihood that many of the collected metrics are not relevant to the situation being assessed.

In addition, the International Security Assistance Force (ISAF), the ISAF Joint Command (IJC), and some regional commands have issued “fragmentary orders” (FRAGOs, or FRAGORDs), requiring the collection of large numbers of numeric metrics (see, for example, the Regional Operational Design Effects Assessment [RODEA] structure used for a while by Regional Command [South], described in figure 1).¹⁶ It is beyond the capacity of most forces in the field to collect on the large numbers of demanded metrics and produce a credible product. Frequent attempts by subject-matter experts to reduce the number of metrics by brainstorming are unlikely to help, for two reasons. First, there is no guarantee that the original lists of metrics contained all those required, so any reduced list may miss necessary ones; second, it was mostly brainstorming in the absence of an analytic framework that gave rise to the problem of metrics

FIGURE 1
REGIONAL OPERATIONAL DESIGN EFFECTS ASSESSMENT (RODEA) STRUCTURE



“bloat” in the first place. Trying to correct a problem by doing the same thing that created it is unlikely to succeed.

The second problem, blinkered metrics collection, attempts to avoid the first problem but introduces worse consequences. Some assessments cells told me that they try to identify up front which metrics are hard or impossible to collect, and then set these aside. The problem is that there is no analysis as to whether these ignored metrics should be collected, and therefore there is a risk that critical metrics will be neither collected nor considered. What is worse, commanders are not informed that the assessment is ignoring

metrics whose importance has not been determined. This jeopardizes the accuracy of the assessment and hides from the commander and from later assessors the original decision to ignore hard- or impossible-to-collect metrics. Since the assessment may very well be poor, the credibility of the commander is ultimately placed at risk.

Junk Arithmetic

Using arithmetic on numeric metrics is optional, but the rules of arithmetic are *not* optional.¹⁷ The following examples of junk arithmetic I encountered suffice to demonstrate the broader problem.

Many of the assessments processes I observed in-theater take qualitative and quantitative data, rank order them, and average the rank-order numbers. For example, in the RODEA process, assessors coded answers to questions on a point scale of one through five, similar to the “rating definition levels” used by ISAF and IJC. These codes are not ratio-scale numbers, and therefore, by the laws of arithmetic, functions such as “averaging” cannot be performed on them—it would be meaningless.¹⁸ To put this into a familiar context, officer pay grades are rank ordered by “O number”—that is, pay grades O-1 (second lieutenant) through O-10 (four-star general). But no one believes that a brigadier general (O-7) is the same as a major (O-4) paired with a captain (O-3) just because four plus three is

seven.¹⁹ Averaging ordinal numbers, such as rank orders, within an assessment process is just as nonsensical, and this kind of obvious error subjects the credibility of the assessment, and the command promoting it, to justifiable suspicion.

Values for many metrics are obtained using polls. In Afghanistan, these polls have claimed margins of error of approximately plus or minus 3 percent for nationwide surveys, 5 percent for regions, and 10 percent for districts.²⁰ Given a plus-or-minus 10 percent margin of error, a district metric would have to change by approximately 20 percent before one could claim a trend. When a change is less than approximately double the margin of error, the soundest conclusion that can be drawn is “We do not know whether there has been a change or not.” Unfortunately in most assessments I observed in-theater these margins of error were ignored, and in a significant number of instances officers claimed unjustified trends on small changes of data. An “assessments dashboard” I observed did not even have a symbol for “trend unknown,” just check-boxes for “trend improving,” “static,” and “declining”; it was impossible to report that it was not known whether there even *was* a trend. When I asked about this, the officer in charge replied, “The military does not like to admit we do not know, so we report one of the dashboard options and then caveat the report.” To report “We know that this measure is not changing” is obviously not the same as to report “We do not know if this measure is changing”; the two situations have very different implications for the commander’s decision making. It is doubtful whether anyone remembers caveats after a trend report has been delivered.

Another observed example of junk arithmetic (this time leading to an overly optimistic claim) was an Afghan National Police assessment claiming that the organization was “nearly 100 percent filled.” Examination of the underlying data showed that patrolmen were overmanned while officers and noncommissioned officers were undermanned, by significant amounts. Whether or not the assessment reported the underlying data along with the conclusion, the fact remains that “nearly 100 percent filled” simply did not faithfully summarize the situation in this case. The “nearly 100 percent filled” summary had been derived by applying junk arithmetic to the underlying data, and all too often the underlying data that would reveal the true situation do not make it into the reports.

The use of *any* arithmetic on numeric metrics in counterinsurgency is subject to suspicion, the more so when one attempts to roll up the numbers into some grand score of how well we are doing. Even in a country as stable as the United States, with all the economic data and information one could conceivably ask for and no one trying to kill you as you ask for it, we still do not have credible economic forecast models that can avoid near-catastrophic economic meltdowns. Where then is the credible numbers-based model for governance, development, and security in Afghanistan during an insurgency? Certainly, a professional,

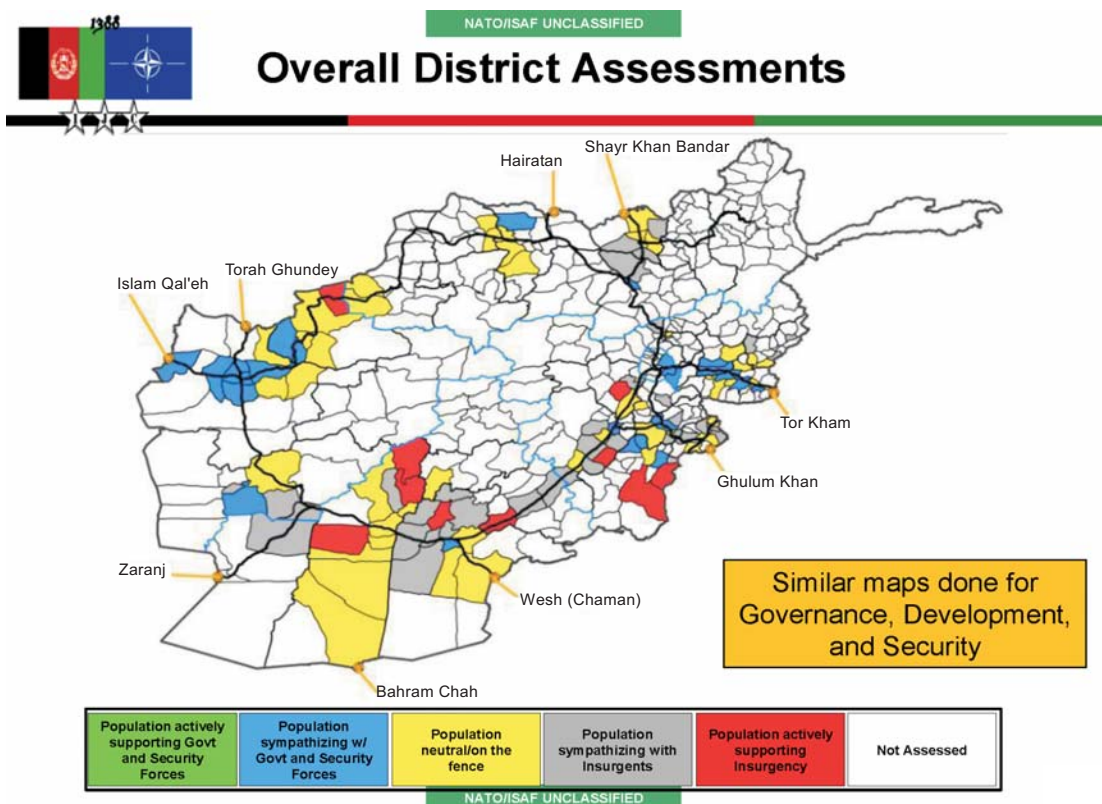
subjective, qualitative assessment of progress will make use of certain numbers, but not by running arithmetic functions over them.

Simplistic Color Coding

Senior commanders' time is a precious and nonrenewable resource; staffs rightfully guard it jealously. The most common approach in-theater to providing senior commanders with the conclusions of assessments is to produce a color-coded map, each district shown in one of five colors indicating the level of success there (see figure 2).²¹ But as has been pointed out and persuasively argued to a working group of senior officers, including generals, at IJC, "The color-coded map dilutes transparency and accuracy and offers a simplistic and misleading representation of the battle space."²²

I have observed that most senior commanders demand narrative explanations from subordinates during briefings and forward narrative assessments up the chain. However, staffs do not usually collect, document, or store these narratives in a database form suitable for later analysis. Since these senior commanders

FIGURE 2
EXAMPLE OVERALL DISTRICT ASSESSMENTS, COLOR CODED BY LEVEL OF SUCCESS



ISAF Joint Command Metrics Workshop, Washington, D.C., 17–18 March 2010, unclassified handout.

apparently believe it is necessary to spend time on narratives, it is necessary to find a way to present them with pertinent narratives in the first place within their time constraints. Claiming that color-coded maps serve the purpose is disingenuous. Any changes in color to a district or province will immediately require a narrative explanation; any nonchange in color will also require an explanation of why, despite effort, the situation has not improved. In either case, it is important in the narrative to answer the questions the briefer knows the commander is going to ask.

In addition, a color-coded map hides information. The single color coding represents an average (not a summary) of a large number of underlying factors, most of which it is nonsensical to average. An average can quite possibly stay the same as some factors improve and others degrade; the color tells us nothing useful about this situation, so one must add narrative explanations. Since smart staffs often provide such narratives anyway, the color-coded map becomes pointless at best and a misleading time waster at worst.

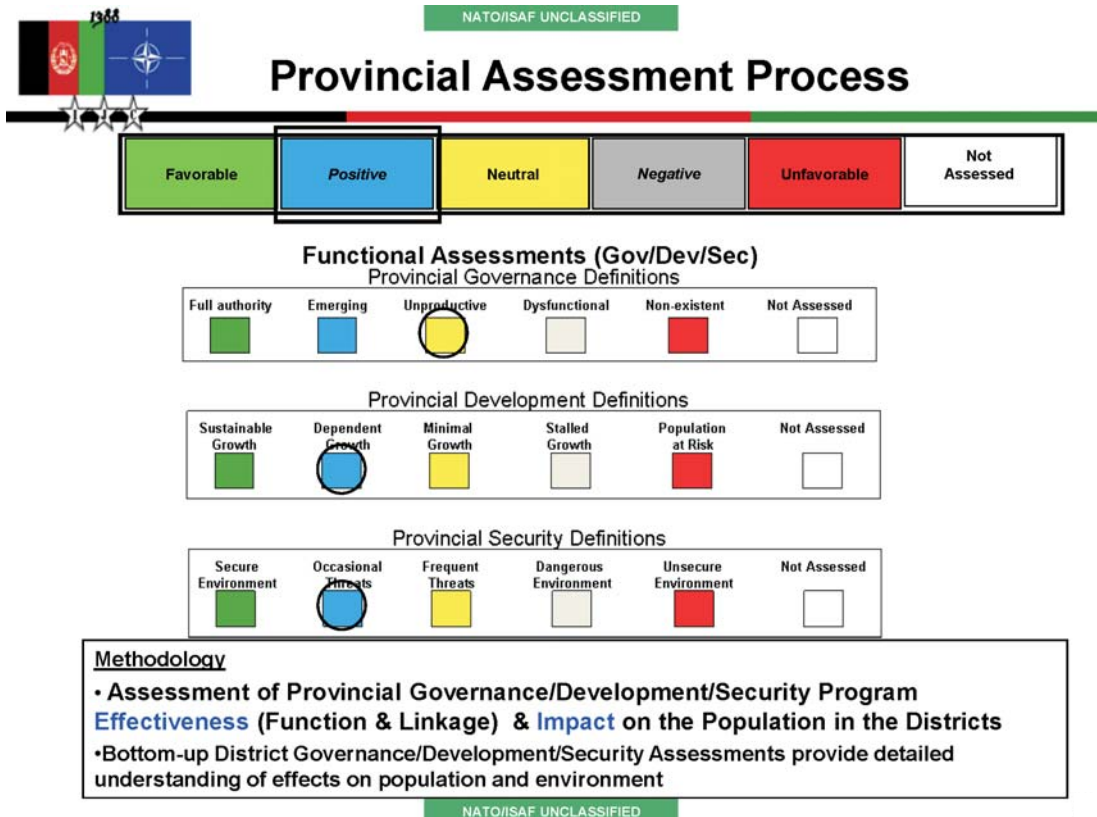
Logic Failures

Even if valid success/fail statistics exist at the district level, validated and documented methods of “rolling them up” into a forecast of success or failure at the regional or national level do not exist. Nor are there any past data on which to base methods for such roll-ups.

No Compelling Combination of Assessments. Furthermore, validated and documented models of how to combine assessments across the instruments of national power (diplomatic/political, information, military, and economic) do not exist. There is no credible model for how to combine the assessments of objectives within the lines of operation (governance, development, and security) into a final assessment. Regional commands appear to be “color averaging” when attempting to combine assessments from separate lines of operation. The regional commands present separate colors for their respective regions for security, governance, and development, then provide an overall assessment color that happens to be the average point on the color-bar chart of the three lines of operation (see figure 3).²³ This is not coincidence; I have observed regional command briefers struggle to explain in operational terms why they had given a particular color to an overall assessment.

The combined coding scale for assessing a district or region (see figures 2 and 4) is drawn directly from the population component of what is known as a “systems thinking” model of counterinsurgency in Afghanistan.²⁴ In that model, each box contains the part of the population described by that box. The model spreads the population across all five boxes. It is incorrect, however, to interpret each box as a position on a scale that describes *the whole population*; the color

FIGURE 3
EXAMPLE COLOR-CODED PROVINCIAL ASSESSMENT



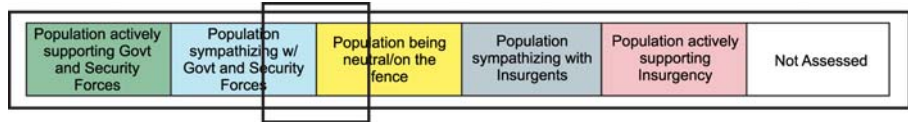
This example combines the colors of the assessments of the three lines of operation.

ISAF Joint Command Metrics Workshop.

scale used by coalition forces, as shown in figures 2 and 4, is an incorrect oversimplification of the original model and cannot describe the majority of popular-support situations.

For example, this representation cannot handle a polarized society where, say, 50 percent actively support the government and security forces and 50 percent the insurgency. One solution proposed by a staff in-theater was to assess a polarized example as being neutral, “on the fence.” But clearly a fractured polarized society is not the same as a neutral society, and this staff was also unable to explain how it would represent and assess a population equally distributed along the support dimension (i.e., with 20 percent of the population in each of the five boxes). This simplistic color-bar approach also cannot handle other very likely distributions of support (with different percentages of the population distributed across all support-level boxes). I observed some regional commands spreading their assessments across two contiguous boxes (as in figure 4); unfortunately, this extension does not solve the problem.

FIGURE 4
AN OVERLAPPED ASSESSMENT EXAMPLE



Combined assessment of population support places the entire population into one of five levels of support for GIRoA and security forces or the insurgency, or into an overlap between two contiguous levels.

No Compelling Connection among Objectives, Metrics, and Assessment. All military forces are employed for reasons—in order to achieve end states laid out in the commander’s objectives—and progress is measured against those end states. Therefore, the commander’s objectives must determine the metrics used.

I have heard no compelling arguments as to why staffs collected certain metrics and not others. For example, a staff at one command claimed it had carried out a rigorous process but was unable to describe or document what that process was, who executed it, and why their predecessors had selected those specific metrics. Staff members at another command claimed to have decided on its metrics “from a book or paper, added some from Iraq, and then added a few of our own.” Another command has had metrics pushed on it multiple times by various academics, think tanks, and civilian agencies and has been obliged to take these seriously, since their authors have political influence.

Since lines of operation, taken alone and out of context, are broad, any conceivable metric is plausible. For example, during an IJC Metrics Evaluation Meeting held in Washington, D.C., on 17–18 March 2010, one participant claimed that “child mortality” was an appropriate metric under “Development.” Asked to explain how this metric supported counterinsurgency, the participant replied simply, “Afghan families care about their children.” Unless on such occasions military commanders want to argue in public that they do not care about the lives of Afghan infants (or whatever other such topics are under discussion) or are willing to collect on and measure any metric proposed by whoever happens to be in the room, they had better have clear, cogent reasons for how the metrics they are using are tied to their objectives, and therefore why other metrics are *not* being collected by their commands.²⁵ Unfortunately, most commands do not appear to have clear connections between their objectives and the metrics they are collecting;²⁶ and at this conference no sound answer was forthcoming from the officers present as to how infant mortality was *or was not* tied to their counterinsurgency objectives.

Staffs that are unable to justify the metrics they are collecting cannot justify *not* collecting others foisted on them by an entire cottage industry of academics, think tanks, newspaper columnists, retired officers, politicians, and

congressional staffers, or by history books about Iraq, or even brainstorm “by whoever is in the room at the time.” I observed a staff at higher command attempting to reconcile at least four collections of metrics to see which ones were duplicates and which were new, instead of determining what metrics they should be collecting, based on the objectives.

No Compelling Connection between Assessment and End State. For operations in Afghanistan, the end state is, loosely speaking, a region (or district, or the entire country) that is suitable for transition to full GIRoA control, where “suitable” means there is some, good chance that the GIRoA will be able to keep it stable and secure. However, unless we have a credible theory that links the level of active support for the insurgency to the likelihood of GIRoA successfully running an area, we have no connection between the rolled-up color-coded assessment and the desired end state. Therefore, the assessment does not provide senior leaders tasked with judging the suitability of a region or district for transition with a credible assessment of its suitability. What it does provide to those decision makers is information from which to argue either way, depending on political convenience.

Higher-Command Demands for Objective Assessments

ISAF and IJC (supported by higher civilian authority) demand a “set of indicators that complements the commander’s qualitative assessment of the environment.”²⁷ Unfortunately, in practice, “indicators” are all too often interpreted as being “quantitative” (or “numeric”) and thus “objective,” whereas the “commander’s qualitative assessment” is seen as “subjective.” An example is a report from one provincial reconstruction team that referred to “overreliance on qualitative and subjective assessments” as a challenge.²⁸

An objective numeric assessment for Afghanistan requires a credible numbers-based theory of counterinsurgency that is applicable to Afghanistan—a way of computing from the metrics the probability that if we transitioned an area to GIRoA the desired end states would endure.²⁹ For such a theory to be credible, one must be able to apply it to past insurgencies with known metrics and with known outcomes; at the very least, one would want such data to construct a model. However, although there have been hundreds of counterinsurgency wars in the past century, comprehensive data on the operational environments and how they changed over time have been kept for only a small number of them;³⁰ we do not have enough data for credible success/failure statistics. Some would argue that the data we have from the approximately four hundred districts in Afghanistan could provide a numbers-based statistical analysis; the problem is we do not have success/fail outcomes for any of these—the war is not over yet.³¹ Therefore, we have no outcome data for Afghanistan. Although statistical

models are applicable at the tactical level (we have large amounts of data concerning tactical/small-unit engagements), they are not applicable to assessing success or failure at the operational or strategic level (since we do not have statistically valid sample sizes).

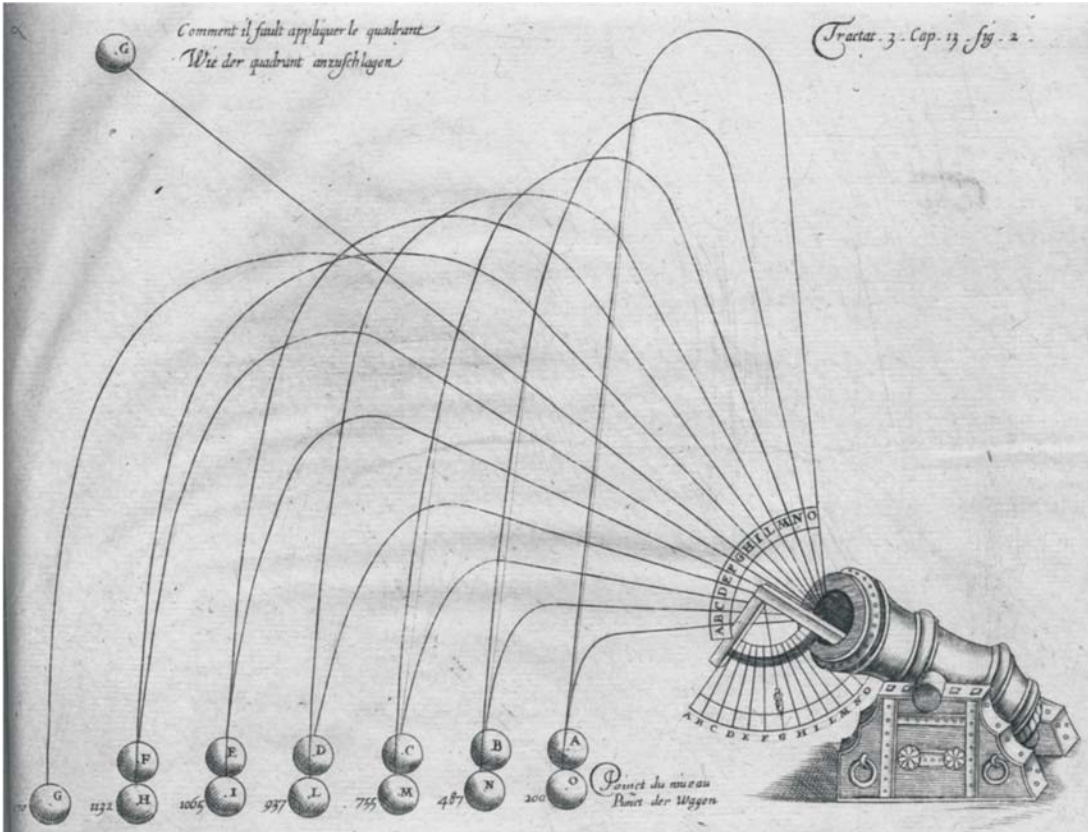
Consider the assessment problem faced by Robert Norton, an artilleryman of the seventeenth century (see figure 5). Despite Norton's skill at mathematics and his recognition that mathematics was important to the artilleryman, he did not have an objective numbers-based theory of external ballistics—Newton had yet to develop his theory of gravity and equations of motion. Norton could collect all the numeric data he liked concerning the present (for example, muzzle velocity, weight of ball, amount of powder, and angle of elevation), but without a numbers-based theory of ballistics he had to apply subjective professional judgment to those objective and subjective numbers to assess where a shot would fall in the future. In fact, the subjective judgment of artillerymen in pre-Newtonian days was superior to the many “objective” but wrong numbers-based assessments published during those times.³²

In the absence of a credible numbers-based theory of counterinsurgency in Afghanistan, there is no objective, numbers-based assessment for military operations there. *Pretending otherwise gives the illusion of precision without the reality of accuracy.*

For tactical and small-unit actions, performance directly generates predictable effects using Newtonian physics (ballistics, logistic flows, time and distance calculations, etc.) and the statistics of millenniums of documented lessons learned from small-unit engagements. Furthermore, the time lag between performance and effect is short, seconds to days. One can use measures of performance as proxies for measures of effect and rapidly check for the necessity of changes of plan, since valid numeric theories of physics and valid statistical theories of small-unit tactics exist. However, there are no valid numeric theories for dealing with operational and strategic levels of counterinsurgency; one must build logical connections between current actions and future effects on the objective and then generate the required metrics from the connections. *The obsession with objective assessments is tactical thinking applied to strategic problems.*

Although purely objective (and numbers-based) predictive theories of the physical world are possible, the likelihood that the same will become true for operational- and strategic-level complex social interactions—such as insurgency and counterinsurgency, terrorism and counterterrorism, and warfare—in time to be useful in Afghanistan is extremely small. Therefore, operational/strategic counterinsurgency assessment in Afghanistan must be subjective, based on senior leaders' subjective professional judgment of pertinent qualitative and quantitative data.³³ Even if all relevant data were available and all of

FIGURE 5



Artillery illustrations published before the existence of a numbers-based physics of gravity show impossible exterior ballistic trajectories. Artillerymen had to assess fall of shot subjectively, on the basis of objective measurements.

Norton, *Gunner*. See note 32.

them accurate, numeric, and objective, assessing what they mean for success is still a professional military subjective judgment call, since there is no credible, objective numbers-based theory of counterinsurgency.

DISTRUST GENERATED BY POOR ASSESSMENT PRACTICE

The military's focus on good-news stories, obviously flawed arithmetic and logic, and lack of transparency causes the press to distrust military statements. It then looks for and writes about what appear to be contradictions between military statements and reality; higher civilian authority then demands explanations from ISAF commanders. Both higher civilian authority and IJC accordingly demand "objective assessments using metrics to complement the commander's subjective in-the-field assessment" and then push metrics systems down the chain for collection. The message is clear: *they do not trust military commanders' assessments*. A reaction I observed was a demand from higher command to the relevant regional

command to provide good-news stories to counter media claims of military failure. This makes the military look defensive, makes the distrust worse, and spreads the distrust to the domestic population, upon whose political support for the mission the military relies for funding.

In my opinion, the number of metrics demanded overwhelms the collection capacity of regional commands' partner civilian organizations and major supporting commands. Furthermore, neither those organizations nor supporting commands appear to trust the value of collecting on those metrics or of assessments done using them. For example I was openly told by a head of planning in one civilian two-star-equivalent organization that in response to his regional command's request for assessment metrics he makes up what he does not have and does not check the quality of what he does have. An additional reason given me for not taking metrics seriously was the absence of feedback from requesting organizations. Another example is the attitude one colonel encountered when he asked troops in the field whether they "collected all the requested metrics or made stuff up"; the response was foot-shuffling and "Is this a trick question, sir?" Additionally, civilian partner organizations have expressed annoyance and suspicion of the military when the military lines of operation overlap civilian ones (such as governance and development).

WHAT IS TO BE DONE?

A regional commander and staff must separate the task of responding to higher commanders' requests for information to feed their assessment processes from the task of assessing the regional command's own progress with its own operational war-fighting objectives, to provide its own assessors time for the latter.

The pathologies inherent in higher command's assessment processes, specifically the enormous numbers of metrics demanded, mean that the regional command staff members responsible for responding to requests for information from higher command should not be the same people who must assess the regional command's progress with its own operational objectives. Otherwise there will never be time to do the latter, because the former inevitably comes first. If the same people have to do both tasks, the commander must set a maximum percentage of time they are to spend responding to requests for information from higher command in order to ensure they have time to produce adequate assessments for the regional commander.

Regional commands must decide what metrics are required to support assessment of their operational objectives. As we have seen, at the tactical level Newtonian physics and the statistics of millenniums of small-unit engagements provide a model for generating metrics and using those metrics to assess effects. At the operational level, there is no established numeric model for translating

performance into effect, and as argued, no credible numeric model is forthcoming. The regional command must build a qualitative understanding of the interacting elements of the operational environment in order to understand that environment, identify relevant assessment metrics, and erect a defense against externally imposed, plausible, but irrelevant metrics.³⁴ Build the model by analyzing commander's intent during the mission-analysis phase—that is, answer the question, “Why are the objectives important in the context of the objectives and associated desired end states?”—and collect the answers in concise narrative form. (If they cannot be expressed in clear English, slide-show bullets, “Pentagonese,” and cartoons just hide that fact.) The narrative generates metrics—“things we want more of and things we want less of”—with explanation, and clusters these into topics of interest. The narrative is the summary of subject-matter experts' opinions on how the environment works, within the context of the objectives, end states, ways, and means. The logic within the narrative provides a qualitative understanding of what is likely to happen as regional command actions alter the qualitative and quantitative values of the metrics. The assessment team then uses professional subjective judgment and the logic within the narratives to assess the implications of the collected metric information against the assessment question we have already postulated.

Identify Operational/Strategic Objectives and End States from the Relevant Operation Orders, FRAGOs, or Other Planning Documents. If the objectives are at the operational/tactical, or lower, level of war, use lines of operation instead of objectives. For each objective, identify the end states; if these are not explicitly available in the planning documents, analyze the documents for implied end states. Identify explicit and implicit “critical requirements” and “in order to” statements attached to the objectives in the planning documents; these are the critical requirements for success.

Write the “Assessment Question.” For operations in Afghanistan this will likely be of the form given above: “What is the likelihood of, and what are the risks to, the conditions for the specified end states occurring or remaining stable if the region transitions from coalition force control to GIROA?”

Analyze Commander's Intent. For each critical requirement, ask the question, “Why is this important to coalition forces, to the insurgents, to the population, to GIROA, to our partners, to our governments?”³⁵ There will likely be more than one answer, but restrict yourself to the important answers; they help identify strengths, weaknesses, and vulnerabilities for friends, enemies, and other stakeholders in the context of the specific objective. For each answer, in turn, ask the same question—“Why is this important?”—and continue until you can make a logical and clear link to the end state. Write a narrative expressing the chain of

links between the objective and the end state. There will be a temptation when analyzing intent to focus on “what we want to happen.” However, the real question is, “Why is this important?”—and so it is also necessary to consider what we *do not* want to happen. In addition, one should consider the most dangerous and the most likely actions or reactions of the stakeholders, including the “upsides” of any “downside,” and the downsides of any upside.³⁶

Use a variety of sources to generate the answers and to build the narrative, including planning documents, conversations and interviews with colleagues, internal advisers, external subject-matter experts and organizations, literature and databases, and your own professional military knowledge. Although regional command staffs are the experts in military matters, they are not the experts in politics, economics, social information, or infrastructure and must seek external assistance for these areas.³⁷

Note that the analyses of intent of the various objectives will overlap each other. This is expected, since multiple objectives apply to the same area of operations and support the same set of end states. Overlap in the analysis of intent represents the linkages between the objectives. However, one should write each narrative as though the reader were going to read only that one—introduce the overlaps but focus on the core objective (or line of operation).

Identify Topics and Metrics. For each narrative, identify the actors and their actions, as well as their strengths, weaknesses, and vulnerabilities implied in the narratives. Pay attention to “who is doing what to whom” and to “who wants, or does not want, what”—that is, to the verbs and nouns (hence the emphasis on declarative English). From these factors derive your topics.³⁸

Analysis of commander’s intent explains why these topics are important and guides collectors in the field. Analysis of the narratives concerning the topics will provide the specific metrics. For example, a critical capability for a governance-related objective might be “tax base of the GIRoA,” resulting in a “tax base” topic. Metrics might be “tax revenues,” “tax revenues skimmed,” “taxes avoided,” etc., to give the required picture of the “tax base” topic. Just as commander’s intent is part of military operational art, the selection of metrics to cover topics is also a matter of art guided by analysis of commander’s intent.

Each narrative for an objective may contain more than one discussion linking the objective to the end state. It may be useful to prioritize them within each narrative, thus prioritizing the topics and metrics, in order to discard less important topics and metrics if their numbers grow too large. Important topics may include some that are risky to collect (for example, likelihood of casualties suffered during collection), expensive to collect (because, for example, they require resources not currently allocated), or impossible to collect (inherently unknowable). Identify

these and inform leadership: uncollected pertinent metrics introduce risk to the quality of the assessment and hence to the credibility of the command and of the commander. “Unknowables” aside, it is up to the leadership to decide whether to accept the risk of an incomplete assessment by ignoring topics and metrics or to accept the risk of collecting on them.³⁹

Provide Guidance to Collectors. Provide collectors the topics (not the full collection of metrics) and the intent behind them (the analysis of commander’s intent); do not ask for numeric metrics;⁴⁰ ask for patrol reports, with whatever numbers and narrative they can provide about the topics. This approach avoids overburdening troops or other collectors in the field by pushing initiative down to the lowest possible level, and it allows collectors in the field latitude to interpret what they should collect and can deliver within the context of local conditions and commander’s guidance.

Make the Assessment. Argue the case both for (optimistic) and against (pessimistic) a successful outcome, and then make a final judgment based on the two cases. As would be done in legal proceedings, provide, along with the final judgment and the reasons for it, both “for” (optimistic) and “against” (pessimistic) arguments and all the evidence. If the resources are available, have separate teams do the optimistic and pessimistic assessments and argue their respective cases to a senior assessor for final assessment. Otherwise, do the pessimistic assessment first.

Gather all the evidence that supports the negative answer to the assessment question—that is, that the likelihood of success is low and the risks are high. Using professional military judgment, pessimistically assess the risk to different areas of the commander’s intent were the district to transition from coalition force to full GIRoA control on some specified date. Make as persuasive a professional argument as possible for a pessimistic answer; record it, along with the evidence. Be rigorous and ruthless when doing the pessimistic assessment; any squeamishness here will result in challenges to the final assessment in (probably) an embarrassing arena.

Then gather all the evidence that supports a *positive* answer to the assessment question—that the likelihood of success is high and the risks low. Again, using professional military judgment, assess the risk to different areas of the commander’s intent, but this time optimistically, were the district to transition from coalition force to full GIRoA control on some specified date. Again, make as persuasive a professional argument as possible, recording the evidence along with the optimistic assessment.

Finally, examine the two cases, their arguments and their evidence, and decide on an overall assessment of likelihood and risks. Pay particular attention to

pessimistic items that overwhelm positive ones and to positive items that fix negative ones. Record the final assessment and the reasons for it. When one produces good “for and against” arguments before making a final case, not only is the quality of the assessment improved but opposition to the assessment is more likely to focus on the interpretation of evidence and not the integrity or competence of the assessors.

The flaws in the operations assessment processes I observed in-theater clearly produce untrustworthy decision support; they are so manifest that commanders place their own credibility at risk when they support the resulting assessments. Regional commanders have the authority and means to fix operations assessment within their commands. However, doing so requires institutionalizing a rigorous process and separating it from the task of responding to higher-command requests for information. If the regional commander decides that this separation is unacceptable or does not have the time or staff resources to implement it, an alternative is to base the regional command’s operations assessment entirely on its commander’s subjective professional judgment combined with that of the region’s civilian provincial reconstruction team and of other regional stakeholders. The continued use of junk arithmetic and flawed logic robs decision makers of the most essential requirements that assessment is supposed to supply—sound, verifiable, and accurate information upon which to make life-and-death decisions.

NOTES

I thank the personnel of I Marine Expeditionary Force (Forward) for their friendship and assistance while I was with them in Afghanistan. They are truly ferocious fighters and courteous colleagues. I am grateful to Ben Connable, Jonathan Schroden, Peter Perla, other colleagues in the field too numerous to mention, and the faculty and staff at the Naval War College for their support and criticism.

1. U.S. Joint Staff, *Joint Operations*, Joint Publication [hereafter JP] 3-0 (17 September 2006, incorporating change 2, 22 March 2010), p. IV-30, available at www.dtic.mil/; U.S. Joint Staff, *Joint Operation Planning*, JP 5-0 (26 December 2006), p. III-57, available at www.dtic.mil/; U.S. Army Dept., *Counterinsurgency*, Field Manual [hereafter FM] 3-24 (Washington, D.C.: 15 December 2006), para. 5-90,
- describes assessment in similar terms; available at www.fas.org/ (FM 3-24 is also issued by Headquarters, U.S. Marine Corps, as Marine Corps Warfare Publication [MCWP] 3-33.5).
2. JP 3-0, p. IV-31.
3. Although “forecast” and “predict” are often used interchangeably, I use “forecast” as a broad description of a likely future rather than a precise “prediction” of exactly what will happen.
4. Note that one can replace “region” with “district” or “Afghanistan,” depending on the level of the assessment. The commander’s plan should specify the end states and will probably include items dealing with security, stability, economic development, governance, al-Qa’ida, etc. Also, the “specified date” can

- be any date that is useful, either in the future or the present. See JP 3-0, sec. IV, for a broad discussion of end states and objectives.
5. It is worth noting what the assessment cell's task is *not*. The assessment cell does not decide how to achieve the operational objectives or what to do to achieve the conditions for success. Those are planning functions. The assessment cell's mission is to assess how well the executed plan is achieving the end states that the objectives are supposed to enable and the risks to that achievement.
 6. The three lines of operation for coalition forces in Afghanistan are "governance," "economic development," and "security." Developing the Afghan National Police or army is a separate topic, and I do not deal with it in this article. See FM 3-24, chap. 5, for a broad discussion of the "Logical Lines of Operation" during counterinsurgency operations.
 7. Belief in pseudoscience and conspiracy theories and inability to use valid reasoning are disturbingly frequent in the American population—see "Science and Technology: Public Attitudes and Public Understanding," *National Science Foundation: Science and Engineering Indicators 2002*, www.nsf.gov/, esp. "How Widespread Is Belief in Pseudoscience?" It would be unwise to assume that excellence in leadership is incompatible with these kinds of thinking failure. Such thinking failures are not a cause for concern when those involved are facing familiar operational and strategic situations of the kind they have successfully dealt with in the past. However, experience with past operational (or strategic) situations is only as relevant to the current situation as the past and current situations are similar.
 8. See J. Scott Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners* (New York: Springer, 2001), for a good overall introduction to forecasting. See also *Forecasting Principles: Evidence-Based Forecasting*, www.forecastingprinciples.com/.
 9. I have also observed these flaws throughout both the U.S. Department of Defense and civilian commercial organizations—as would be expected, since the root causes are present throughout the Defense Department and civilian worlds.
 10. FM 3-24 is the closest the U.S. military has to a theory of counterinsurgency; however, it does not even come close to providing a numbers-based theory, and no one pretends that it does so.
 11. I use the broader, explanatory form of pessimism and optimism concerning value judgments on the agreed facts, rather than the dispositional form, concerning one's confidence in the success of an endeavor—hence the traditional "glass half-empty versus glass half-full" value judgment rather than any implication as to the effects of the glass's state of emptiness on the success of an endeavor.
 12. See, for example, Karl R. Popper, *The Logic of Scientific Discovery* (New York: Basic Books, 1959), and *Stanford Encyclopedia of Philosophy*, s.v. "Karl Popper," plato.stanford.edu/.
 13. See Richards J. Heuer, *Psychology of Intelligence Analysis* (Washington, D.C.: Center for the Study of Intelligence, Central Intelligence Agency), chap. 8, available at <https://www.cia.gov/>. Note that the very public arguments criticizing analysis of competing hypotheses (ACH) deal mostly with failures to execute the method properly or with the difficulties of executing the method. See Tim van Gelder, "Hypothesis Testing: What's Wrong with ACH?," *Tim van Gelder: Bringing Visual Clarity to Complex Issues*, timvangelder.com/, and "How Not to Free Your Mind," *The Interpreter: Lowy Institute for International Policy*, www.lowyinterpreter.org/. Both these criticisms hold for any valid analysis method, and the idea that the validity of a method depends on the ease of its use is ludicrous.
 14. See, for example, *Stanford Encyclopedia of Philosophy*, s.vv. "Naturalized Epistemology" and "Underdetermination of Scientific Theory."
 15. The use of negative words such as "fail," "disprove," and "weakness" may be problematic to some people's professional ethos. However, the rules of logic are not optional, the value of the scientific approach is proven, and one ignores either at one's peril.
 16. Maj. Jonathan Roginski (10th Mountain Division) informs me that "under the guidance of William Upshur and David Kilcullen the Afghanistan Assessment Group of Regional Command (South) replaced RODEA"

- in September 2010 “with a system of 17 indicators—informing 11 metrics—describing the environment and critical conditions specific to southern Afghanistan in the spring and summer of 2011, narrowing the process to only that which is central to the environment and the mission.”
17. The rules of arithmetic—including the fact that adding or averaging rank orders is nonsense—were established over two millennia ago by, among others, Pythagoras and are taught in every elementary school worldwide.
 18. The attempt to get around this by scoring metrics using Likert-like items (for example, the five-point rating definition level) fails, since with every point defined by a text description the numbers associated with each text item are rank-ordered ordinals that, by the rules of arithmetic, cannot be averaged (or have any other arithmetic function used on them).
 19. It may be that in certain instances one can replace a brigadier general by a major paired with a captain, but I suggest that in these cases one has other problems that are beyond the scope of the assessments process.
 20. These margins of error appear to be those that would be computed from the population sizes if there were no corruption, fraud, or intimidation involved in the data collection. Therefore, these margins of error are at best the minimum error, and in my opinion it is highly probable that they are much greater within the context of Afghanistan.
 21. “IJC Metrics Assessment” (IJC briefing to the ISAF Joint Command Metrics Workshop, Washington, D.C., 17–18 March 2010), slide 15. Note that, in addition to the five colors, “white” indicates “not assessed.”
 22. By Ben Connable, at discussions held 5 May 2010. See also Ben Connable, “Afghanistan Assessments and Metrics: Trip Report Analysis and Recommendations for ISAF and ISAF Joint Command” (document MG 1086, RAND Corp., 2011).
 23. “IJC Metrics Assessment,” slide 29.
 24. See PA Consulting Group, *Dynamic Planning for COIN in Afghanistan* (London: 2009), available at msnbcmedia.msn.com/.
 25. They may very well be important or even vital to *other* commands or government agencies, but unless they are tied to *this* command’s objectives, they are not important to *it*.
 26. Other than the promiscuous approach of “This metric appears to have *something* to do with the objective, so therefore we must collect on it.”
 27. IJC guidance to the ISAF Joint Command Metrics Workshop.
 28. Coffey International Development, “Introduction to the Helmand Monitoring & Evaluation Programme,” Helmand Provincial Reconstruction Team, Afghanistan, October 2010.
 29. In scientific parlance, a “theory” is not an unproved speculation but a hypothesis (or a system of hypotheses) that has been so verified through testing that to deny it contingent upon acceptance would be perverse, where a “hypothesis” is an untested (but testable) proposal for how some part of the world might work.
 30. Dr. Jonathan Schroden, in personal conversation, 4 January 2011.
 31. For the figure of four hundred districts, see *Statoids*, s.v. “Districts of Afghanistan,” www.statoids.com/. For outcomes, Dr. Jonathan Schroden points out (note 30) that one could look to some of the districts in Kabul Province to see what “success” looks like—though he counterargues as well that Kabul is a specialized case, given how urban it is compared to the rest of Afghanistan.
 32. See, for example, the trajectory illustrations in Robert Norton, *The Gunner* (n.p.: Humphrey Robinson, 1628), and in Diego Ufano, *Artillerie* (1621), information on both available at www.biografiasyvidas.com/biografia/u/ufano.htm, www.mpiwg-berlin.mpg.de/en/research/projects/DEPT1_Renn-ExpansionTheoreticalMechanics, and www.uh.edu/engines/nortontrajectories.jpg. It was not until over half a century later, in 1687, that Isaac Newton provided an objective numbers-based theory of ballistics in his *Principia*.
 33. A combination of diplomatic/political, informational/ideological, military, and economic leadership and expertise must be involved.

34. “A comprehensive systems perspective considers the interaction between the individual elements of a system and across multiple systems (political, military, economic, social, informational, infrastructure, and others)”;
- JP 5-0, p. III-22. “Joint planners analyze the operational environment in terms of six interrelated operational variables: political, military, economic, social, information, and infrastructure”;
- U.S. Army Dept., *Operations*, FM 3-0 (Washington, D.C.: February 2008), p. 1-5, available at downloads.army.mil/.
35. This does not “second-guess the commander”; it is a drilling-down into the details of the commander’s intent, placing them within the context of the environment and using the commander’s objectives and staff planning documents as the primary inputs. It states what it is about the environment that is critical for a valid assessment; it covers political, military, economic, social, informational/ideological, and infrastructural issues, identifying the relevant stakeholders (own forces, coalition forces, friendly forces, enemy forces, the local population, etc.). It is analogous to the “intelligence preparation of the operational environment” carried out during mission planning.
36. For example, the downside of damaging Taliban finances by poppy eradication might be an upsurge in murder and intimidation by the Taliban to make up for its inability to buy labor from the local population. The upside of Taliban murder and intimidation, in turn, might be an opportunity for the coalition to engage in information operations.
37. Since for most topics there is a surplus of subject-matter experts, grouped into several “schools of thought” that often contradict each other, the choice of which “school” to incorporate into assessments amounts to a command decision. The difficulty is to balance subject-matter experts whose views match the regional commander’s or who are “popular” with those holding opposing or unpopular views. See Philip Tetlock, *Expert Political Judgment* (Princeton, N.J.: Princeton Univ. Press, 2005), for an analysis of how the popularity of an expert is a poor measure of the quality of that expert’s forecasts.
38. For the purposes of this article, “topic” refers to a broad subject (for example, education), and “metric” refers to an item of data (qualitative or quantitative) whose collection tells us something about that broad subject (for example, the number of schools).
39. The latter may require reallocation of resources and perhaps adjustment of the plan.
40. An officer briefing several multistar generals in May 2010 told them there were seven schools in the area. However, another credible source had previously stated that there were three, although no one mentioned that figure at the briefing. It was only afterward that the discrepancy was tracked down: there were three “brick” schools and four “tent” schools. Asking for the “number of schools” will produce a number, but if the quantity of metrics demanded is too high (which it currently is), amplifying and useful narrative will not result. Asking instead for the topic—“Describe the state of schools in your area”—and explaining in terms of the commander’s intent why we are asking is much more likely to generate useful information.