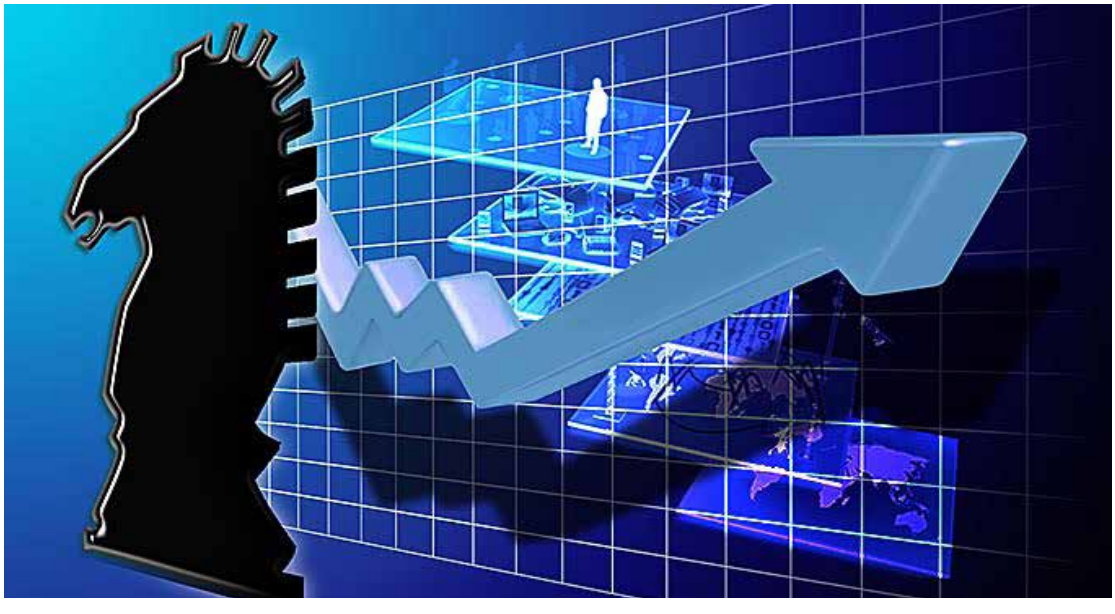
Assessing ISR

Effectively Measuring Effectiveness

Capt Brian A. Hill, USAF



The first step is to measure whatever can be easily measured. This is OK as far as it goes. The second step is to disregard that which can't easily be measured or to give it an arbitrary quantitative value. This is artificial and misleading. The third step is to presume that what can't be measured easily isn't important. This is blindness. The fourth step is to say that what can't be easily measured really doesn't exist. This is suicide.

—Charles Handy's description of Secretary of Defense Robert McNamara Vietnam-era measurement policies, *The Empty Raincoat: Making Sense of the Future*

Intelligence, surveillance, and reconnaissance (ISR) is a critical pillar in decision making, a key driver of operations, and in many ways an operation unto itself. Like all operations, it has an intended effect—generally to inform, shape, and facilitate other operations by providing decision advantage. Joint Publications 2-0, 3-0, and 5-0 all emphasize the importance of assessing operational effects. Assessing ISR effectiveness is often more complex than combat assessment, but it is no less important.[1] This is widely recognized, but although some initiatives are ongoing, not much progress has been made toward correction. Current guidance, including the Joint Publications previously mentioned, offers broad direction on who has the responsibility for assessments and a general framework for what assessments should look like, but literature explaining the nuts and bolts of actually producing effective assessments is scarce.

The RAND Corporation continues to conduct significant research into this domain to "develop detailed mathematical quantities that represent what are generally considered qualitative concepts."[2] During this work, RAND designed several mathematical models for evaluating improvements across ISR capabilities, but they are currently incomplete, appear specific to broad capabilities assessments, and likely cannot be rapidly adapted and used by all units who could benefit from a robust ISR assessments program. The Air Combat Command (ACC) also released an "ISR Assessment Framework" at the end of 2016 which brings guidance that is more applicable to operational and tactical assessments and is more detailed than the Joint Publications listed above.[3] However, this framework still stops short of breaking down the hands-on steps for the assessors themselves.

This article seeks to fill that gap. First, it briefly reviews the problems with ISR assessments, it offers recommendations for breaking down goals and tasks into effective measurements, and it gives examples of applying these methods in an ISR context. It finishes with additional considerations that will enhance ISR assessments, including changing the way we interpret numerical values in the assessments and, most importantly, by establishing a dedicated training program for ISR assessments.

Current guidance breaks assessments into measures of performance (MoP) and measures of effectiveness (MoE). According to JP 1-02, MoPs are criteria used to assess friendly actions that are tied to measuring task accomplishment, such as whether a sortie arrived at a location on time, or whether it collected all of its assigned images. MoEs are criteria used to assess changes in system behavior, capability, or operational environment that is tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect.[4] For example, if the goal of an ISR mission is force protection, one possible MoE is how often a forward operating base (FOB) is attacked without warning. A common distinction is "MoPs ask if we're doing things right; MoEs ask if we're doing the right things." MoPs are most useful at the tactical level to units executing ISR operations, while MoEs are most useful to those planning and coordinating ISR operations, but both are inherently connected, and ISR assessors at all levels should constantly coordinate to develop holistic assessments.

An overreliance on MoPs is a recognized problem in ISR assessments, illustrated in examples of a system focused on whether collection occurred and not whether it

met the commander's intent.[5] The reason for this overreliance is ease: the percentage of tasked targets collected and number of sorties flown are obvious, easy to measure, and require no extra effort to quantify. In every assessments discussion, working group, and conference I've attended, everyone recognizes the need to focus on good MoEs to determine if ISR is achieving our desired effects. Some progress is made by individual units, but none have created an objective, quantitative, and repeatable method for developing proper MoEs.

One of the biggest roadblocks to proper MoEs is the pervasive idea that they are inherently qualitative and subjective, in contrast to easily quantifiable MoPs. Some assessment teams compensate with "false quantification," where they assign numerical values to subjective ideas. For example, they may use a weighted scale that assigns values based on what they feel is the "significance" of the collection: a zero if they deem it to have no significance, a one for low significance, and so forth. This is a somewhat backward way to quantify. It assigns significance to the intelligence and uses that to determine if the intelligence was effective, but, in reality, we can't know if it was significant unless we know it had an effect. This can result in a "false impression of accomplishment," noted in JP 5-0's Appendix D.[6] While this is still an improvement—since it acknowledges the need for objective quantification—the actual method is still subjective and privy to the opinions and moods of the analysts making the assessment. As Handy stated, "The second step is to disregard that which can't easily be measured *or to give it an arbitrary quantitative value*. . ." (emphasis added). To be clear, qualitative assessments and subjectivity do have a place, but only after a proper quantitative assessment is done, which is demonstrated later in this article.

There are two main assumptions people make in assessments that contribute to the myth of MoEs being inherently qualitative:

1. The thing being measured is not and cannot be well-defined, so objective quantification is impossible.
2. The method of empirical observation doesn't exist or isn't known.

Solving the first is relatively straightforward but can require significant intellectual effort. Problems must be well-defined, and once well-defined, they can be broken down into quantifiable measurable factors. Once they have been broken down, the second problem usually solves itself—often more easily than anticipated. Instead of focusing on these two problematic assumptions, assessors should start with these basic principles, adapted from Douglas Hubbard's model for measuring intangibles in business:[7]

1. If it matters, then it must have a detectable or observable effect.
2. If it is detectable, then it can be detected as an amount or range of amounts.
3. If it can be detected as an amount, then it can be measured.

An example in the business world is "employee empowerment." This is certainly not well-defined and by itself not obviously quantifiable. We have to determine what measurable factors we expect to see if employees are empowered. Empow-

ered employees should be able to make decisions at lower levels, which should mean faster decisions, quantifiable by time. Likewise, empowered employees should be open to formulating and pursuing their own ideas, so we can measure the number of new independent projects springing up. Both of these result in hard numbers, not arbitrary weights or gut feelings.

The Measuring Progress in Conflict Environments (MPICE) framework brings this closer to the military intelligence sphere. It provides excellent examples of how to break down less tangible concepts, like diminishing political grievances into specific measurable factors—the number of incidents of political violence, prosecution rates, and percentages of representation for various identity groups.[8] In this framework, nearly every measurable example used, even if not identified as "Quantitative Data," is in fact in a quantitative and objective form. This doesn't directly translate to ISR; by its nature, ISR generally requires additional levels of assessment since it informs the action that produces the effect, rather than producing the effect itself. However, the following examples use a similar approach to the MPICE framework, combined with Hubbard's model for business, to achieve a usable, quantitative assessment.

Quantifying Goals

Let's take a simplified ISR scenario: the fictional nation of Wadiya has made a number of threats to its neighbors, including US allies. It has five airfields, all of which can support bomber operations and three of which currently host bombers. The nation has three road-mobile ballistic missile garrison locations with 10 known dispersal/deployment sites for each, or a total of 30. Each garrison location owns 10 launchers. There are three fixed surface-to-air missile (SAM) sites and three more air defense units with mobile SAMs, each of which has four launchers and one radar.

The USAF is conducting two daily sorties, one collecting signals intelligence (SIGINT) and the other collecting geospatial intelligence (GEOINT), on Wadiya with the following goals:

1. Provide indications and warnings (I&W) of Wadiyan attacks.
2. Prepare the battlespace for strikes in the event of Wadiyan hostilities.

The MoPs are easily established, as usual. Did the sorties arrive and depart on time? Did they collect all assigned images (GEOINT) and total tasked hours (SIGINT)? How many additional (ad hoc) images were taken? Were there any cross-cues between the intelligence disciplines? These are all valuable questions. However, success across these performance metrics does not mean we are achieving our desired goals. We need MoEs. Let's take a look at our desired goals again:

1. Provide I&W of Wadiyan attacks.
2. Prepare the battlespace for strikes in the event of Wadiyan hostilities.

If we ask if the ISR provided I&W, or if it prepared the battlespace, then the slightest bit of intelligence can make the answer a "yes" depending on the point of view of the analyst. Here's where the subjectivity myth comes into play since the

objective measures aren't obvious. One analyst may look at a day's collection and say "this contributed greatly," while another may say it gave us something, but not anything particularly interesting. Instead, we need repeatable, objective measures. To do that, we have to break down the goals, much like how priority intelligence requirements are broken down into essential elements of information (EEI). In fact, if that process is done exceptionally well, the EEIs themselves can be quantifiable MoEs, but this is not always done correctly. In any case, the breakdown could look something like this:

1. Provide I&W of Wadiyan attacks.
    A. Monitor weapons of mass destruction (WMD) posture.
        1. Track location and posture of ballistic missiles.
        2. Track location and posture of bombers.
2. Prepare the battlespace for strikes in the event of Wadiyan hostilities.
    A. Monitor locations of attack capabilities.
        1. Track location of ballistic missiles.
        2. Track location of bombers.
    B. Monitor location of air defense capabilities.
        1. Track location of mobile SAMs.
        2. Monitor status of fixed SAMs.

Some of these ended up as repeats, leaving us with six items we're trying to track: (1) the location of ballistic missiles, (2) the posture of ballistic missiles, (3) the location of bombers, (4) the posture of bombers, (5) the location of mobile SAMs, and (6) the status of fixed SAMs.

Now we determine what constitutes an ideal "effect" state for each. For the first, recall that there are a total of 30 ballistic missiles launchers, 30 known dispersal sites, and 3 garrison locations. Naturally, we want to know the location of all 30 launchers. Did we successfully locate and image all 30? If so, then that day's ISR was successful at achieving the *effect* of informing leadership of the location and posture of the ballistic missiles. The key is to come up with questions that have quantifiable answers.

1. The location and posture of ballistic missiles
    A. How many of the known launchers were located (xx/30)?
        1. How many were imaged?
        2. How many were found by SIGINT but not imaged?
    B. Were any previously unknown launchers discovered, and how many?

Now say that we only got 28 out of 30, and sensor limitations prevent us from imaging three of the dispersal sites. The GEOINT was not 100 percent effective in

achieving its effect, but say the SIGINT platform was able to collect daily communications between the garrison and the dispersal site that located the 2 missing launchers on a dispersal exercise. On the one hand, even though the GEOINT platform alone wasn't 100 percent effective, the overall ISR was. On the other hand, perhaps every known site was imaged, two launchers were still missing, and SIGINT didn't fill in the gap. Something prevented 100 percent effectiveness, but now the cause is less obvious. Perhaps analysts failed to identify equipment at a site, or there's an undiscovered dispersal site. Leaders, planners, and analysts must now determine if and where adjustments need to be made to close that gap. In all of these cases though, the effectiveness was determined with objective, quantifiable MoEs.

The order of battle intelligence is fairly straightforward to measure in this way. A more difficult example is force protection. Unlike known equipment in a country, it's impossible to say how many attack plots one will discover before 100 percent are collected. Even if no surprise attacks occur, there may be undiscovered plots that are never executed. In this example, we can demonstrate the appropriate use of qualitative assessments and subjectivity after the quantitative assessment occurs. Context and qualifiers must be added, but the foundation of the assessment should still be completely quantitative, and the method for determining that quantification is the same.

Qualified Quantification

The primary goal of force protection ISR is to provide intelligence in order to protect a base or unit. For this demonstration, we'll break that down into discovering vulnerabilities and threats. Then we break those down into measurable factors, just like the previous example.

1. Discover weaknesses.
    A. How many gaps in perimeter defenses were found?
        1. How many were corrected?
        2. Were any exploited by an attacker prior to discovery? If so, how many?
2. Discover threats.
    A. How many external attack plots were discovered?
    B. How many attacks occurred without warning?
        1. How many of those attacks appeared to be planned versus spontaneous?

Again, each of these questions have objective, quantifiable answers. The numbers used to answer them are not analyst opinions, they are hard facts. However, they illustrate where in the assessments process subjectivity and qualifiers come into play. In this scenario, there is no "100 percent" goal because it is impossible to determine, but this is a key point: in many cases, there doesn't have to be a target number. We must divorce ourselves from the idea that certain numbers are inherently good

or bad and just use them as what they are: data to drive a decision. Using "discover weaknesses" as an example, say we found five gaps in perimeter defenses during the first week, one the week after, and none after that. Clearly the quantitative effectiveness of that ISR sortie in discovering weaknesses is down, but that doesn't mean there's anything wrong with it. Here we can probably add the qualitative assessment that there are simply fewer weaknesses left to discover, if any. Now ISR planners and base commanders must make a subjective recommendation and decision respectively. Is there anything we can adjust in our ISR that may allow further discovery? Also, is it worth it to keep that sortie examining base defenses, or should it be moved to a different task? The numbers themselves are still purely objective, it's in the recommendations and reactions stemming from those numbers where subjectivity comes into play.

Scaling Up

The previous examples are simplified and predominantly tactical in nature, but these principles can be scaled up to more complex scenarios or to operational and strategic levels, although it requires more intellectual effort to ensure the assessment is still based on objective, quantitative measures. Much of the conversion from a lower to higher level of warfare centers on the fusion of metrics across the battlespace. As noted in a 2014 article by Col Jason Brown, "the adversary's primary objective, or end, is *not* to shoot down aircraft; it is to *prevent* getting bombed."[9] We can take a similar thought process to convert the previously discussed tactical MoEs to operational MoEs. Let's go back to the force protection example.

The operational goal of force protection ISR is not to find vulnerabilities, it is to improve the security of US forces. One of the tactical-level quantifiable measures we used was the number of vulnerabilities in base defenses identified with the implication that they are then fixed. Let's assume similar ISR missions were flown around three additional FOBs across a region, with similar results. If after correcting all identified vulnerabilities, the bases experienced fewer successful attacks, we have a quantifiable, objective MoE at the operational level that force protection ISR is achieving its desired goals. If, instead there is an increase or negligible change in successful attacks—either across the board or at some of the bases—then that implies the operational objective is not being met despite the tactical ISR success at one or more bases.

Additional Considerations

Five changes will go a long way to improving ISR assessments across the Air Force. The first is already illustrated: changing the mindset that certain numbers are inherently good or bad. The previous two examples show when numbers should be considered that way, as in tracking 100 percent of the order of battle, and when they are neutral data to feed a subjective decision, as in force protection. This can be difficult to solidify. Once something is designated as a measure, our first inclination is to maximize that result. If, in the force protection example, one ISR planner

oversaw the first week and found five gaps in defenses and a new planner oversaw the second week and found only one, someone can imply the new planner was less effective, but this is an incomplete interpretation of the numbers. This is similar to units and ISR operations teams that tout their ability to add more and more targets to their collection decks regardless of whether those targets have value. Sometimes there needs to be a goal number, and sometimes more is better, but assessors and commanders both must recognize when that is not the case.

The second change is a shift to structured data formats for intelligence when possible, making it easy to query, discover, and add to the database. This is especially useful in the order of battle collection. In a real-world scenario, the time required to comb through every text report and image to count up the number of SAMs found during each collection mission and compare it to the known order of battle is astronomical. Shifts to structured data would make it much easier to search and compile and will turn this into a simple task taking only a few minutes. The intelligence community has tried shifting to structured data approaches before with varying levels of success. This is once again gaining momentum in the ISR community via structured observation management tools and elements of activity-based intelligence initiatives. Continued incorporation of advances in data science and artificial intelligence will further accentuate these benefits. This will not streamline all measurements for all types of ISR assessments, but will remove a massive manpower burden for some of the most tedious tasks.

Third is a requirement for constant communication and feedback. Customers, the air operations center, the collection units, and the production units must all constantly exchange information on the results of ISR. An exploitation node won't necessarily know about friendly changes to the battlespace unless those engaged in that space communicate. I repeatedly saw units identify and correct this problem during Operation Enduring Freedom. The change always brought about benefits to ISR assessments but eventually all units seem to slip back into not communicating. The necessity of customer feedback is a vital component in assessing ISR.

Fourth, we must recognize the need to continuously reexamine previous assessments of ISR performance in phases further and further removed from the ISR itself. The first effect of ISR is generally to inform, but the true effect is removed by additional steps. At a tactical level, a strike may have the effect of killing a high-value target; this effect is one step from the action that caused it. The ISR that led to the strike is removed by an additional step. This chain gets longer as the level of the effect transitions from tactical to operational and beyond, with the causal ISR always requiring additional connections beyond a combat assessment. For example, if ISR identifies a target that is later struck, that is often used as a marker of "effective" ISR. However, the next step must be assessing the effects of that strike. If ISR identifies a series of supply lines and shows that "destroying these will severely degrade insurgent materiel stores," we can't know for sure if that was accurate until we see indicators of reduced insurgent supplies *after* the strikes. Likewise, we'd then want to know whether the reduced supplies actually reduced insurgent activities. This chain becomes still longer for tracking operational-level intelligence and beyond. We're now several steps away from the original ISR, yet this remains a marker of the effectiveness of that ISR and the accuracy of the analysis based on it, and if

broken down properly can be measured in a purely quantitative way. It's possible these assessments take place at each piece of the chain, but the ISR unit, or units higher in the chain, may not follow up to link the pieces together in order to truly understand the effectiveness of the ISR.

Finally, ISR assessments must become a dedicated skillset. The ACC's recent ISR Assessment Framework notes the importance, and it bears repeating here.[10] This is the most important of these five additional considerations and replicating good ISR assessments across the Air Force depends on it. The rise of big data and analytics offer a huge opportunity for effectively quantifying measures and assessments. Despite the daunting vision of advanced math, using these tools for ISR assessments doesn't require a degree in the subject. However, it does require familiarity that many intelligence analysts don't currently possess. Further, the art of breaking down overarching goals into objective metrics requires training and practice. Different units have different ways of doing assessments due to varying mission sets, but the basic mindset and principles should be the same across the board. Also, the joint nature of operations necessitates a joint understanding of effects. Thus, a standardized training program for ISR assessments that incorporates joint capabilities, quantitative effectiveness assessment frameworks, and familiarization with data science and quantitative measures should be established.

Conclusion

Proper ISR assessments are vital to maximizing the effects of our ISR in a resource-constrained environment. MoPs are simple and widely used, but good MoEs remain a problem. To create good MoEs, we must stop seeing them as inherently qualitative and subjective and develop measurable, quantifiable, and objective ones. This can be done by thoroughly defining the goals of an ISR operation and breaking those goals down into objective, measurable factors with the mindset that any observable effect can be measured. This process can be intellectually intensive and requires creativity at times, but it is always possible. These purely quantified measures must form the core of assessing effectiveness, but it can then be framed with qualitative information to inform subjective recommendations. To get there, we must make several changes in our thinking. Most importantly, a robust training program for ISR assessors must be put in place to teach the art of establishing good MoEs and the science of knowing how to measure them. Intelligence is meant to provide decision advantage. Good measurements are the foundation of good decisions, and recognition of these points is the next step to strengthening the decision advantage that our war fighters deserve. ✪

Notes

1. US Joint Chiefs of Staff, Joint Publication 5-0: *Joint Operation Planning*, Appendix D (Washington, DC: US Department of Defense, 2011), D-6.

2. Walter Perry, David Signori, and John Boon, *Exploring Information Superiority: A Methodology for Measuring the Quality of Information and its Impact on Shared Awareness* (Santa Monica, CA: RAND Corporation, 2004), xxv.

3. Air Combat Command (ACC), "ISR Assessment Framework," 27 September 2016.

4. US Joint Chiefs of Staff, Joint Publication 1-02: *Department of Defense Dictionary of Military and Associated Terms* (Washington, DC: US Department of Defense, 2010), 151.
5. Jason Brown, "Strategy for Intelligence, Surveillance, and Reconnaissance," *Joint Force Quarterly* 72, (January 2014), 40.
6. US Joint Chiefs of Staff, Joint Publication 5-0: *Joint Operation Planning*, Appendix D (Washington, DC: US Department of Defense, 2011), D-9.
7. Douglas W. Hubbard, *How to Measure Anything: Finding the Value of "Intangibles" in Business*, 3rd Edition (Hoboken, NJ: John Wiley and Sons, Inc., 2014), 39.
8. John Agoglia, Michael Dziedzic, and Barbara Sotirin, *Measuring Progress in Conflict Environments (MPICE): A Metrics Framework*, (Washington, DC: US Institute of Peace Press, 2010).
9. Brown, "Strategy for Intelligence," 40.
10. ACC, "ISR Assessment Framework," 15.

Capt Brian A. Hill, USAF

Captain Hill (BS, Syracuse University; MA, American Military University) is the senior intelligence officer and an instructor and course manager at the joint Defense Nuclear Weapons School (DNWS) at Kirtland Air Force Base, New Mexico. He is responsible for running the Nuclear Policy Course and the Joint DOD/DOE Nuclear Surety Executive Course, DNWS's flagship course for senior nuclear leadership. Before this assignment, he was the assistant director of operations for the 693rd Intelligence, Surveillance, and Reconnaissance (ISR) Group, where he was in charge of establishing the group's ISR assessments program and represented the unit at the Twenty-fifth Air Force's ISR Assessments Working Group. He also served as the interim director of operations and flight commander of multiple signals intelligence and geospatial intelligence operations flights, 24th Intelligence Squadron.