ASSESSMENT CRITERIA

Last Updated: 04 November 2016

Criteria define the attributes and thresholds for judging progress toward the end state and accomplishment of required tasks. Development of assessment criteria is the critical component of the assessment process and should be accomplished before specific measures or data requirements are defined. Developing measures without a clear understanding of how they fit into a judgment of the effectiveness of an overall strategy often leads to laborious data collection and analysis processes that provide little to no value to decision-makers. Spending additional time to thoroughly consider and develop meaningful and relevant assessment criteria help avoid this pitfall.

Criteria help focus data collection by ensuring that assessment measures relate clearly to the elements of the strategy being assessed. As data are collected, the criteria translate those data into meaningful insights on the commander's strategy, which may be presented in a variety of ways to visually display progress (or lack thereof) to the commander.[1] Criteria should objectively indicate trends of significance and should be things that can be measured by known means. Determining them prior to commencement of operations allows for the establishment of baseline values for friendly and adversary forces and actions, which will facilitate objective reporting of changes, as well as rates of change.

All criteria have strengths and limitations. Which is used will depend in some part on the personality and preferences of the Commander. However, a variety of means should be used to comprehensively display progress toward (or away from) objectives and avoid losing relevant data by artificial form limitations. Criteria should be developed for the ends, ways, and means at each level of assessment. Well-written criteria should adhere to some basic attributes:

- **Relevant to the effect or action being assessed**. The criteria should relate directly to the commander's end state, tasks, and success thresholds as outlined in the strategy.
- **Mutually exclusive across the assessment categories** (e.g., good, marginal, poor) for a given effect or action assessed. This ensures that only one category is appropriate for a given outcome.

[1] These may include a wide variety of presentation formats, as detailed in Air Force Tactics, Techniques, and Procedures (AFTTP) 3-3.AOC, *Operational Employment-Air Operations Center*.

- **Collectively exhaustive across the range of outcomes for a given effect or action**. This helps ensure that most, if not all, potential outcomes are covered by the criteria.

- **Well-defined**. Specific and relevant definitions should be developed for any confusing or ill-defined terms used in the criteria. Planners should attempt to define success thresholds and the boundaries between assessment categories objectively whenever possible (e.g., what are the criteria for transition between the 'good' and 'marginal' categories?). Nonetheless, judgment is always necessary when assessing the overall strategy.

For example, if the commander's objective is to gain and maintain air superiority in a given operational area, criteria for the ends (i.e., objectives and other effects) should directly address to what degree enemy air defenses have interfered with friendly operations. Planners should select criteria that give the commander meaningful insight into the degree of interference and use these criteria to judge progress toward the objective. Similarly, planners should determine meaningful criteria for establishing whether the tasks undertaken to achieve air superiority have been accomplished. In this example, the commander and planners would want to know if enemy air or air defense operations occurred, whether or not they posed risks to friendly air operations, whether or not air bases, surface-to-air missile sites, and radars were manned, operating, communicating, or emitting, or if such forces were moving.

Some additional criteria selection guidelines may help planners:

- The lines between categories are often hard to determine, especially with some commonly used assessment display techniques like "stoplight" charts (for instance, it may be hard to answer, *"when do we go from good* ["green"] *to marginal* ["yellow"]?") Planners should set objective and concrete boundaries as much as possible, recognizing that some degree of subjectivity (and hence judgment) will always be necessary.

- Try to select criteria that allow depiction of trend data, which may ultimately be among the most meaningful criteria. (For example, "*effectiveness is still marginal on this air tasking order (ATO), but the trend is rapidly improving, so we can probably allocate a lower weight of effort to air superiority on future ATOs, despite the current status*.")

- Try to avoid arbitrary terms like "some," "prohibitive," and "significant." They do not lend themselves to objective definition. (In the example above, for instance, criterion boundaries could hinge on percentages of desired area, mission-capable assets, and desired timeframe.)

- Sliding scales can often be a useful display format, since it helps show relative magnitude of differences. For instance, on a one-to-ten scale, eight may not be much better than seven, but is considerably better than five, even though eight may be "good" or "green" on a stoplight chart, while both five and seven are "marginal" or "yellow."

When assessing complex military operations, subjective data in the form of subject matter expert (SME) inputs often provide the most meaningful (or only available) data. To avoid personal biases and ensure an adequate level of consistency in the assessment, SME inputs should simply provide the information necessary to address

the relevant measures of effectiveness. For example, when assessing the achievement of air superiority, it is more effective to ask a SME about the degree to which adversary air has interfered with their operations, rather than asking directly whether the Air Force has achieved an appropriate level of air superiority.