# LESSONS LEARNED FROM

# THE DEBRIEF

BY JON BECKER



**Elite units spend a great deal of time and money studying and refining their process to ensure they select the right people and help them achieve the required performance.**

## THE SCIENCE OF SELECTION: LESSONS LEARNED FROM ELITE UNIT A&S

**PART ONE**

For the past 15 years, John Dowd has worked at the intersection of elite tactical performance and personnel selection. A former Navy SEAL, Dowd founded SOFware LLC, a specialized consulting and software company that works with the most elite military and law enforcement units, helping them to design and refine how they identify, evaluate and develop talent. From this unique vantage point, Dowd has developed insights into the assessment and selection (A&S) methodologies that are most effective for high-performance organizations.

At its core, A&S encompasses the hiring and promotion process used by an organization. A&S is how people are hired, assigned, evaluated, promoted and even removed from the organization. As a result, elite units spend a great deal of time and money studying and refining their process to ensure they select the right people and help them achieve the required performance.

This is the first of a two-part series that will explore what Dowd and his team have learned from 15 years of working with our nation's most elite units. In this part, we will examine the definition of A&S, its role in establishing team culture, and the key components for an effective A&S process. Part two will look at common pitfalls and errors made in A&S processes. While there is certainly no perfect way to conduct A&S — and the process does (and should)
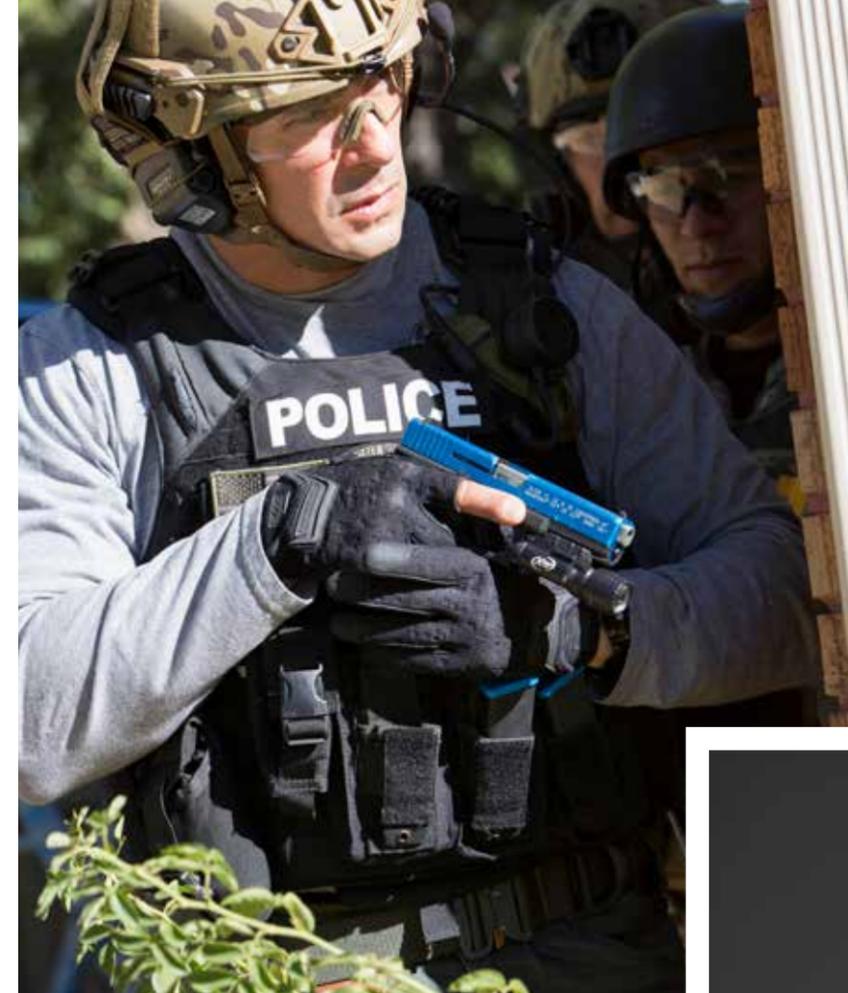
vary from team to team — the recommendations discussed in this series are anchored in the best practices they have seen work for high-performing units.

### A&S – Defining your culture

The foundation of assessment and selection is the understanding that building and maintaining organizational culture is the cornerstone of elite performance. Put another way, A&S is where organizational values meet operational reality, and without an effective A&S process, high levels of performance are extremely difficult to achieve.

While it is obvious that tactical units need individuals capable of performing exceptionally difficult jobs, they often overlook how essential A&S is to that process. As the old saying goes, "you can't make a silk purse from a sow's ear." Achieving the high levels of technical skills, physical capabilities, and psychological resilience required by tactical operations demands an A&S process that accurately identifies, nurtures and develops candidates who are able to achieve those performance levels. Perhaps less obvious than individual performance, but more important, is the understanding that A&S helps to shape and define organizational culture.

By systematically choosing individuals who embody specific values and behaviors, organizations define their identity. The people a unit selects for membership, how they

choose to evaluate their team members, and how they define their standards are all rooted in A&S. It is not surprising, then, that the methodologies developed by elite units are not just about finding competent performers, but also creating a process that defines, reinforces and perpetuates organizational culture.

For decades, tactical units have defaulted to the idea that by making selection difficult, the right people will rise to the top. But toughness and grit, valuable as they are, are not the same as suitability. As Dowd put it, "You can be tough and toxic." Elite organizations in the special operations community learned this years ago. They realized that the hardiest candidate wasn't always the best teammate, leader or long-term cultural asset. Now, many of these units have moved beyond simple physical selection processes based on "who survives" and also have started selecting based on "who contributes."

### Deciding who decides — choosing the evaluators

The first critical step in building an effective A&S process is to ensure it is being conducted by the right people. The quality of assessment data directly correlates with the capability of the evaluators. Dowd places great emphasis on the question: "Who should decide?" The right evaluators are



Former Navy SEAL and founder of SOFware LLC, John Dowd

essential to the process, and choosing the right evaluators is the crucial first step to building a good A&S process.

Dowd discusses this through the lens of three different evaluator archetypes: Gatekeepers, Santas and Senseis.

**Gatekeepers:** These are individuals who maintain hostile attitudes toward candidates, determined to ensure that no one passes unless they exceed the evaluator's own perceived performance. As Dowd characterizes them, these are "bitter veterans who believe no one can earn what they did. They set the bar impossibly high — not to find the best, but to protect their ego." Although gatekeepers will prevent the erosion of standards, they often will do so at the expense of the organization and the candidates. Gatekeepers will tend to under-select candidates and keep out individuals who may have proven very valuable to the team.

**Santas:** These are individuals who struggle to enforce standards, giving candidates unwarranted benefit of the doubt. Santas are too generous, too lenient and unwilling to say no. Santas struggle with candor, enforcing cultural norms, and will ultimately degrade operational capabilities.

**Senseis:** These are individuals who seek to maintain rigorous standards while genuinely hoping that candidates succeed. They hold the line, yet they encourage growth, provide good feedback, are good teachers and maintain high standards without arrogance or ego. Senseis are the gold standard, and all A&S programs should prioritize placing senseis in leadership positions. Any evaluation with subjectivity in it is best left to the senseis, while limiting gatekeepers and Santas to objective measurement roles where their internal biases are less damaging.

This approach ensures that assessment serves the organizational needs rather than the evaluators' egos. "You want senseis more than gatekeepers or Santas," Dowd emphasizes. "Ideally you'd have all senseis, but it's probably not going to work out that way. But at a minimum, know who your senseis are, your gatekeepers, your Santas, and place them in the right roles."

## Building your A&S — defining the criteria

The next step to an effective A&S is precisely defining the role you're selecting for and the behaviors it requires. Although defining the role may seem straightforward (e.g., "new SWAT operator"), it is essential to avoid the temptation to select for behaviors only required by more senior roles. In other words, don't select for team leaders when you should be selecting for new operators. Dowd emphasizes that attributes have no role at this stage, only behaviors: "Fifteen years ago we were big advocates of starting with attributes, but time and again saw how their abstract nature derailed the process and left each assessor with a different picture of what was being selected for. Eliminate those issues by focusing on the concrete, observable behaviors that matter for the target role."

To accomplish this, Dowd advises: "Go find the high-reputation gray beards in your community. The guys who have been there, done that, and ask them, 'For this role, what are the concrete observable behaviors that you've witnessed with your own eyes that made a difference in a workup cycle, on the street, or serving a high-risk warrant late at night?'"

Then use these behaviors to define selection criteria and test events that will distinguish those who have the desired behaviors from those who do not. The commander's role should be to review this list to ensure alignment with the organization's aspirational culture, potentially adding or modifying behaviors to account for future requirements.

## Categorizing ideal behaviors and testing for them

Once identified, these ideal behaviors should be organized into three categories: individual performance, team impact and community devotion. Each will require an appropriate and different evaluation methodology.

### 1. Individual performance — operational capabilities

Individual performance encompasses the skills and capabilities needed to perform the job: physical fitness, technical proficiency, grit and specialized knowledge are all examples of individual performance items. These behaviors typically have no "best self" vs "true self" gap — you can't fake your way to more pushups or a better marksmanship score. To evaluate individual performance, engineers test events that reliably, candidate after candidate, produce an insight into the presence or absence of the given behavior. In many cases, this will be easy. If accurate marksmanship during physical exertion is an important behavior, engineer a test that starts with an obstacle course and concludes with a shooting drill. Other behaviors require more creative or involved engineered tests. As Dowd notes, not-quitting-when-things-suck "is not a build, that is a buy. It's some intersection of your desire to be part of this community, your desire to be worthy of the guy next to you, and just your ability to withstand suck." This can only be reliably tested via a "crucible event" — a prolonged period of intense discomfort that can be easily ended by two words: "I quit."

Dowd notes that individual performance is the easiest category to measure, is frequently over-measured, and matters less than team impact and community devotion.

### 2. Team impact — protecting the culture

How an individual affects group dynamics and collective performance is an essential component for a good teammate, irrespective of individual performance. "I don't care if you can run a 4-minute mile and shoot the lights out from a mile away," Dowd emphasizes. "If we add you to that team and you degrade it in some way, your negative team impact outweighs your positive individual performance."

Team impact behaviors are notoriously difficult to measure. First, there is no truly objective way to measure team impact in an engineered event like there is for performance. We cannot simply tell someone to "do as many nice things as you can in two minutes." Moreover, evaluators are rarely in the best position to evaluate team impact. Everyone going through selection will be on their best behavior around the team cadre. As a result, team impact behaviors have a large "best self" vs. "true self" gap — candidates will present an idealized version of themselves when directly observed by assessors.

> **How an individual affects group dynamics and collective performance is an essential component for a good teammate, irrespective of individual performance.**

Instead, organizations must place candidates into prolonged, intense contact with each other, then harvest the desired insights through surveys completed by their peers afterwards. Dowd's teams embed structured, anonymous peer reviews throughout their pipelines. These aren't popularity contests — they are carefully designed to measure the team impact behaviors that matter. Who helped under stress? Who avoided responsibility? Who maintained a positive influence even when things went sideways? These reviews become a living snapshot of the team dynamic, and they often reveal things the cadre misses.

### 3. Community devotion — protecting the organization

Community devotion is the final area that must be examined for effective A&S processes. Community devotion describes an individual's commitment to organizational values and willingness to prioritize collective success over personal gain. As Dowd puts it, "We need folks, especially as they ascend in rank, who will always sacrifice their own interests for those of the community, who will make the unpopular decision with their guys because it's the right thing for the community."

While these behaviors are critically important, this dimension is exceptionally difficult to measure during initial A&S timeframes. Although the worst offenders may be identifiable, beyond these obvious cases, organizations must harvest insights via surveys of peers and assessors and consult with psychological professionals over time to truly measure community devotion.

Dowd advises maintaining humility about assessments in this domain, recognizing the limited visibility into candidates' true motivations during initial selection timeframes. Getting a deep reading may not be possible until after initial selection. Yet this is a domain that is ignored at the team's peril. This is why it may be an area more appropriate for the training cycles that follow initial selection; for example, during basic training and while new members are on initial team probation. Perhaps no area better exemplifies why A&S must be seen as an ongoing process rather than a single discrete event. You may not be able to assess community devotion appropriately in a short initial selection process, but over several months and numerous peer assessments, true devotion to the community becomes very apparent.

### The buy vs. build framework

To evaluate performance, Dowd advocates the use of a framework that's especially valuable: "Buy vs. Build." The idea behind this is that there are some traits and capacities (e.g., marksmanship, movement, fitness, etc.) that can be "built" after selection through training and mentorship, while others (grit, honesty, or reliability) simply must be present at entry (i.e. "bought").

For each behavior desired, organizations must determine whether that trait can be built and, if so, what their "buy vs. build threshold" is. This is a function of whether the desired behavior is buildable and the time and resources are available to do so. If an item can be developed to the required levels in the time available, then it can be "built." For example, "If they show up running 6:30 miles, we can certainly 'buy' at 6:30 and 'build' to 6:00." But, if we cannot build to the desired levels in the available time, it simply must be chosen for in the selection process ("bought") at the required level. To continue the example, if somebody shows up running 8-minute miles, there is no way we can build that into 6-minute miles in training.

For each behavior, it is essential to ask three questions:

1. What's the minimum required proficiency to be useful on the team?
2. Can the behavior in question be developed at all?
3. If so, given the time and resources we have, how much can we build, and thus what's the minimum level of proficiency they must display at selection?

A common temptation is to treat every key behavior as a "buy," but mature organizations realize many can and must be built.

### Building your assessments — evaluation methods

Of course, to assess each of the three domains discussed above, it is crucial to create the right types of tests for the behaviors being measured. These structured assessments, occurring at specific times, must be designed to measure capabilities in each domain. Ideally, at least some measurements should come after physical stressors like fatigue and discomfort as well as following psychological stressors like failure, disappointment or uncertainty. After all, these are all factors likely to be present in the operational environment.

Engineered assessment events are most appropriate for the Individual Performance domain and come in two varieties: quantitative and qualitative.

---

**A common temptation is to treat every key behavior as a "buy," but mature organizations realize many can and must be built.**

**Quantitative assessments:** These measure behaviors that can be objectively counted or timed. For example, push-ups, runs, marksmanship, etc. These assessments have clear success and failure criteria as well as a means of scoring candidates against one another. Whether you are using the assessment to measure a "buy threshold" (can they do 20 pull-ups) or to differentiate the candidates from one another (shooting scores from 250-300), quantitative assessments are the easiest to design and implement. This makes them easier to conduct for any assessor, including Santas and gatekeepers.

**Qualitative assessments:** These evaluate subjective behaviors that require personal judgment from the evaluator. As a result, they are much harder to design and implement and more prone to failure. For these assessments, A&S leadership must "own the yardstick" or clearly define what the "right behavior looks like" before measurement. Qualitative evaluations must be made concrete and observable. For example, "He remained composed in the face of hostile audience questions" instead of abstract ideas like "compartmentalization" or "resilience." This will ensure that individual assessors do not get to idiosyncratically define what "right" looks like.

### Building effective qualitative assessments

To be effective, qualitative assessments must incorporate four key design principles: They must use multiple assessors, they must be designed to create differentiation between candidates, they need to induce then observe failure, and they must use the right type of questions for the behaviors being measured.

**1. Use multiple assessors:** For qualitative events, multiple people should evaluate the same candidate simultaneously. It is critical to capture the perspectives of several different individuals who are assessing the same event to ensure the evaluation is a reliable measure of the candidate and not the evaluators' biases. By utilizing multiple evaluators, we can ensure that we have inter-rater reliability (IRR). For effective tests, IRR will be very high. However, if IRR is poor, organizations must determine what went wrong in the evaluation and then ensure they implement a better-designed evaluation event.

**2. Create differentiation:** Qualitative assessment events must be designed to reliably distinguish between candidates. "If you have an assessment and everybody gets the same grade," Dowd emphasizes, "you've assessed absolutely nothing. That is a completely useless test." Well-designed assessments will always create a spread between how candidates are scored. As a result, the assessment must be intentionally

designed not to drive a standardized and middle-of-the-road answer from evaluators.

**3. Induce and observe failure:** Well-designed selection processes intentionally place candidates in situations where they will experience failure, then measure their response. Great programs intentionally create failure events: communication breakdowns, moral dilemmas, time traps. And then they watch what happens next. Who takes ownership? Who blames others? Who keeps working without recognition? These are the behaviors that predict how someone will perform under pressure, not just on a range, but in real-world operational chaos. "Failure is the mask remover," Dowd notes. "It's the moment you meet the real person."

**4. Use the right types of questions:** Survey questions can utilize two different types of questions. The first, Behaviorally Anchored Statements, focus on specific observable actions (e.g., "They show up on time, ready to train"). These are always safe to use but are narrow in scope. The second, Outcome-Oriented Statements, assess comprehensive impressions (e.g., "I would actively recruit them for my next team"). These questions provide much more value but can only be used with trusted cultural agents like experienced team members. Outcome-oriented statements can be particularly effective for detecting risk.

### Conclusion

Effective assessment and selection largely is defined by how well a unit can define its desired behaviors and then build appropriate assessment events around individual performance, team impact and community devotion. To accomplish this, however, it is essential that the correct evaluators are employed to do the assessments, and that each assessment is designed utilizing the right measurements.

In part two of this series, we will examine common errors and pitfalls that undermine the effectiveness of A&S processes.

**Effective assessment and selection largely is defined by how well a unit can define its desired behaviors and then build appropriate assessment events around individual performance, team impact and community devotion.**