LESSONS LEARNED FROM
THE DEBRIEF

BY JON BECKER

> If you allow a sadistic environment during selection, you're teaching candidates something toxic about what the organization values.

COMMON PITFALLS IN ASSESSMENT & SELECTION: LESSONS LEARNED FROM ELITE UNIT A&S

PART TWO

This is the second of our two-part series on Elite Unit Assessment & Selection (A&S) with John Dowd, CEO and founder of SOFware LLC. Dowd is a former Navy SEAL, and for the past 15 years, he and the SOFware team have been helping our most elite units improve their A&S and culture-shaping processes.

In Part 1, we took a high-level look at A&S and some of the key concepts and frameworks that Dowd and his team have developed over the years. In this installment, we will explore the most common errors and pitfalls Dowd has seen, as well as mitigation strategies to prevent these issues. We will be looking at these pitfalls from three different perspectives:

Programmatic pitfalls: Errors and flaws in how the overall A&S process is governed, who runs it, how success is defined and how the process evolves over time.

Assessment pitfalls: Mistakes in how events are designed and conducted, including unclear standards, biased evaluations and poorly built scenarios.

Survey pitfalls: Failures in how peer and cadre feedback is gathered, interpreted or applied — often the most misunderstood part of the process.

Programmatic pitfalls

By far the biggest area where selection pitfalls occur is at the leadership and governing of selection processes. These errors can cause widespread problems throughout the process and compromise A&S event results.

Failing to recognize your cadre archetypes: As we discussed in part 1, Dowd and his team have identified three main archetypes among A&S staff: gatekeepers, Santas and senseis. As Dowd puts it, "You want senseis more than gatekeepers or Santas. Ideally, you'd have all senseis, but it's probably not going to work out that way. At a minimum, know who your senseis are, your gatekeepers, your Santas, and place them in the right roles."

Mitigation: Organizational leadership should make it clear that the A&S process will have a sensei culture and should ensure all key A&S leaders have the sensei disposition. Gatekeepers and Santas should receive direct feedback and mentorship so that the A&S culture is continually becoming more sensei-oriented.

Creating a sadism sandbox: Every selection process involves hardship — by design. Tactical units need people who can endure real discomfort and stress. But there's a thin line between testing resilience and simply creating a sadism sandbox where candidates are tortured for no reason. This line is dangerously easy to cross because A&S cadre hold absolute authority, candidates are desperate to succeed and toughness is genuinely a requirement for the job.
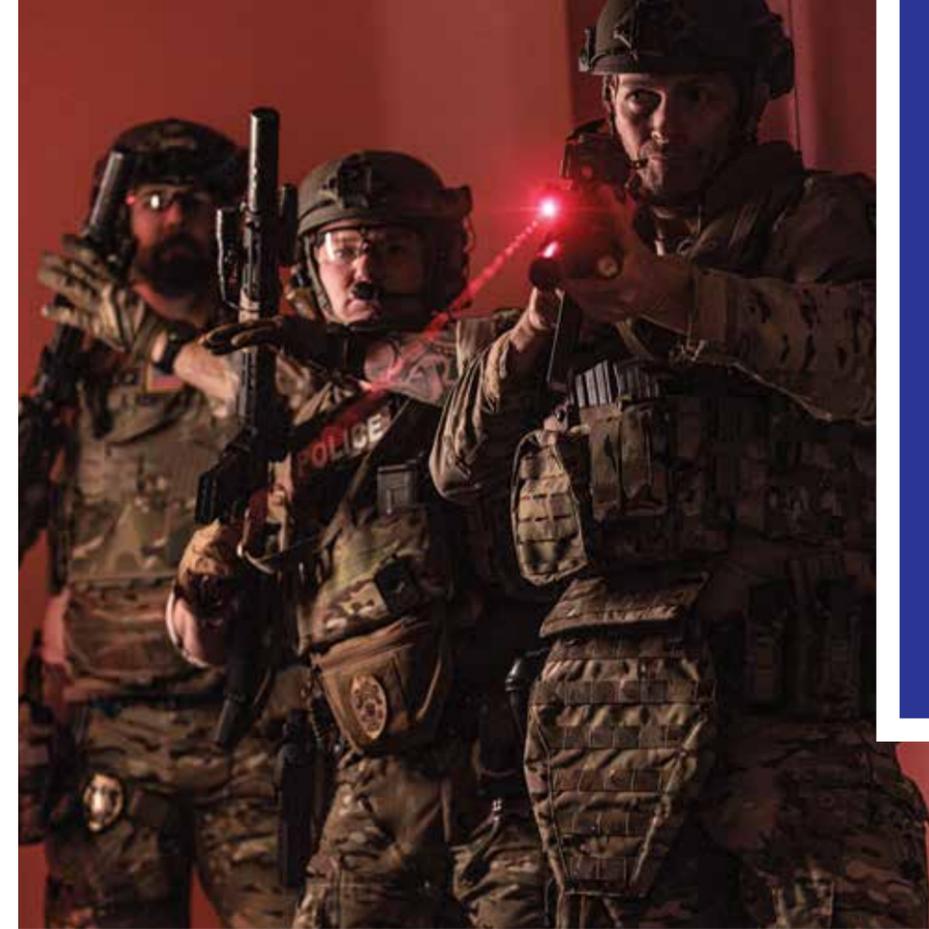
An unmonitored, gatekeeper-heavy culture can quickly devolve into pointless cruelty rather than purposeful hardship. That's a mistake — not only because it risks avoidable injuries and the occasional training death, but because your A&S is each candidate's first exposure to your culture. Dowd points out that if you allow a sadistic environment during selection, you're teaching candidates something toxic about what the organization values.

Dowd warns, "Every fraternity hazing incident or in the Stanford Prison Experiment there's a group of people in charge, a group of people who have to be at their beck and call, and those conditions left unchecked can veer quickly into sadism." These environments not only eliminate people who might have been very good, but the graduates of such programs quickly learn "as their very first lesson about your community that sadism is cool here. Sooner or later, you will regret that."

Mitigation: Focus on a sensei-led process with a relentless drive toward purpose-driven assessments. Every challenging event must be justified by clear links to specific, desired behaviors. Ask constantly: Does this event effectively measure behaviors we value? Or is it suffering for suffering's sake?

Treating everything as a "buy" behavior: Dowd notes that many A&S programs treat every desirable behavior as a pure "buy," meaning candidates must show up fully developed on that dimension. In reality, every behavioral dimension in selection involves establishing a buy vs. build threshold, determined by two key factors: how critical the behavior is, and how much your organization can realistically develop that behavior post-selection.

For skills like shooting scores or running times, your program likely has significant capacity to improve candidates through training after selection. Therefore, your "buy threshold" is the desired performance you need minus what your process can reliably build.

But some behaviors (e.g. integrity, reliability or a selfless mindset) have little or no margin for building after the fact. For these traits, your buy threshold is your final standard. Dowd illustrates this with a concrete example: "If they show up running 6:30 miles, we can certainly 'buy' at 6:30 and 'build' to 6:00. But if somebody shows up running 8-minute miles, there's no way we can build that into 6-minute miles in training."

Mitigation: Clearly identify the precise buy vs. build threshold for each desired behavior. Determine what your team genuinely can (or cannot) develop after selection, and establish initial entry standards accordingly. This ensures you maintain necessary rigor where it matters most (character and cultural attributes), while maximizing your talent pool by allowing trainable skills space for realistic improvement post-selection.

Balancing false positives vs. false negatives: Every selection process faces two competing risks: false positives (i.e., selecting someone who isn't truly suitable) and false negatives (i.e., rejecting someone who would have been outstand-

ing). Many units instinctively fear false positives because the immediate damage to team culture or performance feels severe. But it's critical to remember that A&S isn't a one-time event; it's part of a larger, ongoing developmental pipeline for team personnel. If your pipeline allows multiple opportunities to screen candidates, the correct approach shifts dramatically depending on where you are in that pipeline.

Early in the process, your goal should be to avoid false negatives, which means prematurely rejecting candidates who may just need more observation or training. Later in the pipeline, as you get closer to actual operational assignments, the priority flips and preventing false positives must take precedence.

Mitigation: Calibrate your risk tolerance along the length of your pipeline. Early events should cast a wider net, minimizing false negatives by giving borderline candidates a chance to grow. Later events become progressively stricter, protecting the team from the cultural or operational damage of false positives.

Narrow selection windows: The pipeline for adding new personnel typically has three distinct phases: administrative steps (applications, time-on-force requirements, etc.), a dedicated assessment & selection event, and initial operator training. Dowd points out that many units only permit screening candidates out during the first two phases. He argues that this is a mistake as some disqualifying behaviors can only be detected over time periods longer than the duration of typical A&S programs.

If you remove the ability to deselect during initial operator training, you're locked into candidates who become cultural liabilities or operational risks. Your A&S staff will fear this outcome and will prematurely focus on preventing false positives during the A&S in an attempt to protect the culture even though they won't have sufficient data at that point.

Mitigation: Maintain the flexibility to deselect candidates throughout the entire pipeline. Clearly communicate from the outset that initial operator training remains a continuous evaluation phase. Protect your culture by empowering the cadre to act on the fuller picture that emerges only after extended observation.

Lowering the wrong standards: Sometimes, operational demands mean you need more candidates to graduate your pipeline successfully. In Dowd's experience, people reflexively scream, "We can't lower standards!" But, graduating more candidates by definition means lowering standards, and this reflex makes it impossible to lower the right standards thoughtfully.

The consequences of accepting even one narcissistic, self-serving teammate far outweigh the minimal impact of accepting a slightly slower runner or less technically polished candidate. As Dowd puts it, "You can teach tactics. You can build fitness. You can't fix a guy who poisons the team room."

Mitigation: When it is necessary to expand selection, consciously relax standards around behaviors you can realistically develop later, like technical proficiency, specific tactical capabilities or physical fitness benchmarks. However, fiercely defend and protect the standards related to character, teamwork and cultural impact.

Mishandling non-selects: Every selection program produces more non-selects than graduates. As a result, your process will disappoint far more potential team members than it will make happy. Unfortunately, many units handle these non-selects poorly and often dismiss them with little or no explanation. This misses a critical opportunity.

Non-selects hold valuable insights into your process. Conducting structured exit interviews allows cadre to provide non-selects meaningful, actionable feedback to guide their future improvement. Just as importantly, these interviews offer cadre valuable feedback on their own process to see what's working, what's confusing, or what may unintentionally push good candidates away.

Mitigation: Treat exit interviews as a required step, not an afterthought. Ensure non-selects leave understanding exactly why they weren't chosen, and what they can do differently in a future selection. Equally critical, openly solicit their feedback on your selection process. This creates goodwill, strengthens the credibility of your program, yields continual insights for improvement and offers an important guard against the threat of a sadism sandbox.

Assessment pitfalls

The term "assessment" in this context refers to an engineered event designed to reliably produce an opportunity to measure the presence or absence of a given behavior. Assessments can be either quantitative (e.g., counting pushups or timing runs) or qualitative (e.g., arrest scenarios or public speaking), and it is essential that the design of assessment methodology varies depending upon the type of selection desired.

Using gatekeepers/Santas for qualitative assessments: While gatekeepers and Santas can be trusted to count push-ups or time runs, they should never be assigned to conduct qualitative assessments. Their inherent biases (e.g., being impossibly demanding or inappropriately lenient) contaminate subjective evaluations and produce unreliable data. "Gatekeepers and Santas are fine for timing runs and counting push-ups," Dowd explains. "But any evaluation with subjectivity in it is best left to the senseis." The problem becomes particularly acute in leadership roles, where these archetypes can shape the entire assessment culture of the unit in their image.

Mitigation: Restrict gatekeepers and Santas to objective measurement roles where their internal biases are less damaging. Reserve all qualitative assessments for senseis and ensure that senseis occupy key leadership positions in the assessment process.

Non-selects hold valuable insights into your process. Conducting structured exit interviews allows cadre to provide non-selects meaningful, actionable feedback to guide their future improvement.

Time-slicing during evaluation: One of the most insidious problems in qualitative assessment is time-slicing, which is produced by having evaluators assess multiple candidates simultaneously by rotating their attention between them. The reason this is dangerous is that small glimpses at several candidates can create a scenario where several competent evaluators arrive at completely opposite conclusions about the same candidate based on perceiving different behaviors.

Dowd describes this common scenario: "Time-slicing is, we're all going to grade all of them at the same time, which means I'm watching Billy from zero to 30 seconds, and then I slice over to somebody else. So, I see Billy do something awesome and then I move over to Johnny, and then you look at Billy and you see him do something dumb. I give him a five, you give him a one."

Mitigation: During qualitative assessments, each evaluator must observe the entire performance of each candidate they're tasked to grade. This may require more evaluators or longer assessment windows, but the alternative is unreliable data masquerading as valid assessment, which undermines the entire process.

Failing to induce failure, uncertainty and disappointment: Many assessment programs avoid deliberately creating scenarios where candidates will experience failure, disappointment or uncertainty. This represents a fundamental misunderstanding of what tactical environments demand.

"Great programs intentionally create failure events: communication breakdowns, moral dilemmas, time traps," Dowd explains. "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."

The key insight is that failure is the mask remover. It's the moment you meet the real person. How candidates respond to setbacks reveals crucial information about their resilience and adaptability that cannot be assessed through successful performance alone.

Mitigation: Deliberately engineer failure into assessment events while maintaining their authenticity. Candidates must believe they could have succeeded with better performance. Measure not the failure itself, but the candidate's response to failure in subsequent, unrelated tasks.

Measuring qualitative assessments via abstract attributes: A fundamental error in qualitative assessment design is using abstract attributes like "compartmentalization," "resilience" or "leadership" as evaluation criteria. These concepts mean different things to different evaluators, creating a situation where assessment results reflect evaluator interpretation rather than candidate performance.

Mitigation: Use concrete, observable behaviors instead of abstract attributes. E.g., replace "compartmentalization" with "remained composed in the face of hostile audience questions." Replace "leadership" with "took initiative when group lacked direction." This ensures evaluators are grading the same observable phenomena rather than their personal interpretations of abstract concepts.

Poor Inter-Rater Reliability (IRR): When multiple evaluators assess the same candidate and arrive at dramatically different conclusions, it indicates a fundamental problem with the assessment design. Poor IRR suggests that evaluators are either measuring different things, applying different standards or observing different performance segments. "The degree to which they differ suggests that everyone was on the same yardstick or not," Dowd explains. "If everybody grades John between three and four, that's a very different situation than if they grade him between one and 10."

Mitigation: Design assessment events to maximize IRR by using concrete, observable criteria and ensuring all evaluators observe the complete performance. Monitor IRR as a quality metric and when it's poor, investigate whether the problem lies with the event design, evaluator training or assessment criteria. Make sure qualitative assessments have multiple graders per candidate, so that IRR can be measured.

Over-measuring the easy-to-measure: Quantitative assessments are easier and cleaner than qualitative assessments, so many units over-measure things that can be counted or timed. The most valuable data, however, is nearly always qualitative.

"We tend to over-measure job performance relative to the other stuff," Dowd observes. The problem is that team impact behaviors (i.e. the factors that determine whether someone enhances or degrades group performance) require more sophisticated assessment techniques that many organizations avoid implementing.

Mitigation: Deliberately balance your quantitative and qualitative assessments. Recognize that the most important behaviors for team success are often the hardest to measure, and invest the additional effort required to assess them properly.

Failing to assess your assessments: You have limited time and resources to conduct your A&S, so every event must yield insight. Yet many programs fail to systematically evaluate whether their assessment events do so. Dowd emphasizes two key metrics: "Do they generate good inter-rater reliability (IRR) and do they produce meaningful differentiation?" Events that fail these tests are worse than useless; they consume valuable resources while providing no actionable information. "If you have an assessment and everybody gets a three, you've assessed absolutely nothing. That is a completely useless test," Dowd notes.

> **How candidates respond to setbacks reveals crucial information about their resilience and adaptability that cannot be assessed through successful performance alone.**

Mitigation: Systematically evaluate each assessment event after completion. Events with poor IRR or little differentiation should be re-engineered or eliminated. Invest assessment time only in events that reliably produce useful data for selection decisions.

Survey pitfalls

Dowd's team uses the term "survey" to refer to harvesting insight from the candidates themselves. Rather than a discrete engineered assessment, surveys capture insights gathered over weeks or months of prolonged interaction. Several pitfalls occur around surveys, most of which result from poor design.

Asking about abstract attributes: Just as with assessments, asking peers to evaluate each other on abstract concepts like "integrity" or "work ethic" produces unreliable data. Decades of psychometric testing have established that humans can't hold a stable definition of abstract attributes in their heads, leading to inconsistent and meaningless responses.

Mitigation: Use concrete, observable behaviors in survey questions. Instead of asking about "integrity," ask, "He tells the truth even when it disadvantages him." Instead of "work ethic," ask, "He completes his share of unpleasant tasks without being reminded."

Using outcome-oriented statements when you don't have trusted cultural agents: There are two types of survey questions: behaviorally anchored statements and outcome-oriented statements. Behaviorally anchored questions focus on specific observable actions, while outcome-oriented statements assess comprehensive impressions.

The critical error is using outcome-oriented questions with populations that lack the cultural knowledge to answer them meaningfully. As Dowd explains, "The 17-year-old who was in high school last week should not be asked who he would actively recruit for his next platoon. He knows nothing about your culture."

Outcome-oriented statements like "I would actively recruit him for my next team" can only be used with trusted cultural agents — experienced team members who understand and have bought into the organization's culture.

Mitigation: Stick with behaviorally anchored questions unless you have confidence that respondents understand and have bought into your culture. Reserve outcome-oriented statements for experienced team members who can make meaningful cultural assessments.

Insufficient time and challenges together, away from cadre eyes: Before surveying candidates about their peers, they need plenty of opportunity to interact with one another with no cadre around. The goal is for candidates to drop their facade and reveal their true selves to one another.

"Everyone going through selection will be nice to the evaluators and be on their best behavior around team cadre," Dowd notes. "You've got to set some conditions, get out of the room, and then come back to their peers and harvest the information."

This requires more than just housing candidates together, it requires creating challenging situations where their ability to maintain their "best-self" facade erodes under stress.

Mitigation: Design extended periods where candidates work together on challenging tasks without cadre supervision. Create conditions that naturally erode their ability to maintain facades: physical fatigue, frustrating assignments, limited resources and time pressure. Only then survey them about their peers' actual behaviors and impact.

Conclusion: Learning from the mistakes of others

The pitfalls outlined in this article represent hard-won lessons from some of the world's most sophisticated selection programs. They're not theoretical problems — they're real failures that John Dowd and his team have observed repeatedly across elite military and law enforcement units.

The good news is that these mistakes are entirely preventable through thoughtful design and disciplined implementation. The frameworks and mitigation strategies outlined here provide a roadmap for avoiding the most common traps that undermine A&S effectiveness.

Perhaps most importantly, these pitfalls highlight why selection cannot be treated as a simple screening process. Done correctly, A&S becomes a strategic capability that defines, reinforces and perpetuates organizational culture. Done poorly, A&S can actively undermine the very culture it's meant to protect.

As Dowd reminds us, "Your whole career is a selection event." The principles governing formal selection programs should extend throughout professional development, creating organizations that continuously evaluate, develop and reward behaviors aligned with their aspirational culture.

The stakes are too high to accept preventable failures in personnel selection. Whether building a special operations unit or a corporate team, the science of selection offers proven methods for organizational excellence, but only when implemented with the rigor and sophistication these elite units have learned to demand.

Done correctly, A&S becomes a strategic capability that defines, reinforces and perpetuates organizational culture. Done poorly, A&S can actively undermine the very culture it's meant to protect.