Using Behavioral Measures to Assess Counter-Terrorism Training in the Field

V. Alan Spiker[1] and Joan H. Johnston[2]

[1] Anacapa Sciences, Inc.
Santa Barbara, California
[2] Naval Air Warfare Center Training Systems Division
Orlando, Florida

Abstract. Development of behavioral pattern recognition and analysis skills is an essential element of counter-terrorism training, particularly in the field. Three classes of behavioral measures were collected in an assessment of skill acquisition during a US Joint Forces Command (JFCOM)-sponsored course consisting of combat tracking and combat profiling segments. These included situational judgment tests, structured behavioral observation checklists, and qualitative assessments of the emergence of specific knowledge-skills-attitudes over the course of training. Evidence was present in all three types of measures to indicate that behavioral pattern recognition and analysis skills were successfully acquired by most students (a mix of Army and civilian law enforcement personnel). The paper describes both the types of skills acquired and the statistical evidence that supports their acquisition over the course of field training. Implications for broader training of these critical skills are also discussed.

Keywords: Situational judgment tests, behavioral observations, scenarios, knowledge-skills-attitudes, profiling, tracking.

1 Introduction

In 2008, the US Department of Defense placed Irregular Warfare (IW) on an equal footing with conventional warfare in future military planning and operations [1]. Among IW mission objectives are developing capabilities to address asymmetrical threats and the challenges they pose for counter-terrorism (CT). Whether practiced by military ground units or law enforcement personnel (e.g., Custom and Border Patrol), reading the human terrain – such as through behavioral pattern recognition and analysis (BPRA) skills – is an essential element of CT training. BPRA is a set of skills and techniques that a profiler uses (e.g., cues and indicators of behaviors) to spot people and events *before* the situation becomes lethal. Staying 'left of bang' by constructing these behavior profiles in a proactive fashion is now considered to be a protective element for small units, and is every bit as important as body armor and weaponry [2]. As a result, behavior profiling techniques have become a valuable addition to small unit tactics, techniques, and procedures (TTPs).

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2011, LNAI 6781, pp. 461–470, 2011.

Nowhere are these skills more highlighted than in Combat Hunter (CH) training conducted by the United States Marine Corps School of Infantry. A 10-day course, CH is taught in two segments, combat profiling and combat tracking, by subject matter experts (SMEs) in these respective fields. Each segment has a classroom academic portion and scenario-based field exercises where the skills are developed, applied, refined, and reinforced. Historically, it is hard to measure training development in the field since the observation conditions are difficult, curriculum is not always standardized, objectives are not always well-specified, and instructional delivery by instructors is inconsistent.

To document training acquisition, a wide array of behavioral methods and measures are needed. In this paper, we summarize the results of a field study where we observed development and acquisition of BPRA skills by military and law personnel as part of formal counter-terrorism training in a special offering of CH called Border Hunter. Specifically, we describe empirical results using three different methods of behavior measurement: situational judgment tests (SJTs), behavior observation checklists (BOCs), and knowledge-skill-attitude (KSA) profiles. This discussion addresses both the content of the skills measured and evidence for their acquisition. We conclude with a discussion of the implications for using behavioral measures to support CT training within a broader spectrum of training technologies.

2 Training Counter-Terrorism Skills

2.1 Combat Hunter

Instruments to collect behavioral measures were developed during the authors' naturalistic observation of CH training at the School of Infantry – West, Camp Pendleton. CH is a ten-day course split equally between Combat Tracking and Combat Profiling. Each class consists of approximately 40 students drawn from the same Marine regiment, though students typically come from different platoons and squads. Both segments are split into academic instruction and field scenario portions. In Combat Tracking, students receive academic instruction in the morning on the fundamentals of tracking (e.g., dynamics of footprints, maintaining track line, interpreting spoor), where the afternoons are devoted to utilizing this knowledge in the field where they track 'quarry' (role-playing instructors) as five-person tracking teams. During these tracking scenarios, students learn to read their enemies' spoor (i.e., footprints, human signs, environmental cues, slight ground disturbances). They are also taught to build social/biometric profiles of their quarry, anticipate their targets' actions by acquiring the mindset of the quarry, and apply TTPs to hunt down their targets. Combat tracking is a human-centric competency particularly useful in IW settings to support offensive operations, intelligence gathering, clandestine movement in hostile areas, and counterinsurgency operations. Over days, the field scenarios increase in complexity as the instructors add in such factors as more difficult terrain, more 'skilled' quarry (e.g., where the role players purposely try to cover their tracks), and more intricate team tracking maneuvers.

Combat Profiling is structured differently, where 3 days of academic instruction precede the field scenarios. Profiling is concerned with perceiving, analyzing, and

articulating critical events within the human terrain. Its main goal is to identify pre-event indicators through human behavior 'left of bang', (i.e., before a destructive event occurs). It trains individuals to look for behaviors that are anomalous, beyond the baseline of a culture or location. Through combat profiling, warfighters and law enforcers learn to be more situationally aware and to accurately interpret subtle cues that forewarn a critical event. During classroom instruction, students are exposed to the basic concepts of profiling, such as fundamentals of optics, pattern recognition, reasoning by analogies, forming prototypes, ethical-moral decision-making, and the six domains of combat profiling (i.e., heuristics, geographics, proxemics, atmospherics, biometrics, kinesics). The practical application portion is conducted in the next two days where students are split into teams and man observation posts (OPs) and observe role-players engaged in varying types of behavior within a village mockup. They practice their profiling skills by observing, at a distance, instances of neutral and insurgent behavior in the context of increasingly challenging scenarios. Typically, 5 to 6 such scenarios are executed during these two days. The training culminates in a 4-hour long final exercise where all teams deploy as maneuver units into the village using their insights gained from the previous scenarios.

The authors observed two evolutions of the CH course; these observations were used to develop draft versions of the three types of behavioral measure instruments [3]. Besides observing students and instructors, the authors interviewed selected students and instructors for further information, to clarify points, and to extract the higher level skills that were being trained. Additional materials were obtained from the instructors that provided further information concerning the theoretical and practical underpinnings of combat profiling and tracking.

This repository of information was then used to create separate versions of an SJT for Combat Tracking and Combat Profiling, as well as a structured BOC for each segment. The SJTs were six-item tests that required participants to think about and decide among six possible response options for each briefly-described scenario. They follow the format typically recommended for industrial applications [4]. The BOCs were structured so that rater/observers would provide 3- or 5-point ratings on a set of basic and advanced profiling or tracking skills. In addition, a taxonomy of KSAs was created that would be applicable to both course segments. The taxonomy would be used retrospectively by observers, after the training day's conclusions, as a way to categorize the free-form observations they made while observing students. Further details on each instrument will be discussed when the empirical results are presented.

2.2 Border Hunter

Border Hunter (BH) can be characterized as a one-off, 'graduate-level' version of CH [4] that also consists of combat tracking and profiling instruction. Twenty-one days long, BH was sponsored by USJFCOM and was conducted by Joint Task Force North at Fort Bliss, TX in April 2010. The 10-day Combat Tracking segment was taught by six highly experienced trackers with a combined experience of 180 years. Similarly, Combat Profiling was taught by an eight-person team of instructors with a collective experience in military and police work of more than 200 years.

Forty-three trainees received the BH training, comprising a mix of US Army, Border Patrol, and other law enforcement personnel. All were highly experienced

with an average of 9 years in military/law enforcement. In addition, 22 soldiers from Fort Bliss were recruited as role players during the Combat Profiling field scenarios.

A 13-person research team was present to capture course content for subsequent packaging, as well as to conduct behavioral research and assess training effectiveness. This included both the academic instruction and field training aspects of the course. The focus of the present paper is with the latter, where on a given day, 3 to 6 individuals were available to collect behavioral data on student training performance. All were highly experienced researchers with advanced degrees.

Students were assigned to the same teams for all Tracking and Profiling field training scenarios (FTXs). Teams 1 and 2 were composed of students from the Custom and Border patrol agencies. Teams 3 and 4 were configured with Army personnel; students in Team 3 were less experienced than their Team 4 counterparts. Team 5 was a hybrid team, comprising a mix of Army and law enforcement personnel.

3 Situational Judgment Tests

3.1 Method

SJTs are low- to moderate-fidelity work sample simulations that assess preferences for appropriate behaviors in a work setting [5]. While SJTs have long been used in industrial settings for job selection and placement, their use as a source of proficiency data from field settings is less frequent. Because SJTs have moderate concurrent validity with performance [6], we elected to use them in this project as a way to assess degree of learning in the FTXs for both Tracking and Profiling.

SJT items ask respondents to assess the effectiveness of various response options. The scenarios are intentionally written so that not all situational cues are known, which increases dependence on one's judgment. This dependence involves a balance between analysis and intuition, where good judgment is the ability to go beyond the information given and rely on broader knowledge and experience [7]. If students have been receiving this experience, by virtue of participating in FTXs, then they should exhibit improved performance on the SJTs between pre-test and post-test.

The SJT instrument is a particularly useful tool because its realistic scenario items reasonably approximate the types of cognitive process improvements expected from repeated FTXs. By using the SME instructor ratings as the 'answer key', we assessed how the students' mental representations of various real-world problems began to resemble the instructors' mental representations.

For the Combat Tracking SJTs, our review identified six skill areas that were particularly important for success and amenable to testing via a short, written scenario: methods for closing the time/distance gap, executing lost spoor procedures, counter-tracking tactics, tactical formations, ground spoor characteristics, and dynamics of the footprint. For each area, a one-paragraph scenario was prepared to set up the problem. Then six alternative course of action options were generated that might be initiated by the tracker in response to the problem. In developing these courses of action, two options were specified that would be superior, two that would be inferior, and two that would have both strengths and weaknesses.

An example item from one of the Tracking SJTs is shown below. The scenario set-up is brief, reducing the reading requirement and leaving desirable information omitted. The student must make inferences and exercise judgment which is aided by the experience obtained during the FTX – the main intent of the SJT.

----------Example SJT Item – Topic Area: Counter-tracking----------

Your team has been tracking an experienced, well-armed band of insurgents for several days. The time/distance gap has been slowly closing to where it is now about 8 hours. You come upon where their tracks should be, but they have been obliterated by tracks of local cattle that cut through the ground spoor from several directions. Please rate the effectiveness of the following six decision options using this five-point scale. Don't hesitate to use the entire scale in judging these choices.

5 = highly effective/4 = moderately effective/3 = neutral/2 = moderately ineffective/1 = highly ineffective

[ ] Have one of your flanker trackers and the rear security tracker back track to the point where the cattle came from to see if the quarry's tracks are intermixed with them

[ ] Initiate a 360-degree lost spoor procedure

[ ] Look at surrounding tree branches in the immediate area for aerial spoor to estimate if/when the quarry had been there

[ ] Change to a Ranger/single file formation to look for any quarry ground spoor that might have escaped obliteration by the cattle

[ ] Change your tracking direction to follow the cattle path with the highest density of tracks

[ ] Slow pace of tracking movement to prepare for counter-tracking tactics

---

We used a similar procedure to create the Profiling SJT. In this case, we thoroughly reviewed our field notes from the Combat Hunter Limited Objective Evaluation, and identified 5 topic areas for focus: six combat domains (e.g., atmospherics, geographics), tactical cunning (think like the enemy), optical devices, tactical patience, and combat rules of 3. As with Tracking, we created two equally-difficult versions of the Profiling SJT, with six response options for each item that covered the range of effectiveness. Each SJT was administered to the students as a group during class time. The pre-test was given just prior to the first field scenario and the post-test was given on the day after the last field scenario.

For scoring, students' SJT answers were compared to those from the instructors who had also taken the SJT, which provided the answer key. To compute a score, we used the sum of the squared deviations from the instructors' response [8]. With this method, underestimates and overestimates of the SME rating are weighted equally, where extreme deviations are weighted more heavily. Higher scores mean worse performance since they are more discrepant from the SME's assessment. To determine degree of learning during the FTX, we computed each participant's difference score as their post-test score minus pre-test score; negative scores correspond to improved performance on the post-test.

3.2 Results

The SJT results are presented in Figure 1 for both class segments. The figure indicates the number of students whose post-test and pre-test difference score fell into one of the bins of size 10. Negative scores indicate a learning effect, as students' deviation score (from the instructors') was smaller on post-test (e.g., a desirable outcome).

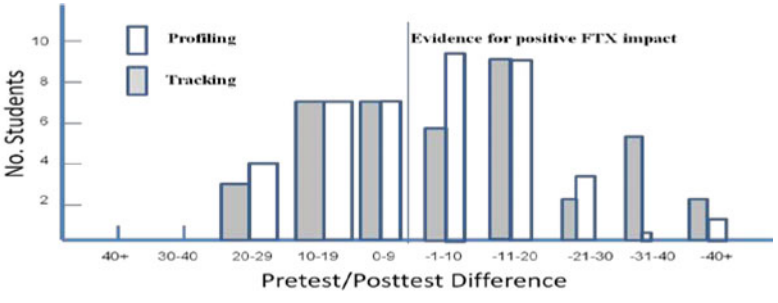


Fig. 1. Frequency distribution of SJT posttest/pretest difference scores

The frequency distribution gives us clues on the locus of the learning effect. For both course segments, there are no instances where students had high (30+) positive scores. On the right side of the Tracking distribution, seven participants had large decreases in their SJT score on the post-test, producing differences of -30 or greater. The FTX experience is to calibrate those students whose initial (pre-test) judgments were askew from the modal representation of the SME instructors. There was statistical evidence of learning as a paired t-test indicated that, on average, students' post-test scores were lower than their pre-test scores ($t = 2.229$, $p < .011$, $df = 41$).

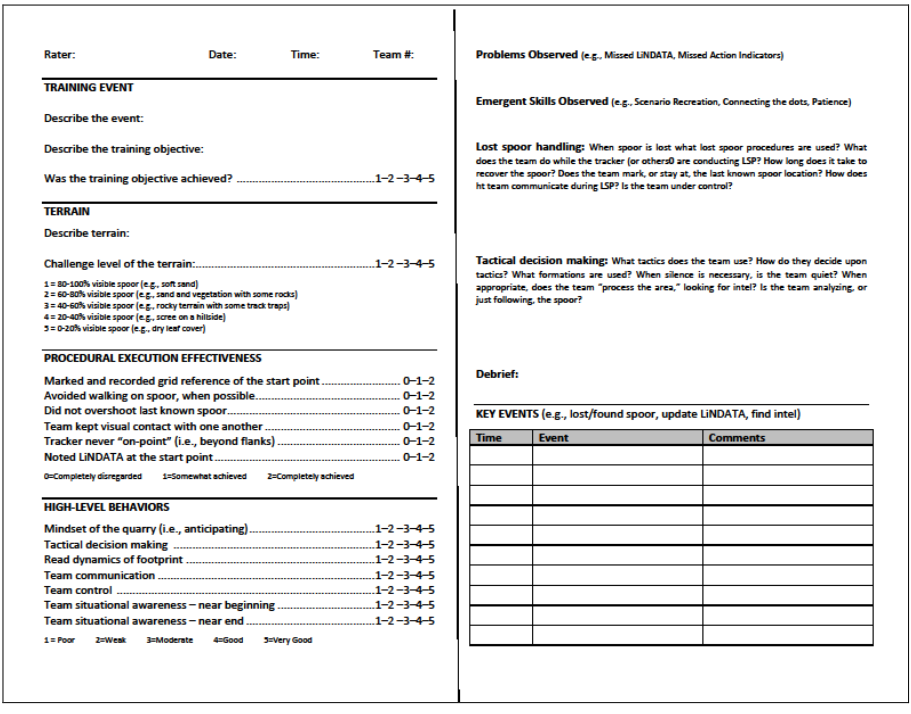
The Profiling effects were similar though smaller in magnitude. This was most evident in the -31-40 bin, where Tracking had five students, but Profiling had 0. The Profiling t-test was not significant, though the trend was in the right direction ($t = 1.114$, $p < .26$, $df = 39$). Part of the reduced effect was likely due to test fatigue, as students took the SJT post-test on the last day of training [11]. On balance, the SJTs indicated that students acquired improved judgment and decision-making from FTX experience.

4 Behavioral Observations

4.1 Method

Behavioral Observation Checklists (BOCs) offer a structured method to collect quantitative and qualitative data on individual and team performance during FTXs. The BOCs used in BH were modified from ones used in the CH evaluation [3]. Separate BOCs were created for Tracking and Profiling, where the former is shown in Figure 2. The instrument was created in a two-column layout so it could be folded for portability. The upper left part of the BOC has space to describe the training event for

that day. A set of 3-point rating scales gauge student proficiency in basic procedural skills whereas 5-point scales capture six high level behaviors. The right column lets the researcher comment on emerging skills, lost spoor handling (an important tracking skill), and decision-making. A similar format was used for the Profiling BOC.



Rater: Date: Time: Team #:

TRAINING EVENT

Describe the event:

Describe the training objective:

Was the training objective achieved? ..... 1–2 –3–4–5

TERRAIN

Describe terrain:

Challenge level of the terrain: ..... 1–2 –3–4–5

1 = 80-100% visible spoor (e.g., soft sand)
2 = 60-80% visible spoor (e.g., sand and vegetation with some rocks)
3 = 40-60% visible spoor (e.g., rocky terrain with some track traps)
4 = 20-40% visible spoor (e.g., scree on a hillside)
5 = 0-20% visible spoor (e.g., dry leaf cover)

PROCEDURAL EXECUTION EFFECTIVENESS

Marked and recorded grid reference of the start point 0–1–2
Avoided walking on spoor, when possible 0–1–2
Did not overshoot last known spoor 0–1–2
Team kept visual contact with one another 0–1–2
Tracker never "on-point" (i.e., beyond flanks) 0–1–2
Noted LiNDATA at the start point 0–1–2

0=Completely disregarded 1=Somewhat achieved 2=Completely achieved

HIGH-LEVEL BEHAVIORS

Mindset of the quarry (i.e., anticipating) ..... 1–2 –3–4–5
Tactical decision making ..... 1–2 –3–4–5
Read dynamics of footprint ..... 1–2 –3–4–5
Team communication ..... 1–2 –3–4–5
Team control ..... 1–2 –3–4–5
Team situational awareness – near beginning ..... 1–2 –3–4–5
Team situational awareness – near end ..... 1–2 –3–4–5

1 = Poor 2=Weak 3=Moderate 4=Good 5=Very Good

Problems Observed (e.g., Missed LiNDATA, Missed Action Indicators)

Emergent Skills Observed (e.g., Scenario Recreation, Connecting the dots, Patience)

Lost spoor handling: When spoor is lost what lost spoor procedures are used? What does the team do while the tracker (or others0 are conducting LSP? How long does it take to recover the spoor? Does the team mark, or stay at, the last known spoor location? How does ht team communicate during LSP? Is the team under control?

Tactical decision making: What tactics does the team use? How do they decide upon tactics? What formations are used? When silence is necessary, is the team quiet? When appropriate, does the team "process the area," looking for intel? Is the team analyzing, or just following, the spoor?

Debrief:

KEY EVENTS (e.g., lost/found spoor, update LiNDATA, find intel)

Time	Event	Comments

Fig. 2. BOC used to collect performance data during Combat Tracking

Performance data were collected by assigning one researcher to each team for each day of the Tracking FTXs. Researchers rotated teams over days to give ample exposure to the teams. A similar strategy was used for the Profiling FTXs. On each day, researchers met with the instructors to understand the objectives and goals of training for that day. In addition, two researchers were assigned to observe the same team on select days of the Tracking FTXs. This was done to allow an assessment of inter-rater reliability (Kappa = .59), which corresponded to 'moderate agreement'.

4.2 Results

Pooling rating data across teams and days yields stable quantitative trends showing how student BPRA behaviors improved. To smooth out daily fluctuations, we pooled data from successive days or scenarios for the basic procedural skills and the higher-level behaviors. A typical result is depicted in Figure 3, where performance ratings for three high level Tracking behaviors (e.g., reading footprint dynamics, adopting a

'quarry mindset', and tactical decision-making) – are plotted across days. Statistical analysis revealed that, despite starting out at fairly high levels, all three measures showed significant increases ($p < .025$) from Days 2-3 to Day 10.

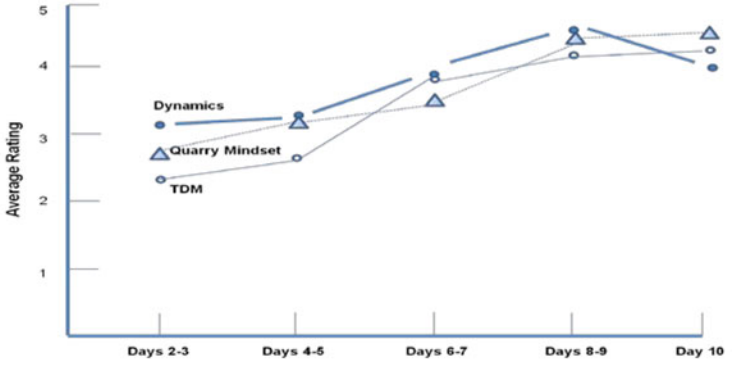


Fig. 3. Average performance ratings for three higher-level Tracking behaviors

Analysis of the quantitative data revealed that most procedural and higher-level behaviors exhibited similar increases across days or scenarios, indicative of a training effect. For Tracking, most of the basic procedural skills (3-point ratings) increased significantly across days, such as avoiding walking on the spoor line, maintaining visual contact, and marking the starting point of a track. For high level behaviors, besides those in Figure 3, we saw that students' performance on situation awareness, communication, and team control increased significantly ($p < .05$) over days. For Profiling, most procedural behaviors (e.g., spreading observations across team members, recording observations, using profiler language, establishing a stable baseline, using criteria to make a positive ID) improved over scenarios. All Profiling high level behaviors, rated on BOC 5-point scales, also improved. These included detecting basic events, adopting an insurgent mindset, interpreting complex events, communication, anticipating upcoming events, and exhibiting tactical patience.

The qualitative data on the BOCs, researcher comments, yielded key insights concerning the content of student behaviors, problems that emerged during the FTXs, and effective instructional techniques. Only a sampling of these results can be presented here; detailed findings are in [9]. For example, the Tracking BOCs show that early in training, students keep their heads mostly down so they can pick up the details of individual tracks (i.e., micro-tracking). While effective for seeing detail, it is slow. Later, students begin to look up more, using the pattern in the track line to discern where the quarry is likely headed (macro-tracking). This is a faster and more efficient technique. Many other aspects of tracking show similar qualitative trends.

Researcher comments on the Profiling BOCs also captured notable problems, trends, shifts in student behavior, effective instructional techniques, and emerging skills. Regarding the latter, students exhibited various skill improvements as they gained scenario experience, such as improved ability to identify high value individuals, synthesize events (i.e., 'connect the dots'), interpret complex events, and predict complex events from early signs. In later scenarios, even more complex skills

were emerging, such as scenario recreation ability, trust building, anticipation to get even more 'left of bang', and adopting the mindset of other cultures.

5 Emerging KSA Profiles

The third behavioral category entailed applying a comprehensive framework of KSAs, developed previously for CH [3], to researcher field notes to capture the richness with which the students' tracking and profiling competencies were emerging over the course of field training. Thirty-three KSAs were defined and organized into a six-category taxonomy [9]: use of enhanced observation techniques, identification of critical event indicators, interpretation of human behavioral cues, synthesis of ambiguous information, proactive analysis and dynamic decision-making, and employment of cognitive discipline. These KSAs apply equally to Tracking and Profiling, as they are both focused on a common set of IW cognitive and behavioral processes; only their manifestations as behavioral markers differ. An example KSA from Cognitive Discipline is shown below.

Table 1. Example of a Knowledge-Skill-Attitude

KSA	Behavior Marker – Profiling	Behavior Marker - Tracking
22. Keep an open mind to the unexpected (recognize there are unknown variables in the situation)	Do they consider the possibility that insurgents might use new tactics (e.g., different IED emplacing) or attempt something completely different than anything that has been tried before?	Do they consider that the hostiles might consider something completely different, like splitting up to rejoin at a rally point further down the track line?

The KSAs served several purposes in BH. They provide a framework to represent course content so training objectives from loosely-defined field scenarios could be established. KSAs link with training outcomes, and are observable, measurable, and trainable. They link Profiling and Tracking, establish validity of the training effort, and facilitate packaging of course materials for other venues.

Using this framework, three researchers took their field notes and instantiated all 33 KSAs for both course segments. A complete, emergent profile was developed that indicated the team (1-5), behavior observed, and day when the behavior occurred. Inter-rater agreement was high and the framework applied to both course segments, only the specific behaviors were different. These profiles were then employed by the larger research team to create a comprehensive program of instruction for BH [4].

6 Conclusions

Collectively, the three classes of behavioral measures reported here – SJTs, BOCs, and KSAs – form a powerful methodology to capture student performance during

FTXs. During BH training, BPRA competencies were assessed using the controlled testing conditions of SJTs, repeated quantitative elements of BOCs, and comprehensive qualitative KSA profiles. Applicable to both Profiling and Tracking, these behavioral measures are well-suited for "reading" the human terrain in CT operations. We believe that all three behavioral methods will be highly useful in other applications where field training of CT skills is taking place. When utilized by well-trained, experienced researchers, and with an appropriate level of setup and incorporation of subject matter expertise into the materials, SJTs, BOCs, and KSAs offer a highly effective method to assess training performance, capture training objectives, and identify the most successful instructional approaches for counter terrorism.

References

1. Department of Defense Directive 3000.07: Irregular Warfare (December, 2008), http://www.dtic.mil/whs/directives/corres/pdf/300007p.pdf (Retrieved from June 28, 2010)
2. Kobus, D., Williams, G.: Training tactical decision making under stress in cross- cultural environments. In: Proceedings of the Conference on Cross-Cultural Decision Making, Miami, FL [CD-ROM] (2010)
3. Spiker, V.A., Johnston, J.H.: Limited objective evaluation of combat profiling training for small units. (Technical Report). US Joint Forces Command, Suffolk, VA (January 2010)
4. Institute for Training and Simulation. In: Schatz, S., Fautua, D.(eds.): Border Hunter research technical report. US Joint Forces Command, Suffolk, VA (July 2010)
5. Gessner, T.L., Klimoski, R.J.: Making sense of situations. In: Weekley, J.A., Ployhart, R.E. (eds.) Situational judgment tests. LEA, Mahwah (2006)
6. McDaniel, M.A., Morgeson, F.P., Finnegan, E.B., Campion, M.A., Braverman, E.P.: Use of situational judgment tests to predict job performance: A clarification of the literature. Journal of Applied Psychology 86(4), 730–740 (2006)
7. Brooks, M.E., Highhouse, S.: Can good judgment be measured? In: Weekley, J.A., Ployhart, R.E. (eds.) Situational judgment tests. LEA, Mahwah (2006)
8. Weekley, J.A., Ployhart, R.E., Holtz, B.C.: On the development of situational judgment tests: Issues in item development, scaling, and scoring. In: Weekley, J.A., Ployhart, R.E. (eds.) Situational judgment tests. LEA, Mahwah (2006)
9. Spiker, V.A., Johnston, J.H.: Border Hunter: Evaluation of field training (Technical Report). US Joint Forces Command, Suffolk, VA (July 2010)