Applied Ergonomics 44 (2013) 680–686

Contents lists available at SciVerse ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Assessing Counter-Terrorism field training with multiple behavioral measures

V. Alan Spiker [a,*], Joan H. Johnston [b,1]

[a] Anacapa Sciences, Inc., 301 E. Carrillo St., Santa Barbara, CA 93101, USA
[b] Naval Air Warfare Center Training Systems Division, 12350 Research Parkway, Orlando, FL 32826-3275, USA

ARTICLE INFO

Article history:
Received 14 March 2011
Accepted 26 April 2012

Keywords:
Situational judgment tests
Behavioral observations
Scenarios
Knowledge–skills–attitudes
Profiling
Tracking

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-sponsored course consisting of Combat Tracking and Combat Profiling segments. Measures included situational judgment tests, structured behavioral observation checklists, and qualitative assessments of the emergence of specific knowledge–skills–attitudes over the course of the training. The paper describes statistical evidence across the three types of measures that indicate that behavioral pattern recognition and analysis skills were successfully acquired by most students (a mix of Army and civilian law enforcement personnel) during the field training exercises. Implications for broader training of these critical skills are also discussed.

1. Introduction

In 2008, the US Department of Defense (DOD) placed Irregular Warfare (IW) on an equal footing with conventional warfare in future military planning and operations (DOD, 2008). Among IW mission objectives are developing capabilities to address asymmetrical threats and the challenges they pose for Counter-Terrorism. Whether practiced by military ground units or law enforcement personnel (e.g., Customs and Border Patrol), reading the human terrain – such as through Behavioral Pattern Recognition and Analysis (BPRA) skills – is an essential element of Counter-Terrorism 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 (Kobus and Williams, 2010). As a result, behavior profiling techniques have become a valuable addition to small unit tactics, techniques, and procedures.

Nowhere are these skills more highlighted than in Combat Hunter training conducted by the United States Marine Corps School of Infantry. Combat Hunter is a 10 day course taught in two segments, Combat Profiling and Combat Tracking, by subject matter experts (SMEs) in these respective disciplines. Each segment has a classroom academic portion and a scenario-based field exercise portion; the latter segment is where the skills are developed, applied, refined, and reinforced. Historically, it is difficult to assess training effectiveness in the field since the observation conditions are difficult, the curriculum is not always standardized, its objectives are not always well-specified, and instructional delivery by trainers is highly variable.

To document training acquisition under these demanding conditions, a wide array of behavioral methods and measures are needed. In this paper, we summarize the results of a field study in which we observed the development and acquisition of BPRA skills by military and law personnel as part of formal CT training in a special offering of Combat Hunter called Border Hunter. Specifically, we describe empirical results based on applying three different methods of behavior measurement: situational judgment tests (SJTs), behavior observation checklists (BOCs), and knowledge–skill–attitude (KSA) profiles.

Each type of behavioral measure contributes valuable information that enhances our understanding of the extent to which BPRA skills have been acquired in a Field Training Exercise (FTX) setting. Thus, SJTs delineate the extent to which higher order cognitive skills have been acquired in the context of controlled pretest/posttest assessments. BOCs, on the other hand, yield quantitative (via ratings) and qualitative (via observer comments) evidence on the acquisition of fundamental and higher-order behavior skills. KSA profiles, tailored to the KSAs anticipated from the FTX, yield indicators of emergent competencies that underlie the success of individual and team training.

* Corresponding author. Tel.: +1 805 729 0148.
E-mail addresses: vaspiker@anacapasciences.com (V.A. Spiker), joan.johnston@navy.mil (J.H. Johnston).
[1] Tel.: +1 407 380 4988.

0003-6870/$ – see front matter © 2012 Elsevier Ltd and The Ergonomics Society. All rights reserved.
http://dx.doi.org/10.1016/j.apergo.2012.04.015

Collection and interpretation of this constellation of behavioral indices is consistent with a multi-measure, multi-method approach that has long been advocated for evaluation of interventions in field settings (Cook and Campbell, 1979). Within this approach, analysts are better able to "triangulate" on a given intervention effect if the results of multiple indices are pointing in the same direction. In the following discussion, we address both the content of the skills measured and the behavioral evidence for their acquisition.

2. Method

2.1. Observing Combat Hunter training to construct instruments

Instruments to collect behavioral measures were developed during the authors' naturalistic observation of Combat Hunter training at the School of Infantry – West, Camp Pendleton. The 10 days of Combat Hunter training were split equally between Combat Tracking and Combat Profiling. Each class consisted of approximately 40 students drawn from the same Marine regiment, though students typically came from different platoons and squads. Both segments were split into academic instruction and field scenario portions. In Combat Tracking, students received academic instruction in the morning on the fundamentals of tracking (e.g., dynamics of footprints, maintaining track line, and interpreting spoor). Afternoons were devoted to applying this knowledge in the field, where students tracked "quarry" (role-playing instructors) as five-person tracking teams. During these tracking scenarios, students learned to read their enemies' spoor (i.e., footprints, human signs, environmental cues, and slight ground disturbances). They were also taught to build social/biometric profiles of their quarry, anticipate their targets' actions by acquiring the mindset of the quarry, and apply their own unit's standard operating procedure to hunt down their targets. Combat Tracking is a human-centric competency that is particularly useful in IW settings to support offensive operations, intelligence gathering, clandestine movement in hostile areas, and counterinsurgency operations. Over days, the field scenarios increased in complexity as the instructors added 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 was structured differently, three days of academic instruction preceded the two days of 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 before a destructive event occurs. This is known as staying "left of bang." It trains individuals to look for behavioral anomalies that are beyond the baseline of a culture or a particular location. Through 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 were 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: heuristics, geographics, proxemics, atmospherics, biometrics, and kinesics (Kobus and Williams, 2010). The practical application portion was conducted in the next two days where students were split into teams and man observation posts to observe role-players engaged in varying types of behavior within a village mockup. They practiced their profiling skills by observing, at a distance, instances of neutral and insurgent behavior in the context of increasingly challenging scenarios. Typically, five to six such scenarios were executed during the two days. The training culminated in a 4 h long final exercise where all teams deployed as maneuver units into the village using their insights gained from the previous scenarios.

The authors observed two evolutions of the Combat Hunter course; these observations were used to develop draft versions of the three types of behavioral measure instruments (Spiker and Johnston, 2010a). In addition to 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 6-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 (Schmitt and Chan, 2006). The BOCs were structured so that observers could provide 3- or 5-point ratings on a set of basic and advanced profiling or tracking skills. For KSA profiles, a taxonomy of 33 KSAs was created that would be applicable to both course segments. These KSAs are higher order competencies that comprise a mix of cognitive, behavioral, and social elements that underlie utilizing BPRA in the field. The taxonomy would be used retrospectively by observers, after the training day ended, as a way to categorize their free form written observations of trainee behaviors. Further details on each instrument are given below.

2.2. Behavioral measure instruments

2.2.1. Situational judgment tests

SJTs are low- to moderate-fidelity work sample simulations that assess preferences for appropriate behaviors in a work setting (Gessner and Klimoski, 2006). 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 shown moderate concurrent validity with performance (McDaniel et al., 2006), we chose to use them in this study to assess degree of learning in the field exercises for both Combat 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 (Brooks and Highhouse, 2006). If students have been receiving this experience by participating in FTXs, then they should exhibit improved performance on the SJTs between pretest and posttest.

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 instructor responses to the same test as the "answer key," we assessed how the students' mental representations of various real world problems began to resemble those of the instructors.

For the Tracking SJTs, our review identified six skill areas that are particularly important for Counter-Terrorism success and amenable to testing via a short, written scenario: methods for closing the time/distance gap (with the quarry), 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 this problem. In developing these options, two options were specified

that would be superior, two inferior, and two that would have both strengths and weaknesses.

An example item from one of the Tracking SJTs is shown in Fig. 1. The scenario setup is brief, reducing the reading requirement and leaving desirable information omitted. Students must make inferences and exercise judgments which are aided by the experience obtained during the FTX – the main intent of the SJT.

A similar procedure was used to create the Profiling SJT. In this case, the field notes from the Combat Hunter observations were thoroughly reviewed and five topic areas were identified: the six combat domains (e.g., atmospherics, geographics), tactical cunning (think like the enemy), optical devices, tactical patience, and combat rules of 3 (having three pieces of evidence to make a positive identification). As with Tracking, two equally difficult versions of the Profiling SJT were created, with six response options for each item that covered the entire range of "effectiveness."

For scoring, students' SJT answers were compared to those from the instructors who had also taken the SJT. To compute a score, we used the sum of the squared deviations from the instructors' modal response (Weekley et al., 2006). With this method, underestimates and overestimates of the instructor rating are weighted equally, where extreme deviations are weighted more heavily. Higher scores indicate worse performance since they are more discrepant from the instructors' assessment. To determine degree of learning during the FTX, we computed each subject's difference score as their posttest score minus pretest score; negative scores correspond to improved performance on the posttest.

2.2.2. Behavioral observation checklists

Behavioral Observation Checklists (BOCs) offer a structured method to collect quantitative and qualitative data on individual and team performance during field exercises. The BOCs used in this study were modified from ones used in the Combat Hunter evaluation (Spiker and Johnston, 2010a). Separate BOCs were created for Tracking (shown in Fig. 2) and Profiling. The instrument was created in a 2-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. Six 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, with items tailored to its skill objectives, was used for the Profiling BOC.

2.2.3. Knowledge–skills–attitude profiles

The third behavior measure entailed applying a comprehensive KSA framework, originally developed for Combat Hunter (Spiker and Johnston, 2010a), to researcher field notes to capture how students' tracking and profiling competencies were "emerging" over the course of the FTX. Thirty three KSAs had been previously defined and organized into a 6-category taxonomy (Spiker and Johnston, 2010b): 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 focused on a common set of cognitive and behavioral processes; only their manifestations as "behavioral markers" differ. Two example KSAs, from Cognitive Discipline (#18) and Synthesis of Ambiguous Information (#28), are shown below.

The KSAs served several purposes in the Border Hunter observations. First, they provided a framework to represent course content so training objectives from loosely defined field scenarios could be established. KSAs also link with training outcomes; they are observable, measurable, and trainable. Third, the KSAs link the Profiling and Tracking course segments, establishing training validity and facilitating packaging of course materials for other venues, such as on-line courseware (Institute for Simulation and Training, 2010).

2.3. Data collection during Border Hunter training

Border Hunter was conducted over 21 days, and was a "graduate level" version of Combat Hunter that includes Combat Tracking and profiling instruction (Institute for Simulation and Training, 2010). Border Hunter was sponsored by the US Joint Forces Command, and was conducted by Joint Task Force North at Fort Bliss, Texas, in April 2010. The first 10 day segment was for Combat Tracking and was taught by six highly experienced trackers with a combined experience of 180 years. Similarly, Combat Profiling was taught by an 8-person team of instructors with a collective experience in military and police work of more than 200 years.

Forty-three trainees, comprising a mix of US Army, Border Patrol, and other law enforcement personnel, received Border Hunter training. All were highly experienced, with an average of nine years in military and/or law enforcement. In addition, 22 soldiers from Fort Bliss were recruited as role players to participate in the Combat Profiling field scenarios.

A 13-person research team had two objectives: conduct behavioral research to assess effectiveness of the course's academic instruction and FTX; and document course delivery via videotape, interviews, and recorded observation in order to produce a prototype online course. Regarding the former, on a given day, three to six highly experienced researchers with advanced degrees in behavioral science were available to collect behavioral data on student training performance.

Students were assigned to the same teams for all Tracking and Profiling FTX scenarios. Teams 1 and 2 were composed of students

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 5-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

Fig. 1. Combat Tracking SJT item for counter tracking.

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.

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 comprised a mix of Army and law enforcement personnel.

SJTs were administered to students as a group during class time. The pretest was given just prior to the first field scenario and the posttest given on the day after the last field scenario. Students' identity was kept anonymous, with each assigned a unique ID number at the beginning of the study. The primary author gave a brief description of the SJT and was available to answer questions while the students completed the test. Students took, on average, about 20–25 min to complete each SJT.

BOC performance data were collected by assigning one researcher to each team for each day of the Tracking exercises. Researchers rotated teams over days to give ample exposure to all students. A similar strategy was used for the Profiling scenarios. 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 exercises. This allowed an assessment of inter-rater reliability for the ratings to the procedural level and high level behavior BOC categories. Separate Kappa statistics (Cohen, 1960) were collected for the two sets of ratings. The resultant Kappa statistics were .590 and .553, respectively, which correspond to "moderate agreement" on the Landis and Koch (1977) scale. Given the limited time available to caucus on use of the BOC, these results were encouraging, missing "substantial agreement" (.61–.80) by only a few percentage points.

Using the KSA framework described in Section 2.2.3, three researchers reviewed the comments in their BOCs and assigned them to one of the 33 KSAs. The researchers had been briefed on the KSAs prior to the start of data collection and used the framework to organize and phrase their comments during the observation periods. A comprehensive KSA profile was created that chronicled the emergence of competencies over training days. 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 Border Hunter (Institute for Simulation and Training, 2010).

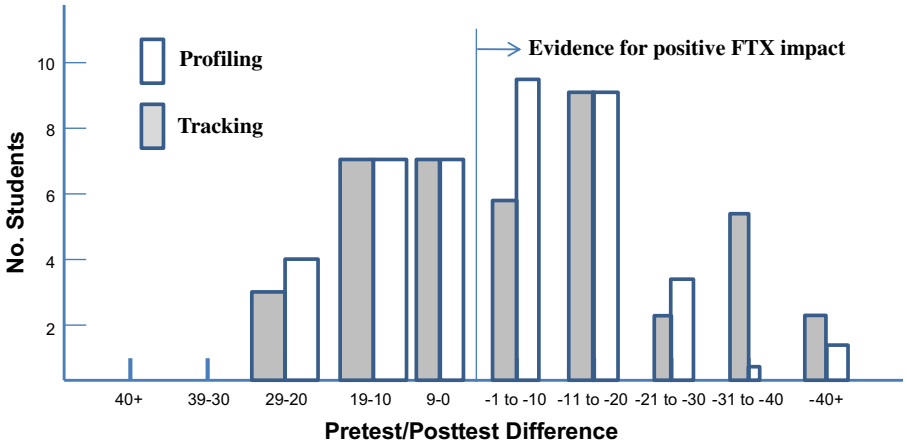

Fig. 3. Frequency distribution of SJT posttest/pretest difference scores.

3. Results

3.1. Situational judgments

The SJT results are presented in Fig. 3 for both class segments. The figure depicts the number of students whose posttest-pretest 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 posttest, a desirable outcome. The bins are thus ordered and labeled to reflect this effect, such that positive learning outcomes (i.e., larger negative difference scores) appear in the right half of the figure. Note that the positive and negative bins have different label boundaries (e.g., 39–30 vs −31 to −40) since we have put the 0-score (no pretest/posttest difference) on the left side of the dividing line, to indicate no learning effect.

The nature of 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 frequency distribution for the Tracking SJT, seven subjects had large decreases in their SJT score on the posttest, producing differences of −30 or greater. The FTX experience, then, appears to "calibrate" those students whose initial (pretest) judgments were askew from the modal representation of the instructors. There was statistical evidence of learning as a paired *t*-test indicated that, on average, students' post-tests scores were lower than pretest ($t = 2.229$, $p < .025$, df = 41).

The Profiling effects are similar, though smaller in magnitude. This is most evident in the −31 to −40 bin, where Tracking had five students, but Profiling had 0. The Profiling *t*-test missed significance ($t = 1.144$, $p < .26$, df = 39), though the trend was in the predicted direction. Part of the reduced effect was likely due to test fatigue, as students took the SJT post-test on the 21st and last day of Border Hunter training (Spiker and Johnston, 2010b). The SJTs suggest that students acquired improved judgment and decision making from the FTX experience.

3.2. Behavioral observations

For BOC data, we pooled ratings across teams and two days to generate stable quantitative trends showing how student BPRA behaviors improved. To smooth out daily fluctuations, data were combined from successive days or scenarios for the basic procedural skills and the more highly skilled behaviors. A typical result is depicted in Fig. 4, where performance ratings for three highly skilled Tracking behaviors – reading footprint dynamics, adopting a "quarry mindset," and tactical decision making – are plotted across days. Statistical analysis (*t*-tests) revealed that, despite starting out at fairly high levels, all three measures showed significant increases ($p < .025$) from Days 2–3 to Day 10.

Analysis of the quantitative data revealed that most procedural and higher level behaviors exhibited similar increases across days (Tracking) or scenarios (Profiling), indicative of a training effect. For Tracking, most 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 the ones in Fig. 4, 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's 5-point scales, also improved. These include detecting basic events, adopting an insurgent mindset, interpreting complex events, communication, anticipating upcoming events, and exhibiting tactical patience.

The qualitative data on the BOCs, obtained from researcher comments, yielded key insights concerning the content of student behaviors, problems that emerged during the field exercises, and effective instructional techniques. Only a sampling of these results can be presented here; detailed findings are in Spiker and Johnston (2010b). For example, the Tracking BOCs showed that early in training students kept their heads mostly down so they could pick

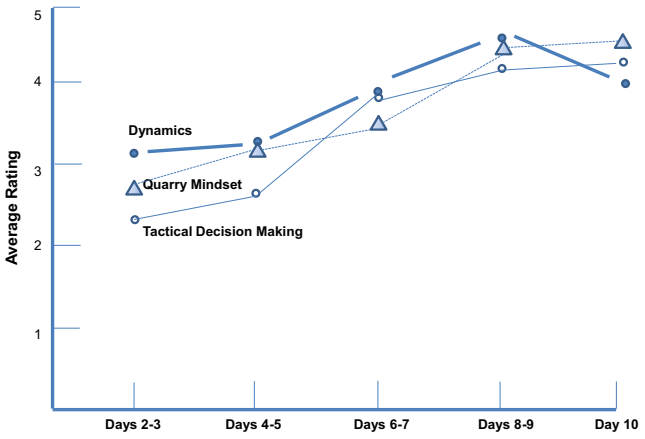

Fig. 4. Average performance ratings for three higher level tracking behaviors.

Table 1
Example of KSA items 18 and 28.

KSA	Behavior marker – Profiling	Behavior marker – Tracking
#18. 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?
#28 Imagining alternative courses of action or alternative event outcomes by what-if mental simulations	Does the profiler try to "think through" what might be happening in an unfolding event (e.g., a possible complex ambush) by running through different alternative outcomes	Does the tracker or team leader try to "think through" what the quarry might be doing (ahead of them) based on the track pattern they are looking at?

Table 2
Two KSA examples of observed behaviors.

KSA	Tracking observation:	Profiling observation:
#18 Imagining alternative courses of action or event outcomes by what-if mental simulations.	Team Leader decides to have team follow main group if quarry splits off (Team 3, Day 4); After quarry moved from initial sighting point, team attempted to predict which ravine they might have headed for. (Team 5, Day 5)	Individual team members work through and talk through the courses action that the Prince might take following the extensive violence in his village, including negotiation, fleeing, fighting, or calling in the US (Team 4, Scenario 6)
#28 Keeping an open mind to the unexpected (recognizing there are unknown variables).	Team stops and looks at action indicator, examining the quarry's formation to determine if they are splitting up and planning an ambush (Team 3, Day 4)	The team begins discussion of what kind of signs might be present to indicate a possible attack on the neutral nomadic village, including sniper-looking individuals lurking by the tents and the possibility that some of them might become more active in village affairs such as attending a wedding with the Prince (Team 4, Scenario 3)

up the details of individual tracks (i.e., micro-tracking). While effective for seeing detail, it was slow. Later, students began using the more efficient technique of looking up more frequently, using the pattern in the track line to discern where the quarry is likely headed (macro-tracking).

Many other qualitative aspects of tracking showed comparable improvements over training days. Notable skills exhibiting learning trends included better team formation despite rough terrain, more efficient communication among team members (particularly between the tracker and team leader), greater awareness of the relation between terrain and quarry's likely intent, and greater integration of binoculars and naked eye viewing to stay on the track line even in rocky terrain.

Researchers' 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 ("connect the dots"), interpret complex events, and predict events from early signs. In later scenarios, more complex skills emerged, such as scenario recreation ability, trust building, earlier anticipation to get even more "left of bang," and adopting the mindset of other cultures.

3.3. Emerging KSAs

The KSA profile was generated by taking researcher field notes and entering specific statements into one of the 33 KSAs; this was done for both Combat Tracking and Profiling. To illustrate the relationship between observed behaviors and KSA assignment, Table 2 presents examples for the two KSAs from Table 1, Imagining Alternative Courses of Action and Keeping an Open Mind.

Because we were not able to collect comments from all researchers for this measure, a detailed analysis of frequency counts is not possible. Nevertheless, several trends deserve mention. First, it was fairly straightforward to extract behaviors from the BOC and assign them to their respective KSAs. This was true for both course segments, and suggests the KSAs as presently stated lend themselves to behavioral representation without much extrapolation. Thus, they are observable because they have fairly clear behavioral markers. Moreover, this was true for all KSAs, where all 33 had at least one, and most had many, observed instances of student behavior over training.

With regard to Tracking, the frequency of observed behaviors was comparable across five of the six KSA categories, with one exception – Use of Enhanced Observation Techniques, which had a count almost double that of the other categories. Three KSAs had frequencies twice as high as the others. These were Using Organic Assets, Taking Someone Else's Perspective, and Anticipating What Will Happen Next. It is difficult to determine if these KSAs were actually more prevalent during the training, if their behavioral markers were easier to spot, or it is a sampling bias from the two researchers providing data. All three KSAs relate to the primary focus of Border Hunter training, so it may represent something more fundamental. Further data collection and analysis are needed on this topic.

Turning to Combat Profiling, we noted that the frequency distribution of behaviors across KSAs and KSA categories differed from Tracking. Specifically, two categories had frequencies substantially higher than the others: Identification of Critical Event Indicators and Proactive Analysis and Dynamic Decision Making. These categories are different than the leader for Tracking, which may suggest a different behavior mix for the two course segments (vs sampling bias). One KSA – Looking for Anomalies – was populated with behaviors twice that of the nearest competitor. This may reflect the importance of detecting anomalies from an established baseline within the Profiling instructional methodology, or perhaps these behavioral instances were easier for researchers to spot. Further data collection and analysis are required to clarify this issue.

4. Discussion and conclusions

The three classes of behavioral measures reported here – SJTs, BOCs, and KSAs – form a powerful methodology to capture student Counter-Terrorism skills during Border Hunter FTXs. 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 Counter-Terrorism operations.

Overall, the SJTs provide a valuable source of data to gauge the learning impact of FTX in Combat Tracking and Profiling. Importantly, the results of statistical testing showed that, on average, students exhibited significantly improved performance on the SJT Tracking posttest. The locus of this effect was the elimination of students having marked discrepancies from the instructors' assessments of the scenarios. This is a major goal of any scenario-based, FTX activity – to reinforce correct judgments of complex, ambiguous situations and weed out discrepant, ill-conceived inferences. While the Profiling SJT failed to achieve comparable statistical significance, it nevertheless revealed a trend in the direction of improved performance on the pre-test.

To the best of our knowledge, this study is one of the first uses of SJTs to measure degree of learning in a FTX environment. Although considerable effort is required to construct, vet, and fine tune the scenario items and the response options, the benefits from this type of testing are considerable. Not only are SJTs a reasonable proxy for job performance, they offer a controlled method to gauge student acquisition of complex decision making, judgment, and inferential processes. As such, they represent a valuable tool in the researcher's arsenal. We believe that SJTs hold enormous promise in future applications of field learning assessments in courses where complex cognitive processes underlie successful skill acquisition.

The BOC data provide compelling evidence for skill acquisition of core tracking and profiling skills based on researcher observations of the FTXs. Improvement as a function of training days or scenarios was evident during both segments of the course, based on both quantitative (ratings) and qualitative (comments, observations) data from the instruments. While this was most notable at the team level, it was clear that this was also due to improved performance of individual trainees who were occupying key roles, particularly team leaders and trackers for Combat Tracking and squad leaders for Combat Profiling.

For Tracking, there was statistical evidence for improvement in basic tracking behaviors (macro-tracking, marking the initial point, not walking on the spoor line) despite experiencing more difficult terrain and more complex training objectives. Evidence was also obtained for consistent improvement in the high level behaviors captured on the BOC, including adopting a quarry mindset, tactical decision making, reading the dynamics of the footprint, communication, situation awareness, and team control.

While statistical support for Profiling performance improvement was somewhat weaker, there was clear evidence that many of the basic Profiling behaviors improved, such as distributing duties, using cue clusters to make a positive ID, and maintaining observational discipline. Higher level behaviors also showed significant increases across scenarios, including adopting an insurgent mindset, detecting basic events, interpreting complex events, external communication, and tactical patience.

With regard to KSAs, three aspects of competency specification have important implications for small unit training in IW generally and Counter-Terrorism particularly. One, the KSAs, by design, encompass a broad range of behaviors such that, once acquired, they should transfer to a trainee's other courses, tasks, and mission areas. Thus, focusing training on these KSAs should lead to improved job performance. Two, the KSAs reflect a mix of cognitive behaviors that can be organized based on a previously established theoretical model for team decision making. In this vein, Spiker et al. (2010) categorized the skills within this model as: identification skills, elaboration skills, and cognitive monitoring skills. Third, depending on how a given skill is operationalized, one can view its manifestation at either the individual or team level, where most of the KSAs have elements of both.

In closing, we recommend these three behavioral methods and tools be used where field training of Counter-Terrorism skills is taking place. When utilized by well trained, experienced observers, and with an appropriate level of subject matter expertise contributing to tool development, SJTs, BOCs, and KSAs offer a highly effective method to identify training objectives, assess trainee performance, and identify the most successful Counter-Terrorism instructional approach.

References

Brooks, M.E., Highhouse, S., 2006. Can good judgment be measured? In: Weekley, J.A., Ployhart, R.E. (Eds.), Situational Judgment Tests. LEA, Mahwah, NJ.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 76 (5), 378–382.

Cook, T.D., Campbell, D.T., 1979. Quasi-Experimentation: Design & Analysis for Field Settings. Rand-McNally, Chicago, I.L.

Department of Defense Directive 3000.07: Irregular Warfare (December, 2008). Retrieved June 28, 2010 from http://www.dtic.mil/whs/directives/corres/pdf/300007p.pdf.

Gessner, T.L., Klimoski, R.J., 2006. Making sense of situations. In: Weekley, J.A., Ployhart, R.E. (Eds.), Situational Judgment Tests. LEA., Mahwah, NJ.

Institute for Simulation and Training, 2010, July. In: Schatz, S., Fautua, D. (Eds.), Border Hunter Research Technical Report. US Joint Forces Command, Suffolk, VA.

Kobus, D., Williams, G., 2010. Training tactical decision making under stress in cross- cultural environments. In: Proceedings of the Conference on Cross-Cultural Decision Making, Miami, FL [CD-ROM].

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

McDaniel, M.A., Morgeson, F.P., Finnegan, E.B., Campion, M.A., Braverman, E.P., 2006. Use of situational judgment tests to predict job performance: a clarification of the literature. Journal of Applied Psychology 86 (4), 730–740.

Schmitt, N., Chan, D., 2006. Situational judgment tests: method or construct? In: Weekley, J.A., Ployhart, R.E. (Eds.), Situational Judgment Tests. LEA, Mahwah, NJ.

Spiker, V.A., Johnston, J.H., 2010a. Limited Objective Evaluation of Combat Profiling Training for Small Units (Technical Report). US Joint Forces Command, Suffolk, VA.

Spiker, V.A., Johnston, J.H., 2010b. Border Hunter: Evaluation of Field Training (Technical Report). US Joint Forces Command, Suffolk VA.

Spiker, V.A., Johnston, J.H., Williams, G., Lethin, C., 2010, December. Training tactical behavior profiling skills for irregular warfare. In: Proceedings of the 30th Interservice/Industry Training Systems and Education Conference. Orlando, FL.

Weekley, J.A., Ployhart, R.E., Holtz, B.C., 2006. 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, NJ.