Assessing Operational Decision Making

David A. Kobus, Erica P. Viklund, Jason M. Kobus, Brian Dister & Matthew Kelly
Pacific Science & Engineering Group, San Diego, CA

davidkobus@pacific-science.com, ericaviklund@pacific-science.com, jasonkobus@pacific-science.com, briandister@pacific-science.com, matthewkelly@pacific-science.com

ABSTRACT

Operational decision making is one of the foremost cognitive demands in the field for infantry. Thus, there is increasing emphasis on training that enables Warfighters to build adaptive operational decision making (ODM) skills. Assessment of the effectiveness of this training is critical for determining its value with respect to both the cost and impact on Warfighters' limited time. To this end, an assessment method was developed and evaluated during a series of studies conducted at the Camp Pendleton Infantry Immersion Trainer (IIT). The method includes pre- and post-training assessments to allow quantification of training-related changes in ODM. Twenty-five IIT training scenarios were decomposed to extract ODM training objectives, from which 26 decision areas were selected. The decision areas fall into four competency areas emphasized during IIT training (Combat Hunter/Every Marine a Collector skills; Communication; Cultural Interaction; Infantry Squad/Team Skills). Multiple decision dilemmas resembling situations that may be encountered during field operations were developed for each decision area. Five potential courses of action (COAs) were developed for each dilemma. An assessment consists of a series of items (dilemmas and COAs); trainees use a 5-point scale to rate the effectiveness of each COA for addressing the associated dilemma. As part of the item development process, the dilemmas and COAs were subjected to a two-stage verification and validation process by separate sets of infantry subject matter experts (SMEs) to ensure face validity and consensus of SME ratings of COAs. Trainees completed one assessment prior to training, and another with similar (but not identical) items immediately following training. Their ratings were compared to those of SMEs. This paper describes the approach in greater detail, along with preliminary assessment results and a discussion of factors contributing to ODM training effectiveness.

ABOUT THE AUTHORS

Dr. David A. Kobus is the Director of Warfighter Performance at Pacific Science & Engineering Group, of San Diego, CA. He has been involved in human performance research and project management for over 25 years. He has served as the principal investigator, or as program manager, on over 30 research projects related to enhancing human performance. These studies have incorporated a wide variety of measurement techniques to assess human performance including behavioral, physiological, and cognitive psychophysiological techniques. He was also a co-winner of the 2010 NTSA Modeling & Simulation Awards, and a finalist for the Governor's Award for Outstanding Achievement in Modeling & Simulation. He is currently involved in several projects related to the assessment of cognitive readiness of small unit decision makers. He has extensive teaching experience at both the graduate and undergraduate level and is a retired professor from San Diego State University. He holds an MS and PhD from Syracuse University.

Dr. Erica P. Viklund is a Senior Scientist with Pacific Science & Engineering. Her interests and research experience center on the application of cognitive, perceptual, and human factors psychology to Warfighter training and performance. She has significant experience in individual assessment and performance measurement, using behavioral, psychophysiological, and neurophysiological metrics, as well as principles of human systems integration (HSI). She is also experienced in functional and cognitive task analyses, scenario based training, and modeling human cognition. In her work with Pacific Science & Engineering, she has provided scientific support for the development and assessment of a training program for enhancing Warfighters' awareness of critical cues in their surroundings, assessed the effectiveness of a recently developed immersive training facility for infantry squads, assisted in the development of metrics for assessing decision making during virtual and immersive training, and provided human factors expertise to support an HSI evaluation of a modernized U.S. Marine Corps logistics system designed to provide enhanced efficiency/effectiveness for meeting supply and maintenance requirements.

Jason M. Kobus is a Research Analyst with Pacific Science & Engineering Group (PSE). He has nearly ten years' experience working with the military in scenario-based training, field data collection, interviewing Subject Matter Experts (SMEs), and immersive environment training. He is also experienced in conducting usability evaluations of both hardware and software products, data analysis, and multi-media product development for both military applications and medical devices. He has captured physiological data and performed analyses assessing decision making in immersive and mixed-reality environments at the squad level. He is also Combat Hunter trained and his field work experience includes the Combat Hunter and Border Hunter Programs. He also serves as the quality assurance lead of product development for the Warfighter Performance area at PSE.

Brian Dister is the Director of Engineering and a Senior Software Architect at PSE. He has spent twenty-five years in Research and Development, within both the scientific community and the high tech sector, developing award-winning software and hardware including Macromedia Authorware and Flash. Brian brings his wealth of experience in user centered design and development to a variety of multi-disciplinary efforts at PSE

Matthew Kelly is a Research Scientist with Pacific Science & Engineering. His experience focuses on using behavioral, physiological, and neurological metrics to gauge performance. He also has experience working within virtual environments (e.g. VBS 2, Second Life). While at Pacific Science & Engineering, he has provided support toward assessment of decision making within immersive training environments, assessment of cognitive performance under physical load, and supported the development of a metric for perceptual training.

Assessing Operational Decision Making

David A. Kobus, Erica P. Viklund, Jason M. Kobus, Brian Dister & Matthew Kelly
Pacific Science & Engineering Group, San Diego, CA

davidkobus@pacific-science.com, ericaviklund@pacific-science.com, jasonkobus@pacific-science.com, briandister@pacific-science.com, matthewkelly@pacific-science.com

INTRODUCTION

> *In many cases, the individual Marine will be the most conspicuous symbol of American foreign policy and will potentially influence not only the immediate tactical situation, but the operational and strategic levels as well. His actions, therefore, will directly impact the outcome of the larger operation; and he will become -- the Strategic Corporal.*
>
> —General Charles Krulak, USMC

Operational decision making is one of the foremost cognitive demands in the field for infantry. United States troops of all ranks and experience levels are required to conduct operations and to make decisions that may have an effect not only on the immediate tactical situation, but also on larger operational and strategic issues (cf., Krulak 1999). The information processing demands placed on Warfighters are enormous, and decision makers in the field are most often not afforded the time to combine sequential, procedural, and comparative methods to solve problems. Crucial to enhancing performance under such demanding conditions is training that quickly advances Warfighters beyond declarative knowledge to a level of mastery that supports expert-level decision making. Numerous training venues (such as the Infantry Immersion Trainers) have been developed that support deliberate practice and enhancement of operational decision making. However, the effectiveness of such training for meeting specific objectives is often undetermined due to a lack of measures and metrics for assessing cognitive skills such as decision making. To truly know if training is effective, empirically based objective measures are needed. Assessment of the effectiveness of training is critical for determining its value with respect to both the cost and impact on Warfighters' limited time.

This paper describes an approach derived from situational judgment tests (SJTs; see Whetzel & McDaniel, 2009; Kobus et al, 2010) and tactical decision games (TDGs; see Schmitt 1994) that provides a method for empirically assessing changes in operational decision making as a result of training. SJTs are a measurement tool commonly used as a way to assess the test taker's ability to solve problems in work-related situations by presenting a series of short decision making vignettes with potential courses of action (COAs). SJTs have high face validity and are often used to assess decision making, tacit knowledge, and practical "know-how" (Weekley & Ployhart, 2006). These tests usually start with a participant reading, viewing or listening to a brief scenario (similar to those commonly used in TDGs), after which they rate or rank a set of actions based on either perceived effectiveness or how likely they would be to perform the action. SJTs allow the assessment of decision making behavior for complex situations by presenting the participant with incomplete or ambiguous cues regarding a specific "real-world" event, and requiring the participant to rate the effectiveness of several COAs. This method is especially useful when there is not a single "right" course of action for the scenario, as is often the case in situations encountered in combat.

Developed correctly, the SJT methodology provides a unique and viable measurement tool for assessing changes in the decision making process as a result of training. In addition, the use of this method allows for the assessment of each squad member's decision making independently, without interfering with the training or the after action review (AAR) process.

Objective: The primary objective of this effort was to develop, validate, and empirically demonstrate objective, diagnostic measures for quantifying changes in operational decision making as a result of training.

The Camp Pendleton Infantry Immersion Trainer (IIT) was the chosen venue for this research because it a) simulates operational infantry tasks, often carried out under stress, b) provides a training environment in which to practice operational decision making, and c) provides an environment that can be somewhat controlled and/or adapted for meeting specific training objectives. Training scenarios used during this evaluation were conducted to meet specific training objectives and were not manipulated in anyway.

Hypothesis

The hypothesis was that squad members' effectiveness ratings of possible COAs for addressing specific decision dilemmas would more closely resemble "expert-like" decision making (reflected by SMEs' ratings of the COAs) after completing IIT scenario-based training than would ratings made prior to training.

METHOD

Development of the Assessment

Operational decision making measures must be developed with the specific training objectives and context in mind. The Camp Pendleton IIT conducts training by identifying scenarios (five scenario categories, each with five levels of complexity) to meet specific training objectives and requirements for I Marine Expeditionary Force (IMEF) units. In addition to addressing specific Training & Readiness (T&R) tasks, each scenario requires multiple decisions for successful mission completion. These decision requirements can also be viewed as training objectives. Researchers collaborated with the IIT staff to identify specific decision areas addressed within each of the scenarios (see Kobus, 2012). Kobus (2012) provides a matrix of specific decision areas identified within each scenario (category by level of complexity). Individual decision making areas were then categorized into four primary core competency areas (see Table 1) that are emphasized during training at the IIT (Kobus, 2012).

Table 1. IIT Core Competency Areas

Core Competency	Types of Decisions	Decision Areas Included
Communication	When, why, what and with whom to communicate	• Monitor and communicate with adjacent/supporting units • Communicate with squad members • Communicate clear, concise messages to higher HQ
Combat Hunter / Every Marine a Collector (EMAC)	Identifying and acting upon anomalies in the environment	• Identify individuals for questioning • Discover/locate suspicious or illegal materials • Recognize and deal with suspicious behavior • React to anomalies in a baseline • Conduct tactical questioning • Collect and disseminate information
Infantry Squad / Team Skills	Related to activities and coordination among two or more members of the team	• Enter/search homes/buildings • Plan a patrol • Prioritize multiple ongoing events • Detain personnel • Apply escalation of force procedures • Respond to contact • Establish security • Establish a cordon • Handle casualties
Cultural Interaction	Related to activities associated with the local populace	• Interact with villagers • Interact with foreign females • Deal with uncooperative/confrontational villager • Respond to request of village elder • Assess crowd behavior • Control an unruly crowd • Exhibit cultural awareness/sensitivity • Integrate/coordinate Interpreter/ANSF/other local national entities

A team of Subject Matter Experts (SMEs) worked in collaboration with researchers to develop individual assessment items for each decision area. Each item consists of a decision dilemma and five courses of action (COAs). Each decision dilemma reflected one that Marines were expected to experience during the IIT training scenarios (and in the field). SMEs then evaluated each dilemma and COA to verify face and content validity. Software was then developed to display each dilemma sequentially, along with the associated COAs. The software requires that the participant read and/or listen to each dilemma and then rate (on a 5-point scale) how effective each COA is for addressing the dilemma. Each of the five associated COAs is rated independently from the others (i.e., not rank ordered), thus providing five data points for each item. An example screen shot of an item is shown in Figure 1.

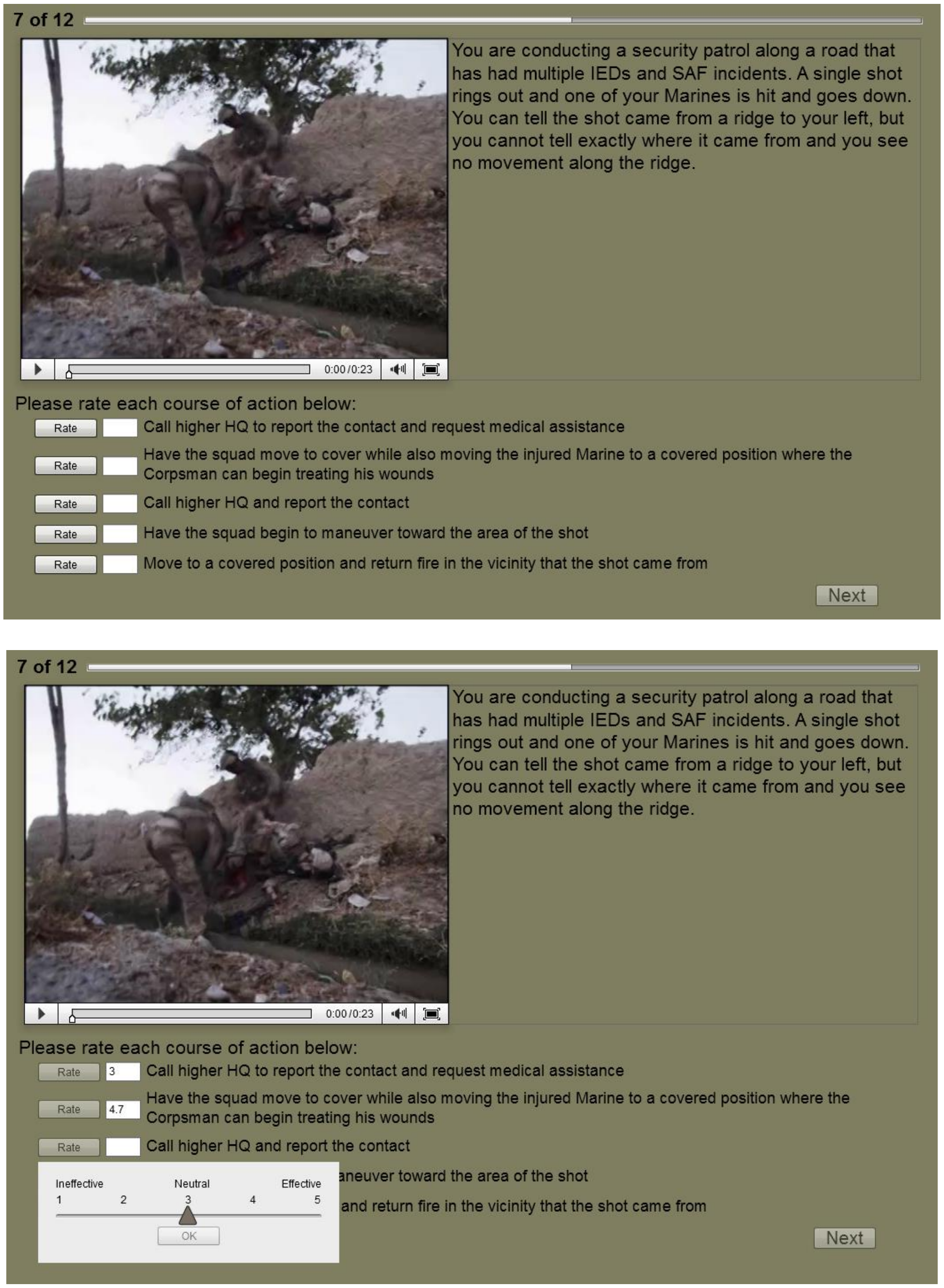

Figure 1. Screen shots of an assessment item requiring Infantry Squad / Team Skills. Top panel shows initial item presentation, bottom panel shows screen shot of participant in process of rating the third COA.

Development and Validation of the SME Database

To develop the SME response database a new set of SMEs was identified. The 11 trainers from the IIT were selected as the SMEs for the IIT assessment items. All trainers had similar backgrounds in that they were former Marines (average time in service = 14.8 yrs.) with combat experience, and were current IIT staff members (average IIT experience = 3 yrs.). Although trainers shared similar experiences it was important to validate that their decision making was in concert. Thus, all trainers independently completed each of the assessment items by rating the effectiveness of each of the COAs for addressing the associated dilemma. An Intraclass Correlation Coefficient (ICC) was determined for the ratings provided for each item to evaluate level of consistency (agreement) in responses across the 11 SMEs. If ICC values were less than .6 (the minimum for very strong agreement) the item was rejected and a new item was developed and re-evaluated by the SMEs. This iterative process continued until an ICC $\geq$ .6 across the SMEs was reached for all items. Once this validation process was finalized, SME ratings for each COA were averaged. These averaged SME values served as the "expert" response database.

Participants

Sixty-seven infantrymen participated in three squad-level training scenarios at the IIT at Camp Pendleton, CA. The primary targeted training objectives of the unit were Infantry Squad / Team Skills. Each member of the unit completed two equivalent versions of the assessment (pre-training / post-training). One participant completed the assessments in a time period too short (< 5 minutes) for him to have read or listened to each item; his data were dropped from further analyses. Therefore, all analyses were conducted using the results of 66 infantrymen.

Pre- and Post-training Assessments

All squad member trainees completed an assessment consisting of 12 items related to the unit's specific training objectives prior to the first scenario. Each squad then participated in three training scenarios. Immediately after the hot-wash of the third scenario, trainees completed a similar set of 12 assessment items. Difficulty of these items (as determined by SMEs) was comparable to the pre-training set. Trainees read / listened to a scenario-based dilemma and then rated the effectiveness of each of five potential COAs. Ratings were made on a continuous five-point scale (1 = ineffective to 5 = effective). Each COA was rated independent of the others (i.e., not rank ordered), thus multiple COAs could be given the same rating. Once all ratings were made for each of the COAs, participants could proceed to the next item.

Analyses

All data were collected on laptops located at the IIT. All scoring and analysis was conducted off-line. Note that each of these analyses is based upon a different number and subset of items, due to the items that were selected to reflect the unit's specific training objectives and emphasis; Infantry Squad / Team Skills = 8 items, Cultural Interaction = 1 item, Combat Hunter / EMAC = 1 item, and Communication = 2 items.

Standardized Average Difference

This analysis utilized the full set of SME and trainee ratings for each of the five COAs for each of the 12 decision dilemmas in the pre- and post-training assessments. Thus, the analysis is based upon 60 data points for each trainee for the pre-training assessment and 60 data points for each trainee for the post-training assessment. In order to control for potential restriction of range issues between individuals using the rating scale, we conducted a z-score transformation of all SME and trainee ratings to standardize the values. Differences between the SME database and the trainee ratings for each COA for a given dilemma were determined and values were squared. The average squared difference from expert-like performance for a specific item was then determined for each trainee, for each pre- and post-training item. The delta between the pre-training and post-training difference scores can be considered a measure of training effectiveness. A lower score (smaller difference from the SME database) was indicative of more expert-like decision making.

PRELIMINARY RESULTS

A two-tailed paired *t*-test was conducted on standardized averaged deviation values of trainees' responses compared to the SME database for pre- and post-training assessment items. Results indicate that as a whole, the unit rated COAs more like experts after training than before training, $t(64) = 5.87$, $p < .01$. Figure 2 displays the mean difference values from expert level for pre/post test scores for all participants.

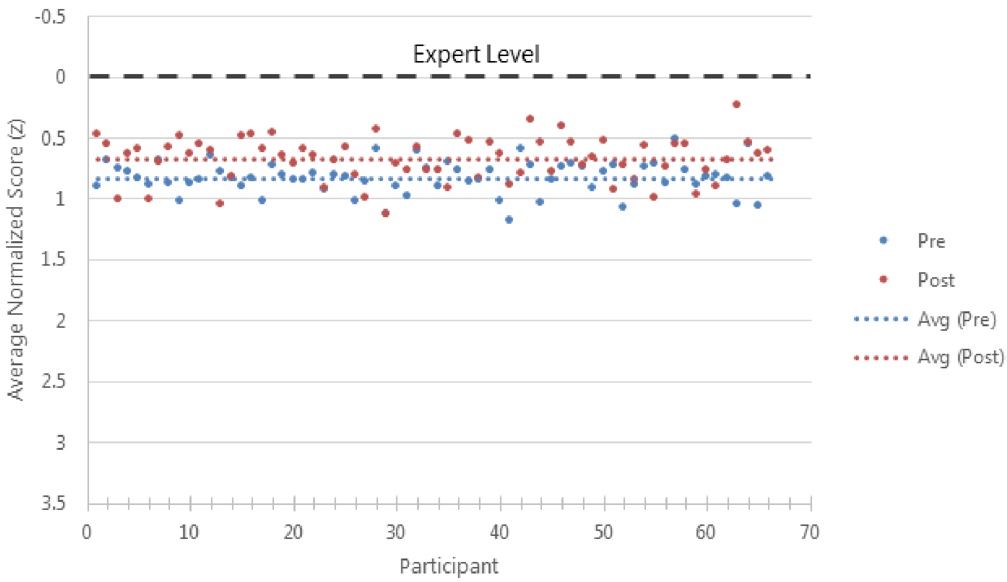

Figure 2. Overall results showing pre- versus post-training difference.

Core Competency Areas

Individual paired *t*-test analyses were also conducted based upon items related to each of the four core competency areas emphasized during training at the IIT (Communication, Combat Hunter / EMAC, Infantry Squad / Team Skills, Cultural Interaction).

Infantry Squad / Team Skills

The primary training objective for this unit was Infantry Squad / Team Skills. The majority of the assessment items (8 of 12) were therefore related to these skills. A paired *t*-test was conducted on standardized averaged deviation values, compared to the SME database, for pre- and post-training assessment items. Results indicate that as a whole the unit rated COAs more like experts after training than before training, $t(64) = 3.95$, $p < .001$. Figure 3 displays results showing pre- and post-training levels for Infantry Squad / Team Skill decisions for each trainee.

Cultural Interaction

Based upon the scenarios used for training, one assessment item was related to decisions involving Cultural Interaction. A paired *t*-test was conducted on standardized averaged deviation values, compared to the SME database, for pre- and post-training assessment items. Results indicate that as a whole the unit rated COAs more like experts after training than before training, $t(64) = 3.22$, $p = .01$. Figure 4 displays results showing pre- and post-test levels for Cultural Interaction decisions for each trainee.

Communication

During the training scenarios, Marines were required to make several decisions pertaining to internal and external communication; two items on the assessment addressed Communication-related decisions. A paired *t*-test was conducted on standardized averaged deviation values, compared to the SME database, for pre- and post-training assessment items. Results indicate the unit's performance changed significantly in the opposite direction from expert-like responses post training, $t(64) = 7.86$, $p < .01$. Figure 5 displays results showing pre- and post-test levels for Communication decisions for each trainee.

Combat Hunter / Every Marine a Collector (EMAC)

One assessment item was related to Combat Hunter / EMAC decisisons. A paired *t*-test was conducted on standardized averaged deviation values, compared to the SME database, for pre- and post-training assessment items. Results indicate that as a whole the unit rated COAs more like experts after training than before training, $t(64) =$

10.67, p = .01. Figure 6 displays results showing pre- and post-test levels for Combat Hunter / EMAC decisions for each trainee.

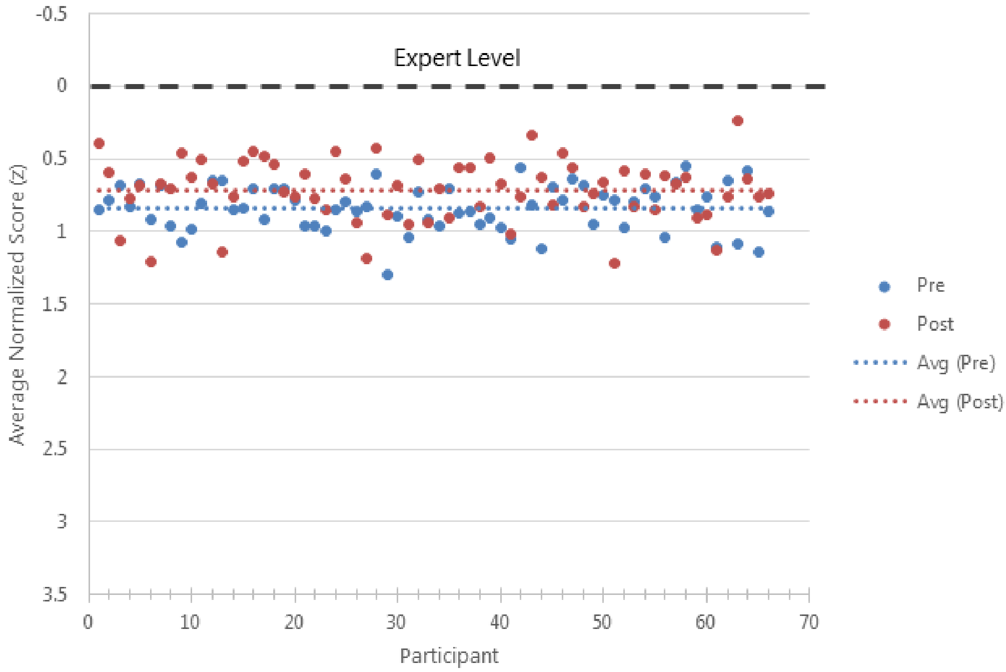

Figure 3. Pre- and post-training results for all participants for items based upon Infantry Squad/Team Skills.

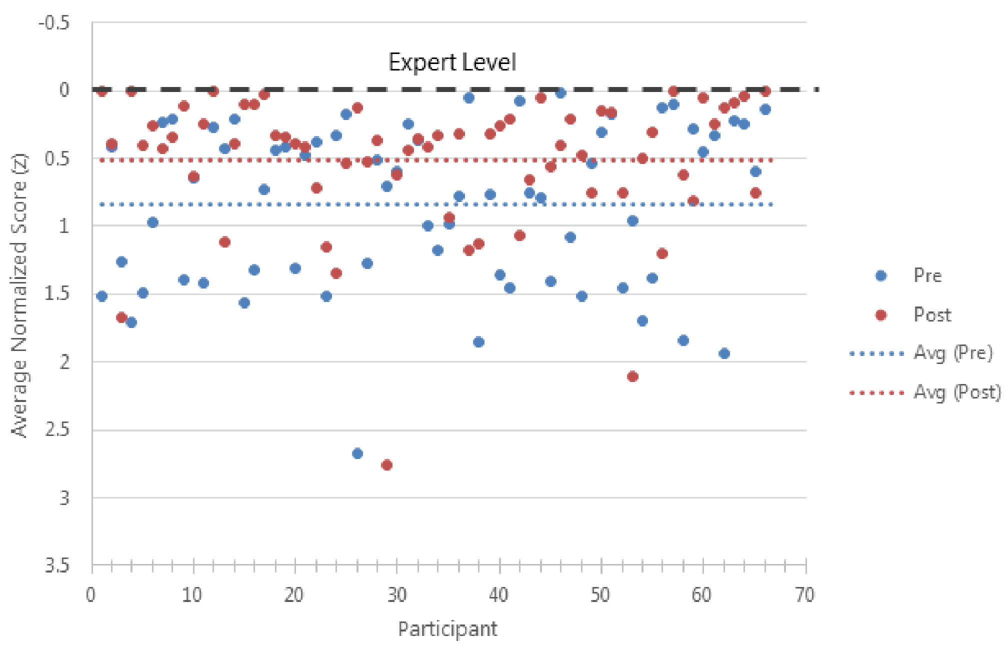

Figure 4. Pre- and post-training results for all participants for the item based upon Cultural Interaction.

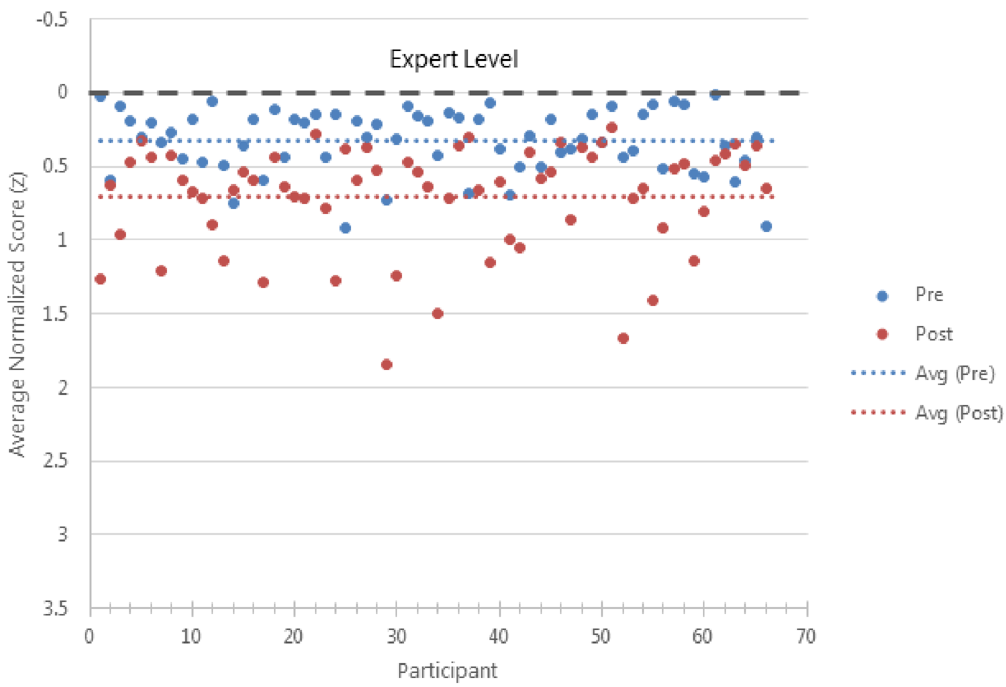

Figure 5. Pre- and post-training results for all participants based upon items related to Communication.

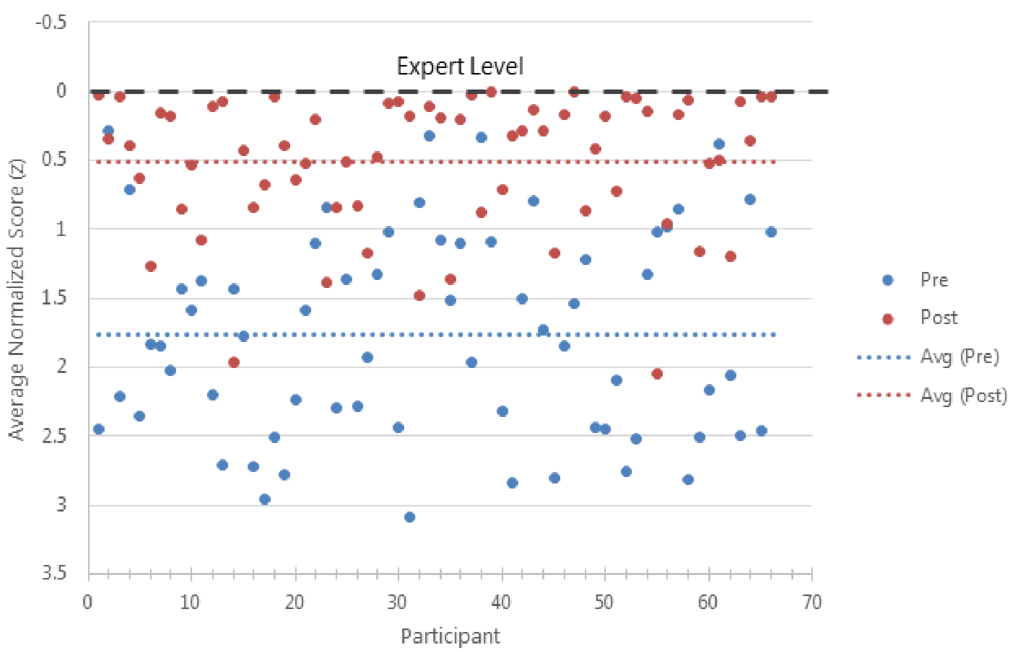

Figure 6. Pre- and post-training results for all participants based upon the item related to Combat Hunter/EMAC skills.

SUMMARY

Although preliminary, the results from this unit suggest the utility of this method for assessing changes in operational decision making as a result of training. While results demonstrated statistically significant pre- to post-training changes in decision making, the level of detail at which the results can be interpreted or further explored for diagnostic purposes requires further research.

The analysis compared each trainee's COA ratings for each dilemma, to the SME database. Results demonstrated that overall decision making became more expert-like (COA ratings more similar to those of SMEs) post-training. Additional analyses were conducted for the primary core competency areas commonly addressed in IIT training scenarios. Although training for this unit was conducted to specifically practice Infantry Squad / Team Skills, operational decision making statistically improved for three of the four competency areas (Infantry Squad / Team Skills, Combat Hunter / EMAC, and Cultural Interaction). Changes found for Combat Hunter / EMAC and Cultural Interaction competencies suggest that although not emphasized, exposure to these areas in the course of IIT training may facilitate operational decision making performance. In contrast, performance was significantly less expert-like for the Communication competency area post-training. While this performance change was not in the desired direction, the result is consistent with the finding for Communication-related decisions reported during Spiral 2 of the Future Immersive Training Environment (FITE) Joint Capability Technology Demonstration (JCTD) (Kobus, Kobus, Ostertag, Kelly, & Palmer, 2010b). Further validation and identification of possible explanations of this performance change will require additional data collection.

Of specific note is the fact that the assessment results indicate that training was effective for the targeted skill-set (Infantry Squad / Team Skills); there was a statistically significant improvement in operational decision making for decisions that addressed this training objective. In fact, over 80% of the trainees demonstrated pre- to post-training improvement. Statistical analyses were only conducted at the unit or collective level, but could also be conducted at the individual level. The diagnostic capability of this method of analysis is one area that needs to be further explored.

The results described in this report are a first step in demonstrating the potential of one method for assessing the effectiveness of decision making training across a variety of competency areas. Although promising, the results need to be corroborated by additional data. Equally important is systematic research to determine how the capabilities of this method, in combination with other measures or metrics, can be fully exploited to best evaluate, assess, and diagnose operational decision making performance in a variety of venues. Full development of diagnostic capabilities will provide valuable tools to identify best practices for decision making training. Some of the many outstanding research questions to be addressed are identified below. The first three relate specifically to development, validation, and analysis of assessments. The remaining items are crucial for efficient, effective decision making training, and can only be addressed by objective measures of changes in decision making performance.

1. How many assessment items are needed to ensure construct validity?
2. What is the correlation between trainer subjective evaluation and objective assessment results for individual and collective training?
3. Which assessment analysis methods provide the most useful information for diagnostic and intervention purposes?
4. How much training (days, number of scenarios) is required to impact decision making?
5. How do prerequisites (or failure to complete them) impact training effectiveness?
6. What effect does context have on operational decision making?
7. Which individuals (squad leaders, team leaders, corpsman, etc.) are most / least affected by training?
8. To what extent does performance during decision making training transfer to the operational environment?
9. What is the sustainability of changes in decision making?

Clearly a significant amount of data will be required to begin addressing these questions. With the growing emphasis on operational decision making training objectives, it is imperative to answer these questions in order ensure the most efficient and effective training is provided to our military.

ACKNOWLEDGEMENTS

This effort is part of the Small-unit Training for Adaptability and Resilience in Decision Making (STAR-DM) program sponsored by Dr. Peter Squire, ONR Code 30; the assessment methods described in this paper will be used to evaluate the effectiveness of new training packages developed under the STAR-DM program.

REFERENCES

Kobus, D. A., Kobus, J., & Ostertag, J., Kelly, M., & Palmer, E. D., (2010b). *Assessing changes in tactical decision making: FITE-JCTD Spiral 2 Operational Demonstration (Tech Report 10-13)*. Pacific Science & Engineering Group, San Diego, CA.

Kobus, J. M. (2012). Training Needs Analysis (TNA) Report: The Infantry Immersion Trainer (IIT) Camp Pendleton, CA, *PSE Technical Report: 12-02.*

Krulak, C. C. (1999). "The Strategic Corporal: Leadership in the Three Block War," *Marines Magazine* (January). Marine Corps Training and Readiness Manual.

Weekley, J.A. & Ployhart, R.E. (Eds.). (2006). *Situational Judgment Tests: Theory, Measurement, and Application*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Whetzel, D.L. & McDaniel, M.A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19,* 188-202.

Schmitt, J.F. (1994). *Mastering tactics: A tactical decision games workbook*. Marine Corps Association: Quantico, VA.