2014

An Analytical Method for Assessing the Effectiveness of Human in the Loop Simulation Environments: A work in progress

Hodges, Glenn A.

An Analytical Method for Assessing the Effectiveness of Human in the Loop Simulation Environments: A work in progress

Glenn A. Hodges, LTC, USA
Naval Postgraduate School
gahodges@nps.edu

Rudolph Darken, PhD.
Naval Postgraduate School
darken@nps.edu

Michael McCauley, PhD.
Naval Postgraduate School
memccaul@nps.edu

Keywords: Assessment, Affordances, Human Abilities, Training Effectiveness, TADSS

Abstract

This paper discusses the development of the Integrated Training Environment Assessment Methodology (ITEAM). ITEAM is proposed as a way to assess integrated human in the loop (HITL) simulation training environment effectiveness (i.e. to determine how well the simulation tools support the deliberate practice of specific skills) at the human ability level. A work in progress, ITEAM has been used to reassess two previously evaluated military simulation environments -- the game Full Spectrum Command (2004) and the Engagement Skills Trainer Heavy Weapons Variant (2008). ITEAM results from both studies closely resemble and often match those concluded in the original empirical training effectiveness analysis (TEA) studies. A third study is currently underway to solidify ITEAM validity.

1. INTRODUCTION

Research into simulation software effectiveness for various analytical and manufacturing efforts has been ongoing for over two decades [Banks 1991; Jadhav and Sonar 2009; Nikoukaran et al. 1999]. Similar efforts in determining the effectiveness of human in the loop (HITL) simulation for training have been a topic of interest for researchers for about as long [Salas and Cannon-Bowers 2001]. Recently, the topic of training simulation effectiveness has attracted renewed interest within the Department of Defense (DoD) as budgets for training and training systems have been cut [GAO 2013]. Historically, research conducted on HITL simulations has involved subject matter experts (SME) and has used the empirical analysis of training transfer to attempt to answer questions about simulation effectiveness [Baldwin and Ford 1988; Ford and Weissbein 1997; Blume et al. 2009]. However, an inability to consistently define and apply useful metrics to evaluate the Modeling and Simulation (M&S) tools used to support training has resulted in few answers and even more questions. Within the military, simulations have many uses. This research effort is focused solely on HITL simulation and the integrated training environments (ITEs) used to develop and practice the skills needed to effectively execute warfare. For our purposes an ITE is defined as *any training environment that includes Live, Virtual, Constructive or Game-based training aids, devices, simulators, or simulations (TADSS) alone or in combination, that support the deliberate practice of skills for defined mission tasks*. The ideas for developing ITEAM came from the study of the literature pertaining to the assessment of simulations within and outside of the training domain, the literature related to training transfer and the systems engineering process.

2. RELATED WORK

Several meta-analytic reviews investigating the state of the art in analytical assessment of HITL simulation have been conducted [Muckler and Finley 1994a, 1994b; Simpson 1995; Tufano and Evans 1982; Wheaton et al. 1976]. These reviews indicate that since about 1960, various forms of training and cost effectiveness models and programs have been developed and implemented. While many of the efforts have merit and some have demonstrated their usefulness in various contexts, few if any of them are currently in use today despite their advantages. The reasons provided for this situation are that the models are too complex and costly to run and maintain, they require extensive data necessarily obtained by empirical means and they require too much time to set up. Efforts to evaluate simulations outside of the training domain have resulted in an overt focus on technical features, knowledge of which is then applied to support simulation acquisition decisions. The evaluation of simulations in support of analytical activities in areas such as manufacturing and operations research have mainly focused on developing methodologies to support the selection of simulation software packages by investigating and rating the strengths and weaknesses of what can be considered ‘surface features’ [Banks 1991; Nikoukaran et al. 1998; Nikoukaran and Paul 1999; Hlupic and Mann 1995; Law and Haider 1989]. While acknowledging that both performance and technical features are important, we recognize that both of these perspectives have not resulted in an effective, long-term and sustainable method for assessing the utility of integrated training environments.

3. INTEGRATED TRAINING ENVIRONMENT ASSESSMENT METHODOLOGY (ITEAM)

The integrated training environment assessment methodology (ITEAM) is a human-centric systems engineering approach for assessing ITEs. Assessment and scoring of an ITE using ITEAM is conducted manually. Future work includes the automation of both the assessment process and ITE scoring. Considering the pieces of a training program (e.g. technology, requirements and humans) it has been established that computer technology evolves the quickest (i.e. Moore's Law). Requirements determination occurs at a much more deliberate pace. Human beings evolve the slowest yet their evolutionary stability is mostly ignored during HITL simulation development in favor of an emphasis on technology. ITEAM suggests that the human should be the bedrock of ITE development rather than the technology. Useful evaluation of an ITE is dependent upon a clear articulation of the need for the training environment, an understanding of what the system(s) will provide/support, and a clear statement of the desired measurable performance outcomes from users of the system. Training program context is important. Training programs include (TADSS) as well as other instructional materials and plans. Understanding where and when a system will be used with respect to the training program helps in determining the simulation functional allocation between the user and the system. While we believe that ITEAM may benefit system developers during their initial development processes, the main emphasis of this research has focused on using ITEAM to assess the utility of existing ITE's. ITEAM may be viewed as a set of three main processes. Each process contains multiple sub-processes (Figure 1). All of the processes are iterative in nature and steps may be abbreviated or skipped depending on the time available and level of detail required.

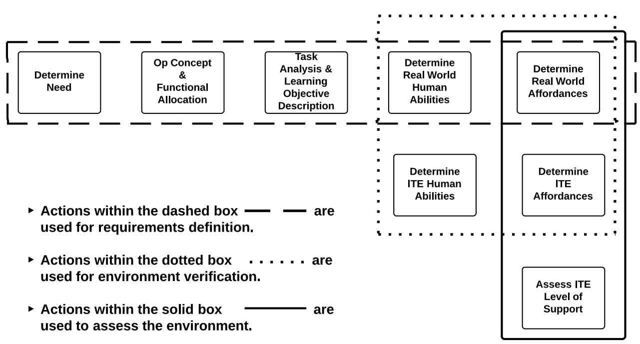

Figure 1. ITEAM main and sub-processes

3.1. Defining the requirements for the ITE

Proper problem description and analysis are critical in the simulation development process. ITEAM groups the activities of determining the need for the ITE, how it will be used, which functions will be performed by the ITE and the human, description of the tasks to be executed during training and the desired learning outcomes within the boundary of requirements definition. Also included is a list of real world human abilities and real world affordance requirements that are necessary to accomplish the training tasks. Human abilities (HA) used to describe work are leveraged here as a tool to illuminate the critical aspects (environmental affordances) required of an ITE [Fleishman and Quaintance 1984]. Affordance theory and affordances, from ecological psychology, are used to describe the attributes of the ITE that are necessary to support the execution of the desired training [Gibson 1986].

3.2. Verifying that the ITE has what it needs

Verification is **"**the process of determining that a model or simulation implementation and its associated data accurately represent the developer's conceptual description and specifications" [DOD 2009]. The sub-processes of ITEAM considered to be useful for verification consist of compiling the identified real world and system-supported human abilities as well as the real world and system-provided affordances. During this phase, the evaluator uses the task analysis to determine the real world human abilities and affordance requirements associated with the tasks to be trained. Then, the ITE is investigated to determine what human abilities it supports and what affordances are available. Comparison of these items provides the basis for an initial judgment on whether or not the ITE will support the execution of the desired training.

3.3. Assessing the ITE for a training purpose

The final phase of ITEAM assesses an ITE's ability to support desired training through a logical process of comparing real world affordance requirements with ITE affordance resources. The quantification of resources provides the customer (user) with an estimate of the level of support that the ITE provides. ITE scoring is based on the evaluator's judgment on the absence or presence of specified affordances using the scale seen in Figure 2 below.

Scale Definition

5- Excellent – the ITE contains all but a few (90-100%) of the affordances determined during the analysis
4- Very Good – the ITE contains a significant portion (70-89%) of the affordances determined during the analysis
3- Good – the ITE contains a good portion (50-69%) of the affordances determined during the analysis
2- Fair – the ITE contains some (25-49%) of the affordances determined during the analysis
1- Poor – the ITE contains very few (0-24%) of the affordances determined during the analysis

Figure 2. ITEAM assessment scale

Two studies have been conducted to determine the validity of the ITEAM approach. Both studies utilized previously conducted training effectiveness analyses (TEA) studies that attempted to empirically evaluate an ITE. The remainder of this paper discusses our observations of the assessment of the Engagement Skills Trainer (EST) 2000 Heavy Weapons (HW) simulator.

4. ITEAM ASSESSMENT OF THE EST 2000 ITE

The EST 2000 heavy weapons simulator is primarily used to practice the engagement of targets using the M2 heavy barrel machine gun and the MK19 grenade launcher. [Hughes and Nau 2008]. The Training and Doctrine Command Analysis Center (TRAC) White Sands conducted a training effectiveness analysis (TEA) of the EST 2000 HW simulator in March of 2008. The focus of the TEA was to determine if using the EST could mitigate the impact of ammunition shortages on soldier proficiency for two machine guns: the M2 .50 caliber and MK19 .40mm grenade machine guns (MG). [Hughes and Nau 2008]. The EST 2000 is a laser-based unit/institutional, indoor, multi-purpose weapons trainer that displays targets, terrain and weapons' effects in a real-time, 3 dimensional presentation on a screen that is 26 feet 3 inches from the firer [Hughes and Nau 2008]. Qualification and crew engagement skills practice are both possible with the EST 2000. The objectives listed in the 2008 TEA for the EST were to determine the effect of substituting EST 2000 training for live fire training for both the M2 and MK19. There were two Essential Elements of Analysis (EEA) for the study. The first was to determine if the quality of training using the EST 2000 differed from live-fire training in terms of the impact on the end of course weapons' proficiency test. The second was to determine if EST 2000 training could effectively substitute for live-fire training in terms of its impact on target engagement proficiency with the M2 and MK19 machine guns.

Review and reflection of the objectives, stated EEAs, scope and focus of the TEA lead us to believe that the original TEA was improperly focused to determine if deliberate practice with the EST could reduce or eliminate the need to conduct deliberate practice using live ammunition, which was the essence of the objectives and EEAs of the original TEA study. The EST 2000 does not explicitly train soldiers; rather, it is an ITE that allows for the deliberate practice of skills such as the four fundamentals of marksmanship (steady position, proper sight picture, proper breathing and trigger squeeze), target identification, range determination and target engagement. Recognizing this fact, our evaluation focused on whether or not the EST environment contained the necessary affordances to facilitate task execution and deliberate practice of the skills involved in engaging targets using both the M2 and MK19 weapon systems.

4.1. Analysis and verification of the EST 2000 ITE

Our analysis began with a review of the weapons manuals for the MK19 and M2 as well as a review of the day qualification shooting standards for both weapons (i.e. Department of the Army (DA) Forms 7518-R and 7448-R respectively). As previously mentioned, ITEAM is flexible; processes of the methodology may be abbreviated or skipped depending on the information available. For this study, we recognized that it was unnecessary to conduct analysis on the first two sub-processes of the requirements determination process. The EST is an established virtual simulation capability whose scope we believe is clear. We began ITEAM by conducting a task analysis (TA). The two high-level tasks initiating this TA were <u>Engage targets with the MK19 in accordance with (IAW) DA form 7518-R</u> and <u>Engage targets with the M2 MG IAW DA Form 7449-R.</u> The supporting tasks and subtasks were described using information from the weapons manuals for both weapon systems. The TA encompassed all of the activities necessary to place the weapons into operation, from loading thru clearing, which goes well beyond simply engaging targets with the weapons. We believe that these additional tasks play a critical role in the operation of the weapons and should therefore be included. Figure 3 displays a representative sample of the TA conducted.

Sub-task: Charge MK19

1. Grasp the charging handles with the palms facing down.
2. Press the charger handle locks in, rotate the handles down and pull them sharply to the rear
3. Return the charger handles forward to their original upright position after locking the bolt to the rear
4. Place safety selector switch on fire and press trigger
5. Repeat step 2
6. Place safety selector switch to safe
7. Repeat step 3

Figure 3. Example of partial TA for MK19

Next, real world human abilities (HA) identified as being necessary to execute the skills and tasks outlined were identified and listed. HAs are categorized within the groupings of cognitive, sensory, psychomotor and physical. Abilities were assigned to the tasks based on the definitions given in [Fleishman and Quaintance 1984] and our experience using the weapons. Figure 4 provides a sampling of the human ability inventory from this study.

COGNITIVE: Problem Sensitivity; Information Ordering; Spatial Orientation; Memorization
SENSORY: Night Vision; Peripheral Vision; Glare Sensitivity; Depth Perception; Near Vision
PSYCHOMOTOR: Arm-Hand Steadiness; Manual Dexterity; Finger Dexterity; Multi-limb Coordination
PHYSICAL: Static Strength; Trunk Strength; Extent Flexibility

Figure 4. Human abilities (HA) for charging the MK19

Upon completion of the inventory listing of human abilities (HAs), we focused our attention on describing the real world affordance requirements needed within the ITE in order for training to occur. From the tasks listed in figure 3 and the HA inventory from Figure 4, the following affordances were identified (Figure 5).

- Weapon resemblance that is an appropriate weight with 2 charging handles one on either side of the weapon receiver
- Charging handles provide haptic and proprioceptive feedback and range of motion similar to those of the MK19
- Weapon bolt resemblance that moves within the receiver that can be locked to the rear of the receiver
- Resemblance of a Fire/Safe lever on the backplate of the weapon resemblance that provides haptic, proprioceptive and visual feedback of location (setting)
- A stand or mount that holds the weapon resemblance in a proper position for loading, unloading, firing and clearing activities
- A butterfly type trigger resemblance located between two handles on the weapon receiver resemblance backplate

Figure 5. Example real world affordance requirements

The next step in our assessment of the EST 2000 was to investigate the actual simulator. This investigation occurred at Fort Hunter Liggett in California. Figure 6 depicts the HA inventory of the EST 2000 for the tasks shown in Figure 3, "**charge MK19**".

COGNITIVE: Oral Expression; Oral Comprehension; Problem Sensitivity; Information Ordering; Spatial Orientation; Memorization
SENSORY: Night Vision; Peripheral Vision; Depth Perception; Near Vision
PSYCHOMOTOR: Arm-Hand Steadiness; Manual Dexterity; Finger Dexterity; Multi-limb Coordination
PHYSICAL: Static Strength; Trunk Strength; Extent Flexibility

Figure 6. Human ability (HA) inventory of the EST 2000

Interestingly, we identified that the simulator lacked support for glare sensitivity but it supported oral expression and comprehension that were not determined to be necessary. While not directly noted as necessary for task execution in the real world, oral expression and comprehension are important in the training environment where it may be necessary to practice team communication skills when employing the weapons. The final step of this stage consisted of cataloging the affordance resources available within the ITE so that they could be compared to the identified affordance requirements. For the example sub-task of "**charge MK19,**" it was determined that the EST 2000 possessed all of the necessary affordances.

4.2. Assessment and scoring of the EST 2000 ITE

ITE assessment results in a comparison and scoring of affordance resources and affordance requirements. This analysis results in qualitative and quantitative information that informs potential ITE users of the level of support they should expect from the ITE based on their training objectives and desired training tasks. As mentioned earlier, the scoring process is currently a manual one conducted by the evaluator. The following paragraphs provide the reader with a brief description of how scoring is conducted and the results from our assessment of the EST 2000 ITE.

4.3. The general approach to scoring an ITE

Each high-level task is broken down into manageable sub-tasks where affordances are determined and catalogued. Sub-task affordance scores are determined for the ITE by averaging the number of affordances present by the number required. This percentage qualitatively represents how well the ITE supports the deliberate practice of the skills necessary to execute the sub-task. Scores range from 1-Poor to 5-Excellent as seen in Figure 2. For the example in Figure 3, we assessed the ITE as having all six of the required affordances seen in Figure 5, resulting in a score of 5-Excellent for this particular subtask. If in this example we had determined that only two of the six sub-task affordances were present (33%) a score of 2-Fair would have been assigned. In cases where sub-tasks occur more than once in a TA, the first value obtained is used throughout. This eliminates unnecessary scoring work.

If a high-level task consists of multiple sub-tasks that have been previously evaluated and scored in other parts of the assessment, the scores from those previous evaluations are used for the high-level task under current investigation. When a condition arises where a high-level task consists of unique sub-tasks as well as sub-tasks previously scored, a 4-step process ensues to determine the numerical value for the high-level task. First, any previous affordance evaluation is treated as a single affordance that is present. Next, a determination of unique affordance (not previously accounted for) presence or absence is conducted. These two steps result in a temporary rating assignment for the sub-task. This temporary value is then combined with the scores from the previous sub-task affordance evaluations. A ratio of the total value of the sub-task scores divided by the number of subtasks is then obtained. This process provides a numerical score between 1 and 5, (see Figure 2) which is assigned as the score for the high-level task. Scores containing 0.5 or less are rounded down to the nearest whole number whereas scores containing 0.6 and above are rounded up.

For TAs and assessments of complex ITEs, it is possible to become mired in a situation where the repetition of the same tasks and their nesting results in a situation that is unproductive. In those instances we have opted to halt our decomposition, for scoring purposes, at the second nested level to streamline the scoring process. This decision results in a slightly different numerical value for the high-level task than would be otherwise calculated but we have determined that the differences do not make a significant impact on the final ITE assessment scores.

4.4. Results of the ITEAM EST ITE Study

The results of the ITEAM evaluation of the two high-level tasks for engaging targets were a 4.0 very good and 4.43 very good respectively. Weaknesses noted in the original TEA, also captured with ITEAM, were the amount of visual feedback provided by the system on the trajectory and impact of rounds and the training environment's inability to replicate natural factors such as heat, cold, wind, rain, etc. Additionally noted was the EST's inability to fully support clearing procedures. Overall, the ITEAM analysis predicted that the EST 2000 would provide very good support to the deliberate practice of the skills necessary to engage targets using both weapons systems. This prediction was supported by the results of the qualification scores that found no statistical difference between the two groups, those who solely used the EST to practice target engagement and those who solely used live fire practice [Hughes and Nau 2008]. Results of soldier opinion surveys also support the ITEAM assessment. Soldiers and instructors were asked to rate the quality of their prequalification training as Excellent, Good, Adequate, Inadequate, Poor and Very Poor. For the M2, 50 percent of respondents rated the EST as Good and an additional 13 percent rated it as Excellent. For the MK19, 41 percent rated it as Excellent and 36 percent rated it as Good.

5. CONCLUSIONS

HITL simulations, consisting of numerous training aids, devices, simulators and simulations (TADSS) are expensive and are relied upon daily to train personnel in many domains. The use of analytical methodologies to assess the effectiveness of these various types of simulations and simulation packages is desirable given the amount of money spent each year obtaining them [Salas and Cannon-Bowers 2001]. ITEAM is novel in its perspective and approach. Instead of focusing on the features of each of the various simulations, simulators and devices individually, ITEAM deliberately looks at the integrated training environment as a whole using human abilities as a lens. Using the human abilities of the trainee, rather than the technology to shape the effectiveness assessment, ITEAM provides insights to both users and developers alike. If used early in the design process, ITEAM may demonstrate value, as a way to define simulation requirements that are more aligned with stakeholder needs and desires. Used during and after initial development, ITEAM may help in highlighting deficiencies in original designs or stakeholder desires. Finally, as demonstrated in this paper, when used as an assessment tool, ITEAM is valuable in providing answers to questions of integrated simulation environment utility. Identified future work includes the automation of the assessment process and scoring of affordances as well as refinement to the methodology to sharpen the level of detail provided by the assessment results.

REFERENCES

Baldwin, T., Ford, J. (1988). "Transfer of training: A review and directions for future research", Personnel psychology, 41(1), (p. 45).

Banks, J. (1991). "Selecting simulation software", In G. Clark, B. Nelson, & W. Kelton (Eds.), Proceedings of the 1991 Winter Simulation Conference (p. 15–20). IEEE Computer Society Press.

Blume, B., Ford, J., Baldwin, T., Huang, J. (2009). "Transfer of Training: A Meta-Analytic Review", Journal of Management, 36(4), (p. 1065–1105).

DOD. (2009). *DoDI 5000-61: Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A) Instruction.* Washington, D.C.: USD(AT&L).

Fleishman, E. A., & Quaintance, M. K. (1984). *Taxonomies of Human Performance: The Description of Human Tasks* (1st ed., p. 514). Orlando: Academic Press.

Ford, J., Weissbein, D. (1997). "Transfer of training: An Updated Review and Analysis", Performance Improvement Quarterly, 10(2), (p. 22–41).

GAO. (2013). "ARMY AND MARINE CORPS TRAINING: Better Performance and Cost Data Needed to More Fully Assess Simulation-Based Efforts", (p. 33). Washington, D.C.

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception* (p. 127–143). Hillsdale: Lawrence Erlbaum Associates, Inc.

Hlupic, V., Mann, A. (1995). “SimSelect: a system for simulation software selection”, In C. Alexopoulos, K. Kang, W. R. Lilegdon, & D. Goldsman (Eds.), Proceedings of the 1995 Winter Simulation Conference (p. 720–727).

Hughes, C., Nau, K. (2008). “Engagement Skills Trainer (EST) 2000 Heavy Weapons Training Effectiveness Analysis (TEA)”, (Vol. 7, p. 56). White Sands Missile Range.

Jadhav, A., Sonar, R. (2009). “Evaluating and selecting software packages: A review”, Information and Software Technology, 51(3), (p. 555–563).

Law, A., Haider, S. (1989). “Selecting simulation software for manufacturing applications”, In E. MacNair, P. Heidelberger, & K. Musselman (Eds.), Proceedings of the 1989 Winter Simulations Conference (p. 29–32).

Muckler, F., Finley, D. L. (1994a). “Applying Training System Estimation Models to Army Training Volume 1. Analysis of the Literature”, Alexandria.

Muckler, F., Finley, D. (1994b). “Applying Training System Estimation Models to Army Training, Volume 2. An Annotated Bibliography 1970-1990”, Alexandria.

Nikoukaran, J., Hlupic, V., Paul, R. (1998). “Criteria for simulation software evaluation”, In D. Medeiros, E. Watson, J. Carson, & M. Manivannan (Eds.), 30th conference on Winter simulation (p. 399–406). Los Alamitos, CA: IEEE Computer Society Press.

Nikoukaran, J., Hlupic, V., Paul, R. (1999). “A hierarchical framework for evaluating simulation software”, Simulation Practice and Theory, 7(3), (p. 219–231).

Nikoukaran, J., Paul, R. (1999). “Software selection for simulation in manufacturing: a review”, Simulation practice and theory, 7(1), (p.1–14).

Salas, E., Cannon-Bowers, J. (2001). “THE SCIENCE OF TRAINING : A Decade of Progress”, Annual review of psychology, (52), (p. 471–499).

Simpson, H. (1995). “Cost-Effectiveness Analysis of Training in the Department of Defense”, (p. 90). Monterey.

Tufano, D., Evans, R. (1982). “The Prediction of Training Device Effectiveness : A Review of Army Models”, (p. 64). Alexandria.

Wheaton, G., Rose, A., Fingerman, P., Korotkin, A., Holding, D. (1976). “Evaluation of the effectiveness of training devices: Literature review and preliminary model”, (p. 138). Alexandria.

AUTHOR BIOGRAPHIES

LTC GLENN A. HODGES U.S.A. is a Ph.D. Candidate in the Modeling Virtual Environments and Simulation (MOVES) curriculum at the Naval Postgraduate School. During his career he has served in a variety of Command and Staff positions. Most recently, he served as the Proponent Officer for the Army’s personnel Functional Area 57, Simulation Operations, at the Center for Army Analysis (CAA). LTC Hodges holds a BS in Business Administration from Old Dominion University and MS in MOVES from NPS. His research interests include the application of systems engineering processes to the development of modeling and simulation and the use of M&S to improve human performance.

DR. RUDOLPH DARKEN is Professor of Computer Science at the Naval Postgraduate School in Monterey, California. He has served as the Director of the Modeling, Virtual Environments, and Simulation (MOVES) Institute and as the Academic Program Chair and Director of Research for the Center for Homeland Defense and Security (CHDS). His research has focused on spatial cognition, virtual environments, game development and agile software development techniques. He has taught executive seminars on the role of disruptive technologies in defense acquisition and he currently teaches graduate courses in simulation and training, innovation and business practices in modeling and simulation, and critical infrastructure protection. He received his D.Sc. degree in Computer Science from The George Washington University in 1995.

DR. MICHAEL McCAULEY has been a Research Professor at the Naval Postgraduate School since 2002. He teaches courses in Human Factors, Anthropometry, and Simulation & Training systems. His recent research has been in the areas of prototype development of a Landing Signal Officer database for carrier landings, ship motion effects on human performance, and visual detection of IEDs. For over 40 years, he has been active in applied research on human performance. After serving in the Navy as an Aerospace Experimental Psychologist, Prof McCauley received his Ph.D. in Experimental Psychology (Perception & Performance) from the University of California Santa Barbara in 1979.