Adaptive Instructor Operating Systems: Design to Decrease Instructor Workload and Increase Effectiveness

Laura M. Milham, James A. Pharmer
Naval Air Warfare Center Training Systems Division
Orlando, FL
laura.milham@navy.mil, james.pharmer@navy.mil

Audrey W. Fok
Kaegan Corp
Orlando, FL
wingsze.fok.ctr@navy.mil

ABSTRACT

Adaptive training systems tailor training to address knowledge and skill deficiencies in response to trainee performance. While research on adaptive training concepts has indicated a strong potential for training effectiveness (Goldberg, Holden, Brawner, & Sottilare, 2011), it may also benefit instructors through the reduction of workload during and after training events. The purpose of this study was to investigate the effectiveness of adaptive training technologies at reducing workload for instructors executing individual level tactical scenario based training (SBT) exercises. Adaptive features on an Instructor Operating System (IOS) were developed to support instructors tasked with monitoring trainees real-time and in preparing for After Action Review (AAR). Instructors from Surface Warfare Officers School (SWOS) Department Head courses were asked to monitor two Tactical Action Officer (TAO) students using one of two systems (either an IOS equipped with adaptive features or one without adaptive features). Performance was assessed by instructors' accuracy in recalling critical student errors; workload was assessed through subjective ratings and task loading (e.g., time required to complete tasks versus time available to complete). Findings indicated that instructors using the adaptive IOS showed a significant improvement in detecting high and medium priority student errors when compared to instructors using a traditional IOS. In addition, instructors reported significantly lower workload in the adaptive IOS condition when compared with the traditional IOS. In fact, instructors estimated 33% time savings for each training exercise if an adaptive system was used versus the traditional IOS.

ABOUT THE AUTHORS

Dr. Laura Milham received her Ph.D. in Human Factors Psychology from the University of Central Florida, specializing in the areas of team performance and team training. She is a Senior Research Psychologist at the Naval Air Warfare Center Training Systems Division where she supports R&D efforts in adaptive team training, Human Systems Integration (HSI), and aviation acquisition.

Dr. Jim Pharmer is a Senior Research Psychologist at the Naval Air Warfare Center Training Systems Division where he is the lead of the Human Systems Integration (HSI) Science and Technology Lab. His research interests are in applying HSI principles to the systems engineering and acquisition processes. He holds a Ph.D in Applied Experimental and Human Factors Psychology from the University of Central Florida and an M.S. in Engineering Psychology from the Florida Institute of Technology.

Audrey Fok is a Research Psychologist at Kaegan and supports the Human Systems Integration **(**HSI) research team at the Naval Air Warfare Center Training System Division. She is a graduate student in the Applied Psychology and Human Factors Ph.D. program at the University of Central Florida. Her research interests include Human Computer Interaction (HCI) and usability in interface design.

Adaptive Instructor Operating Systems

Laura M. Milham, James A. Pharmer
Naval Air Warfare Center Training Systems Division
Orlando, FL
laura.milham@navy.mil, james.pharmer@navy.mil

Audrey W. Fok
Kaegan Corp
Orlando, FL
wingsze.fok.ctr@navy.mil

As training systems become more technologically advanced, they have also increased our ability to collect large amounts of real-time performance data. As a result, it can be a challenge to maintain an optimal balance between automated performance feedback and instructor feedback when dealing with complex scenario based training (SBT) events. In order to develop training systems that provide the most efficient and effective training, it is critical to consider the *roles* of instructors and how best to support them when designing an Instructor Operating System (IOS). Haphazard functional allocation of adaptive training elements that provide a heavy dose of automated feedback to the student without considering the workload and information processing demands on the instructor may limit the instructor's ability to maintain situational awareness in monitoring and assessing student activity, and result in less valuable feedback. As such, a highly automated system may provide minimally effective instruction to the student and overload the instructor who must still contend with a traditional IOS while processing complex, SBT exercises that are adapting in response to student actions. As a result, the instructor may suffer from higher workload thus negatively impacting student through put as data is sorted and processed. The resulting lack of situational awareness may also decrease the instructor's ability to detect critical errors resulting in less effective after action review (AAR).

In response to these challenges, the Adaptive Training for Combat Information Centers (ATCIC) project is developing methods for delivering effective SBT through tailoring training to individual and team performance *while* developing instructor interfaces that facilitate the most efficient level of active instructor involvement. The purpose of this study was to investigate the effectiveness of an IOS with adaptive training components aimed at reducing workload and increasing effectiveness for instructors conducting highly automated SBT.

BACKGROUND

SBT has been used extensively as an effective approach for training the integration of complex skills and naturalistic decision making (Cannon-Bowers & Salas, 1998; Martin, Schatz, Bowers, Hughes, Fowlkes, & Nicholson, 2009). However, while SBT can support student learning within dynamic, relevant, and large scale exercises, it also increases workload demands on instructors, as they must execute and ***monitor*** these scenarios by ***assessing*** individual and team performance indicators of learning objectives, ***diagnosing*** issues with performance, and ***providing feedback*** or remedial action for the trainee(s). The impacts of this are increases in the number of instructors needed to manage the exercise and an increase in workload, which can burden schoolhouses already managing training with limited instructor resources. The end result is a bottleneck that leads to decreased trainee throughput.

To counter some of these pressures on the training community, the Office of Naval Research (ONR) has focused on developing technology for innovative approaches to training that have the potential to increase trainee throughput by decreasing instructor workload in conducting SBT. In part, ATCIC is addressing this need by developing and testing an adaptive IOS that decreases instructor workload and increases instructor effectiveness. As a first step in developing an adaptive IOS system that would mitigate workload issues, a predictive workload analysis was conducted to determine specific sources of instructor workload.

Issues with Instructor Workload

McCracken and Aldrich (1984) suggest that tasks can be decomposed into their Visual, Auditory, Cognitive and Psychomotor (VACP) components to predict the types and degree of workload that will be experienced by individuals engaging in the task. For this effort, Prospective Commanding Officer (PCO) and Tactical Action Officer (TAO) instructor duties were broken down into the following tasks to identify instructor workload issues:

- Performance assessment: Instructors monitor the student as s/he goes through the scenario. While monitoring, instructors observe and evaluate all actions completed by the student(s) and determine performance errors.

- Diagnosis of whether learning objectives are met: Instructors determine if the student has mastered the skill required.

- Remediation of performance deficiencies: Instructors provide feedback/makes adjustments to the training to address specific deficiencies.

Results indicated that current instructor tasks required high amounts of visual and cognitive resources for assessment and diagnostic tasks. To address this, the ATCIC team developed requirements for an adaptive IOS that contains interface features that facilitate assessing, diagnosing, and providing feedback to students in situ, based on reviewing and leveraging design guidelines from empirical and theoretical work (e.g., Tactical Decision Making Under Stress [TADMUS] program; Cannon-Bowers, Burns, Salas, & Pruitt, 1998).

Adaptive IOS Design Functionalities

Based on this analysis, challenges in ***performance assessment*** and ***diagnosing*** trainee deficiencies for instructors centered on the need to track and calculate when student performance deviates from standards. While it can be clear when a trainee is not performing to standard, it can be significantly more challenging to determine *why* there are performance issues. One reason is that, after training, it can be difficult for an unaided instructor to mentally step back through an exercise to determine a trigger event that cascaded into a performance breakdown.

To support this, an IOS was designed that presented and highlighted key performance measures both during scenario execution and after the training session. Adaptive diagnostic support was provided through a table that illustrated whether performance on detect to engage (DTE) actions within the scenario had deviated from expected actions (see Figure 1). Timelines were designed to show individual and summary errors across time, and whether they led to failing a learning objective (see Figure 2).

Figure 1. Table of Detect-to-Engage Breakdowns in Performance (real time)

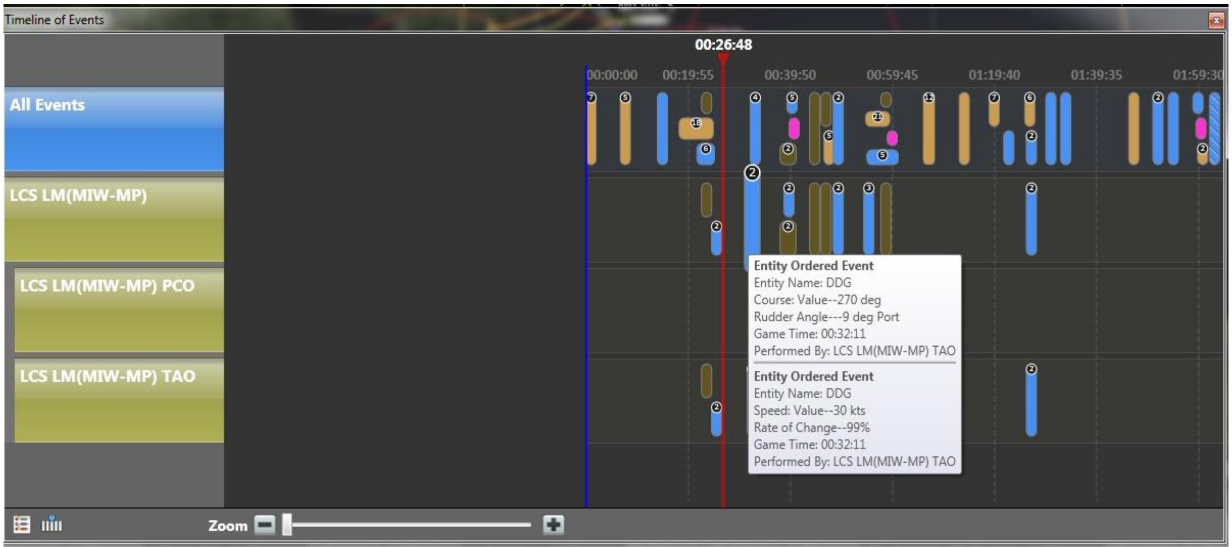

Figure 2. Timeline of Student (Individual or Team) Performance

The final component of the adaptive IOS design is related to ***real time feedback***. Instructors currently determine when and how to provide feedback based on the type of performance breakdowns that occur, and if the issue is more effectively addressed in real time or after action. In support of this, we developed an adaptive real time feedback capability which utilized pre-scripted remediations triggered by diagnostic rules tied to scenario events, which can be instantiated at the instructor's command directly, by permission, or by negation (see Figure 3).

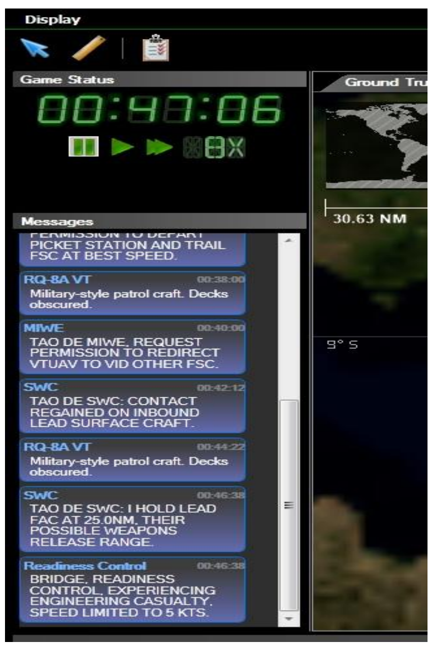

Figure 3. Instructor View of Remediations

Summary

The development of the adaptive IOS took a top down and bottom up approach to identify predicted sources of instructor workload and to design a system using adaptive features to address those issues. The resulting IOS software, known as Portable Adaptive Training and Remediation Instructor Operator Terminal (PATRIOT), was created as an advanced IOS for exploring instructor workload reduction and supporting functionality within individual and team-level adaptive training concepts. To ensure that the design was effective, the ATCIC team conducted an experiment to assess the impact of the design on instructor workload and performance.

Hypotheses

The overarching questions were a) would an adaptive IOS reduce perceived instructor workload when compared to a traditional IOS, b) whether adaptive features of the IOS would increase instructor effectiveness in determining deviations in student performance when compared to a traditional IOS, and c) if instructors perceived that the adaptive IOS would increase the effectiveness and efficiency of training events.

Hypothesis 1: Workload

For instructor workload, we predicted both an overall reduction in workload for conducting the exercise, and a workload reduction in reviewing student performance before AAR for those instructors who were in the adaptive IOS condition vs. those who had a traditional, non adaptive interface. In addition, we predicted that the instructor would need to provide less remediation in the adaptive IOS condition than when in the non adaptive IOS condition.

Hypothesis 2: Instructor Performance

With respect to performance, we predicted that the adaptive IOS display would assist recall accuracy of errors after the scenario, as instructors reviewed performance in preparation for AAR.

Hypothesis 3: Training Effectiveness and Efficiency

An additional hypotheses related to the degree that instructors felt that the adaptive IOS supported the training exercise and after action. Specifically, it was hypothesized that instructors would indicate that using the adaptive IOS would result in a more effective training event and could be used to save time and increase throughput of students.

METHOD

Two IOS conditions (Adaptive IOS and Non Adaptive IOS) were evaluated for this experiment. Consequently, two variations of PATRIOT were used for this study.

Non adaptive PATRIOT IOS

The non adaptive IOS consists of a tactical situation display map depicting geographical information and ground truth representations (i.e., all contacts regardless of whether they are within sensor range of any asset) of the air and surface contacts.

Adaptive PATRIOT IOS

For the adaptive IOS condition, the IOS included the above functions, as well as interface features (e.g., tables and visual displays) and resources that will support instructor participants in maintaining situational awareness, performance assessment, diagnoses, providing feedback and preparing for AAR (see Figure 4).

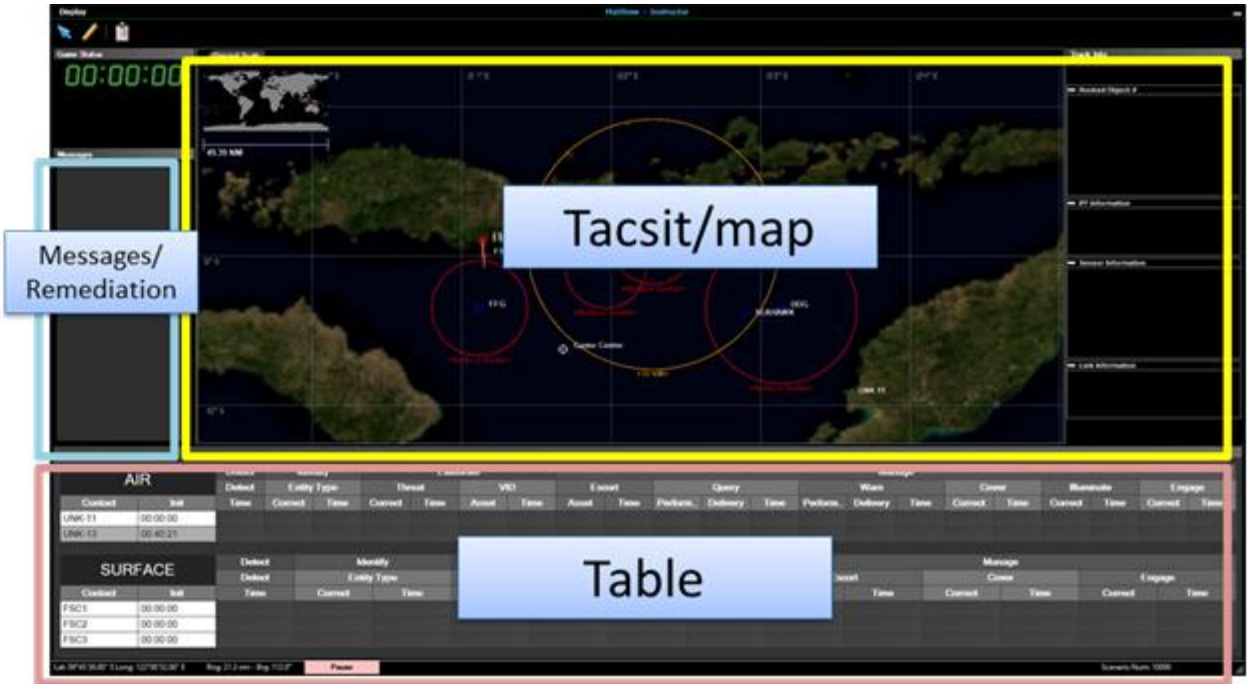

Figure 4. Overview of PATRIOT IOS real time station

Instructors utilized both versions, consecutively, in order to conduct two concurrent training exercises. The relatively small population of Department Head instructors at Surface Warfare Officers School (SWOS) necessitated using a within-subjects design commonly used for experiments with small sample sizes. Because all participants receive all treatments, and differences due to participants can be removed from the error term, overall variability is reduced and, as a result, statistical power is increased (Howell, 1992). The increased statistical power design allowed us to determine the impact of the adaptive training interventions on ability to assess performance issues in real time, instructor workload, recall accuracy of performance, diagnosis and remediations needed, and time to prepare for performance review.

Participants

This experiment was conducted at SWOS, with twelve SWOS Department Head instructors as participants, with an average of 15 years in service and 15 months as Department Head instructors. As these instructors are the targeted population and intended users of the training system under development, they are in the best position and have the necessary qualifications to perform the training tasks involved.

Measures

Demographic questionnaire: Participants were given a demographic questionnaire to gather instructor experience.

Workload Measures

NASA TLX: The National Aeronautics and Space Administration (NASA) Task Load Index (TLX) (Hart & Staveland, 1988) was used to assess subjective workload. This scale provides a multi-dimensional rating procedure that derives an overall workload score based on ratings on six subscales. The subscales include Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration.

Number of times instructor indicated that they needed to provide remediation: This measure was captured through the instructor pressing the “arrow” key on the keyboard.

Instructor Performance Measures

Post-Exercise Recall of Trainee Performance Questionnaire: Participants were given a paper-based knowledge test on student performance.

Training Effectiveness and Efficiency Measures

Post AAR Training Effectiveness Questionnaire: Participants were asked to rate the degree that the adaptive IOS supported effective training on a Likert scale.

IOS Usability Questionnaire: Participants were asked to provide feedback on the adaptive IOSs' content and interface, and estimated time savings for conducting SBT.

Time to complete Post-Exercise Recall of Trainee Performance Questionnaire: This measure was captured through experimenter's stopwatch.

Training Simulation and IOS

PATRIOT provides runtime control and instructor-in-the-loop evaluation of individual and team-level adaptive training scenarios. It was created using an instructor-centered design approach with capabilities derived from the approach described above.

Specifically, PATRIOT consists of a number of reusable subcomponents that target the various instructor support functions required within an adaptive training system to include a table that highlights deviations in trainee performance (see Figure 1), a remediation display (see Figure 3) and, for use in after action reviews, it provides a timeline of trainee actions (see Figure 2) to prompt accurate recall.

Procedure

Upon arrival, instructor participants reviewed and signed an informed consent form, completed a demographic questionnaire, received a slideshow for training orientation on the simulation system and IOS symbology, and then were assigned to one of two starting conditions in counterbalanced order (adaptive or non adaptive IOS).

For both the non adaptive and adaptive IOS scenario execution trials, the instructor participants were asked to perform SBT tasks consisting of assessing, diagnosing, and remediating the pre-recorded performances of two individual confederate TAOs, each working through the 24 minute scenario exercise. When in the adaptive IOS condition, instructors watched TAOs who received automated remediations; when in the traditional IOS condition, TAOs did not receive automated remediations.

The first confederate TAO recording was of a poorly performing individual and the second TAO illustrated average performance as predetermined by Subject Matter Experts (SMEs). The instructor participants had the ability to interact with the two IOSs as needed to assess and diagnose performance to determine how well the individuals were performing the scenario tasks. Instructors were asked to indicate each time they wanted to provide feedback during a training scenario. Each condition lasted approximately an hour.

Post Trial Evaluations

After each trial, instructors were administered the NASA TLX to evaluate the workload associated with assessing, diagnosing, and determining feedback needed for the 2 TAOs. Instructors were then asked to complete the Post-Exercise Recall of Trainee Performance Questionnaire. Finally, instructors were again administered the NASA TLX to rate the workload of preparing for an AAR through the recall of trainee performance.

Following a short break, instructors were assigned to the IOS condition to which they had not been previously exposed, with a novel set of scenarios. The procedure was the same as described above, but the confederate TAOs illustrated different patterns of trainee performance.

At the end of the second trial, instructors were asked to complete the NASA TLX, the Post-Exercise Recall of Trainee Performance Questionnaire, and the NASA TLX after completing a questionnaire on trainee performance. In addition, they were asked to subjectively evaluate the degree that the adaptive IOS supported effective training and the perceived usability of the adaptive IOS by completing the Post AAR Training Effectiveness Questionnaire, and the IOS Usability Questionnaire, respectively. In all, each participant took 2 hours to complete the experiment.

RESULTS

Workload

During Training

In using the adaptive system to monitor average performing individuals, instructors indicated that they would need to provide feedback significantly less often (2.00 times during the scenario) than they would using the non adaptive IOS (5.33 times during the scenario, $t[11] = 2.45$, $p < 0.05$). These results were consistent with the data on instructor NASA TLX ratings, which indicated instructors viewed their overall workload, effort, and physical demand with the adaptive IOS to be significantly less than with the non adaptive IOS during scenario execution.

After Training

Immediately following the training scenario, instructors indicated that overall workload was significantly less with the aid of the adaptive IOS (Mean = 42.25) than with the non adaptive IOS (Mean = 51.67, $t[11] =1.85$, $p < 0.05$). This was supported with significantly lower workload ratings across the individual NASA TLX mental, physical, temporal demands, effort, and frustration dimensions for adaptive than non adaptive conditions.

Post-AAR Questionnaire

The NASA TLX ratings provided by instructors after completing the Post-Exercise Recall of Trainee Performance Questionnaire were also significantly lower overall in the adaptive (Mean = 32.08) versus the non adaptive (Mean = 61.58) conditions, $t(11) = 3.89$, $p < 0.01$. Further, this result was consistent for each of the individual NASA TLX dimensions.

Instructor Performance

Instructor recall of Student Performance

Table 1 depicts the analysis of the Post-Exercise Recall of Trainee Performance Questionnaire. The analysis revealed that, overall, instructors using the adaptive IOS scored overall 15.2% higher on recalling trainee performance issues after the scenario than those using the non adaptive IOS ($t[11]=5.24$, $p <.001$). Further, instructors using the adaptive IOS scored significantly higher in recalling student performance issues established a priori as high priority (9.99% better scores, $t[11] = 3.32$, $p <0.01$) and medium priority (28.89% better scores, $t[11] =6.46$, $p < 0.001$) than those instructors using the non adaptive IOS. No significant difference was found between groups on recall of low priority issues. It should be noted that this appeared to be accompanied by a speed accuracy tradeoff in that instructors with the adaptive IOS completed the questionnaire significantly slower (10 min vs. 5.75 minutes) than those with the non adaptive IOS($t[11]=3.78$, $p <.01$).

Table 1. Post-Exercise Recall of Trainee Performance Questionnaire

Percentage Correct adaptive vs. non adaptive	Mean Difference
High Priority	9.99*
Medium Priority	28.89*
Low Priority	3.57
Total	15.20*

* p <.05

Training Effectiveness and Efficiency

Instructors reported that the system would support moderate improvements across 5 effectiveness dimensions (See Table 2).

Table 2. Post AAR Training Effectiveness Questionnaire Results

Effectiveness Dimension	Average
Assessment capability	3.17
Diagnostic Capability	3.33
After Action Review Effectiveness	3.67
Instructor Effectiveness	3.58
Instruction Effectiveness	3.17

On average, instructors estimated that the adaptive IOS would save them 33.13% time in real time assessing, diagnosing, and remediating team performance and 38.25% time in preparing AAR.

DISCUSSION

Overall, instructors reported lower workload both during the training execution, and afterwards, in preparing for a review of performance. During execution, instructors did not need to provide feedback as often to students with the adaptive system, reducing the workload of remediating students manually. Finally, when asked to consider time savings if an adaptive IOS system were available, they estimated a 33% reduction in the time needed to provide assessment and remediation during training, and a 38% reduction in time needed after training in preparing for AAR. This has dramatic implications for student throughput.

In addition to workload, instructor performance was also enhanced. While the time needed to complete the questionnaire of student errors increased slightly with the adaptive IOS, the quality of instructor performance also increased. Arguably, this increase in amount of time to complete the questionnaire can be attributed to the addition information that the instructors in the adaptive IOS group provided. Notably, the adaptive IOS facilitated recall of high and medium priority errors exhibited by the students. Specifically, instructors noticed and noted 10% more high priority errors related to major training objectives within the scenario. As the instructors were briefed as to the goals of the scenario, and the performance standards for meeting those objectives, it is a critical finding that the adaptive IOS increased the ability of the instructors to determine if scenario objectives were met. Perhaps more importantly, however, was the 28% increase in recall of medium priority errors. Medium priority errors reflect the enabling objectives, or those that illustrate *why* a student failed at meeting an overall objective. Within training events, process level performance is often the focus of training, as it provides students with stepping stones to achieve mission success. Within the training science literature, it has been suggested that this approach is the most effective, as performing actions correctly, in the right way, the right patterns, etc., are a better predictor of success than solely an outcome measure, which may have been achieved while following incorrect procedures (Cooke, Salas, Cannon-Bowers, & Stout, 2000). Despite their importance, however, they are often difficult to capture, especially in simulation exercises (Cooke, Salas, Cannon-Bowers, & Stout, 2000). As such, the increase in recall of these enabling objectives has the potential to increase the overall effectiveness of the exercise by allowing a more complete and actionable after action review.

This effort is currently being conducted under the Office of Naval Research Capable Manpower Future Naval Capability program. The views expressed herein are those of the authors and do not necessarily reflect the views of the Department of Defense, the Department of the Navy, or the Naval Air Warfare Center Training Systems Division.

REFERENCES

Cannon-Bowers, J. A., & Salas, E. (1998). Team performance and training in complex environments: recent findings from applied research. *Current Directions in Psychological Science, 7(3)*, 83−87.

Cannon-Bowers, J. A., Burns, J. J., Salas, E., & Pruitt, J.S. (1998). Advanced technology in scenario-based training. In J.A. Cannon Bowers & E. Salas (Eds). *Making Decisions Under Stress: Implications for Individual and Team Training (pp. 365-374)*. Washington, DC. American Psychological Assoc.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. (2000). Measuring team knowledge. *Human Factors, 42*, 151-173.

Goldberg, B., Holden, H., Brawner, K., & Sottilare, R. (2011). Enhancing performance through pedagogy and feedback: domain consideration for ITSs. *Proceedings of the 2011 Interservice/Industry Training, Simulation and Education Conference*.

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload (pp. 139-183).* Amsterdam: North Holland

Howell, D. (1992). *Statistical Methods for Psychology* (3rd ed.). Belmont, CA: Duxbury Press.

Lyons, D. M., & Allen, R. C. (2000). Mobile aid for training and evaluation (MATE): a hand-held, configurable set of team performance and measurement tools. *Proceedings of the 2000 Interservice/Industry Training, Simulation and Education Conference*, 661-671.

Martin, G., Schatz, S., Bowers, C. A., Hughes, C. E., Fowlkes, J., & Nicholson, D. (2009). Automatic scenario generation through procedural modeling for scenario-based training. *Proceedings of the 53rd Annual Conference of the Human Factors and Ergonomics Society*, Santa Monica, CA., 1949-1853.

McCracken, J. H., & Aldrich, T. B. (1984). Analysis of selected LHX mission functions: Implications for operator workload and system automation goals (*TNA AS1479-24-84*). Fort Rucker, AL: Anacapa Sciences, Inc.