AUTOMATED INDIVIDUAL, TEAM, AND MULTI-TEAM PERFORMANCE ASSESSMENT TO SUPPORT DEBRIEFING DISTRIBUTED SIMULATION BASED EXERCISES (DDSBE)

Thomas F. Carolan
Micro Analysis & Design, Inc
Boulder, CO

Peter Bilazarian
Lockheed Martin Advanced Technology Laboratories
Cherry Hill, NJ

Long Nguyen
NAVAIR Orlando Training Systems Division
Orlando, FL

Automated, embedded data collection, assessment, and integration capabilities are key requirements of an instructional framework to support performance evaluation and debrief of multiple teams participating in distributed simulation-based exercises. This paper discusses recent progress in the application of automated performance data collection and assessment capabilities as part of a prototype Debriefing Distributed Simulation-Based Exercises (DDSBE) system. The automated data collection process obtains data from local and distributed simulation systems and operator consoles to assess individual, team, and multi-team performance on training objectives during critical and key events. Performance is assessed at the multi-team, team, and individual levels as appropriate. Automated and observer-based semi-automated assessments are integrated into data products suitable for debrief development. Methods, products, and results from the research and development effort to date are discussed.

INTRODUCTION

As the technical capabilities supporting training and practice in distributed simulation-based training environments continue to improve, they are increasingly a viable alternative to live training for maintaining and improving many mission essential knowledge and skills. The application of advanced learning technology is needed to support instructors in effectively and efficiently assessing performance and providing focused learning feedback. The simulation environment provides a rich source of objective performance data that can be used to monitor, quantify and assess elements of performance process that are typically not adequately captured by observer-based measurement systems. Taking advantage of this source of performance information will free expert observer/instructors to focus limited attention on aspects of performance that are not amenable to automated processing. Automated, embedded data collection, assessment, and integration capabilities are therefore key requirements of an instructional framework to support performance evaluation and debrief of multiple teams participating in distributed simulation-based exercises.

One of the challenges involved in automating performance assessment using simulation-based data is determining what to measure. Performance data can support assessment of performance outcomes and monitoring performance process. Many variables from different sources related to relevant scenario state situations, process skills, and mission outcomes can be monitored, time-stamped, and stored. Finding a balance between selective measurement to support current training objectives and exploiting the capability to monitor and assess data relevant to multiple performance requirements is a factor in the effective use of automated performance monitoring and assessment tools.

An issue that constrains automated simulation based assessment is the fidelity of the training system. Lower fidelity may mean different or abbreviated procedures, and limited or degraded information sources and performance cues. Even if data is available to assess a particular behavior, it may not be instructionally useful if the cues that support the behavior in real world situations are not available.

An assessment approach that integrates performance information from multiple sources including data captured from the simulation environment and operator consoles and data from experts observing training exercises provides more complete and objective performance information to support post-exercise diagnosis and debrief. It allows developing historical data that can be used not only to evaluate performance readiness but also to evaluate the effectiveness of training systems and approaches and provide a common metric to evaluate transfer to the operational environment. An integrated approach provides a sound basis for identifying proficiency deficiencies and adapting training to address identified deficiencies. By using a common measurement framework, observation- and simulation-based data can be integrated to provide assessments at multiple levels. Finally, the user community needs the capability to access performance data at each level of analysis and to trace performance information through the system to assure reliability and accuracy.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **SEP 2005**	2. REPORT TYPE	3. DATES COVERED **00-09-2005 to 00-09-2005**

4. TITLE AND SUBTITLE **Automated Individual, Team, and Multi-Team Performance Assessment to Support Debriefing Distributed Simulation Based Exercises (DDSBE)**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Lockheed-Martin Advanced Technology Laboratories,3 Executive Campus, 6th Floor,Cherry Hill,NJ,08002**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**5**	

DDSBE AUTOMATED DATA COLLECTION AND ASSESSMENT CAPABILITIES

Last year at the Debriefing Distributed Simulation-Based Exercises (DDSBE) symposium, we presented an overview of the methods and technologies applied in the first year (Spiral 1) of the DDSBE program to develop automated performance data collection and assessment capabilities to support performance analysis of distributed teams in simulation-based exercises (Carolan & Bilazarian, 2004). In this follow-up paper, we summarize the automated data collection, assessment and integration results from the initial phase of the DDSBE prototype development for distributed Navy E-2C Hawkeye and F/A-18 Hornet teams, and address the plans, and accomplishments to date for the second phase.

The DDSBE assessment system consists of software capabilities to automate performance data collection and reduction, to detect and assess performance deficiencies, and to integrate assessment information from automated and semi-automated capabilities to support diagnostic analysis and debrief development. Figure 1 provides an updated top-level architecture for the DDSBE automated data collection and assessment component. This architecture involves interoperability between the distributed exercise simulation environment and the DDSBE Team and Multi-Team Performance Evaluation and Debriefing training support system. The Performance Evaluation and Debriefing module is constructed on an open, flexible, and scalable software framework. A communications layer and domain model Application Programming Interface (API) provides a reusable, domain-independent, and Object-Oriented C++/Java application integration framework for DDSBE. It extends the single team API design developed for the Navy Advanced Embedded Training Program (Zachary, et al, 1999) to multiple teams.

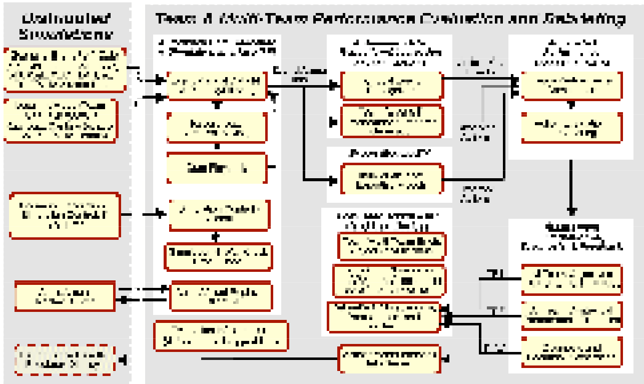

Figure 1. DDSBE Automated Data Collection, Assessment, and Diagnosis Architecture and Interfaces

Automated Data Collection and Reduction/Observation

The Automated Data Collection and Reduction (ADCR) component performs local and global automated performance data recording of human behavior during distributed exercises. Automated data collection involves parsing and filtering operator console manual/keystroke actions and scenario event data received from local and distributed simulation systems. These data typically are drawn from the following sources:

1. Scenario state and event data (e.g., entity 'ground truth' identification, position, and velocity data),
2. Local console, mission computer, and database (e.g., track file) data that record operator manual/keystroke actions (e.g., switch selections, menu selections, pilot 'joystick and throttle' activity) at user workstations.
3. Communication data that record the relay of information between exercise participants.

Additional filtering and aggregation of collected data, by the Automated Observation of Team Actions component, reduces it to the meaningful actions required for performance measurement, assessment, and debriefing. The Critical Event Recognizer determines if a scenario event is key or critical and if so passes it to the Automated Performance Assessment (APA) and Semi-Automated Assessment components opening an evaluation window for assessment.

Automated Performance Assessment

The DDSBE approach incorporates an Event-Based Approach to Training (EBAT) as described in Johnston, Cannon-Bowers and Smith-Jentsch (1995). The EBAT approach enhances the scenario-based learning process by linking learning objectives, performance assessment, diagnosis, and debriefing feedback to key scenario events. For each key scenario event, the set of expected response actions and attributes, and the time window within which those actions should occur, are defined.

During the training exercise, the automated assessment component uses this expected or expert performance data as a basis for evaluating observed performance related to key scenario events. The team performance data includes operator keystroke-based actions and attributes that are automatically captured by the data collection and reduction component described above. Team performance data also includes speech-based actions (voice reports) and associated attributes that are captured by human observers/evaluators using computer-based data collection tools, such as the Virtual Communications Assessment Tool, or VCAT, shown in Figure 1. Automated performance assessment compares the actions observed with the set of expected responses. Actions are assessed at the individual watchstander level, and, for each key event, performance is assessed at the team level. The action assessments are aggregated over each event. At each assessment level an acceptable, not acceptable, or above acceptable rating is assigned, based on predefined standards. Performance deviations and assessments are delivered to the API as assessment products for use by assessment integration and diagnostic processing and by human evaluators.

Assessment Integration

Assessment Integration is an automated process that combines individual, team, and multi-team results from the APA and Semi-Automated Assessment components into meaningful data products suitable for single-team and multi-team debriefings and post-exercise analysis. A major benefit of assessment integration is the automatic and rapid creation of a set of objective-based and performance-based debriefing products with integrated replay for single-team and multi-team debriefs. Assessment integration includes the coordination and integration of assessments associated with team outcomes and team processes to provide integrated products to support diagnosis and debrief. These include cumulative summaries of performance on team objectives, team performance examples, evaluations of teamwork process (such as, information exchange, communications, leadership, and supporting behavior), and higher level rating schemes (such as, Mission Essential Task List).

AUTOMATED DATA COLLECTION AND ASSESSMENT: SPIRAL 1 RESULTS

One of the primary objectives of the initial phase (Spiral 1) of the DDSBE project was to demonstrate a capability to perform real time automated data collection, assessment, and assessment integration in a distributed multi-team environment. The initial test environment to support DDSBE prototype development, testing and demonstration consists of three E-2C Advanced Control Indicator Set (ACIS) consoles, a Joint Semi-Autonomous Forces (JSAF) scenario generation and simulation control environment, a Mission Computer, and the Run Time Infrastructure.

The automated data collection and assessment emphasis was on E-2C team performance. Examples of E-2C operator keystroke-based actions and attributes include keystroke actions that support track monitoring, track information collection and identification activities, and communicating with controlled aircraft. E-2C Team manual/keystroke actions on the ACIS console were collected via recorded messages available in the E-2C Team Simulator's Mission Computer and scenario entity info was captured from the High Level Architecure (HLA) interface. These E-2C actions were converted to XML statements for input to the Automated Performance Assessment (APA) component via standard XML interface mechanisms. We developed initial capabilities for performing dynamic recognition of scenario critical/key events, such as new air tracks that originate from potential hostile airfields and are flying outside of commercial air lanes.

The DDSBE team developed a domain-independent expertise syntax for representing individual and team expected actions during scenario critical and key events. This general expertise representation is applicable to fully ordered, partially ordered, and unordered sets of expected team actions. It can also model and incorporate different acceptable task strategies for responding to critical and key scenario events.

The initial objective for the APA component was to develop assessment software that would automatically differentiate between gradations of performance on a number of different measures to support evaluation of performance process and outcome at the individual and team level. The initial focus was on assessing E-2C team watchstander keystroke based-actions and the overall E-2C team response to key and critical events. The automated assessment process involved identifying deviations from expected keystroke action performance on selected measures, generating assessment scores based on the observed deviations, generating assessment scores for speech actions based on the evaluator inputs, and aggregating these action scores to generate an assessment score for the E-2C team on each key and critical event. Figure 2 illustrates the APA Spiral 1 general design approach.

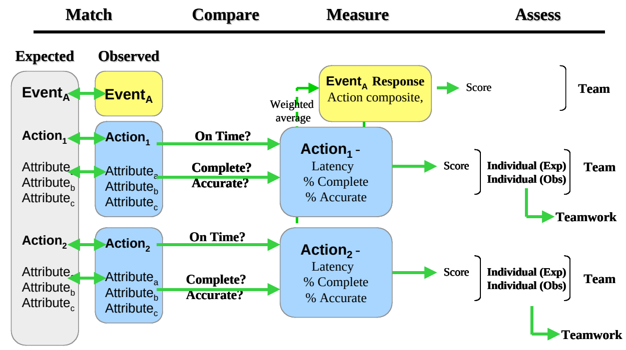

Figure 2. Spiral 1 APA design approach

For Spiral 1, four types of measures were implemented. At the ***task action level***, **individual** performance was assessed for:

- Completeness - was an expected action completed by the expected watchstander and with all the expected attributes provided?
- Accuracy – were all attribute values accurate?
- Timeliness – was the action within the acceptable response time?
- Order – were all prerequisite action requirements met?

At the ***event level,*** the same measures were applied to the set of **team** actions expected in response to an event. At the team level credit is given if any watchstander performs the action. In cases where an individual's performance is deficient in some way but the team as a whole successfully accomplishes the task, teamwork process may account for the difference. An example of teamwork is when one operator performs or corrects actions assigned to another operator who, due to high workload or distraction, misses an action or makes an error. This has been called supporting, compensatory or back-up behavior (Smith-Jentsch, Johnston & Payne, 1990). **Teamwork** measures are typically captured and assessed by the human observers. However, APA can support teamwork assessment. For example, APA can capture instances when a task is performed by a different watchstander than expected (in this case also performed on a different console). Currently we are identifying and flagging these instances.

APA creates a user accessible log of each event and action received from the API with the observed value of each of its attributes. The log allows an observer to view in real time all scenario events and associated expected actions extracted from the Master Scenario Event List. The log also shows all observed events and actions as they are received from the API and the evaluations computed and transmitted for each attribute, action and event. The teamwork panel identifies those E2-C keystroke tasks that are flagged as potential back-up behaviors.

The primary Assessment Integration products generated for Spiral 1 include products at the team and multi-team levels. At the multi-team level, Assessment Integration of semi-automated outcome and process measurements of performance was developed, using training objective ratings supported by inter-team checklist types of performance measures.

The E-2C Team Single-Team products include:

- automated generation of Team-Level Training Objective Summaries, using event-based inputs from APA,
- automated generation and sorting for debrief of knowledge-rich Team Contextual Performance Examples,
- integration of semi-automated measurements of important teamwork processes from an Instructor/Evaluator using VCAT rating scales and checklists on a tablet hand-held computer.

Figure 3 shows a screen shot from the DDSBE Spiral 1 Assessment Integration Graphical User Interface (GUI) that provides an example of an Assessment Integration product involving Team Performance Examples. This figure shows the first of ten automatically generated E-2C Team Contextual Performance Examples, priority sorted by event criticality (with Critical Events listed first, followed by Key Events) and by poor to good team performance over the entire scenario. The relevant event training objective(s) and performance measure(s) are automatically attached to this Performance Example as well as the event time interval and trigger. A concise, qualitative summary-with-context (e.g., specific Track Nos., Reports, Manual/Keystroke Actions cited) of E-2C team performance during this event is provided that uses information from APA to describe the results of how individual E-2C team members (ACO or CICO) performed. Finally, relevant audio, visual, and manual replay data can be attached to this event that would help an instructor decide during AAR preparation whether this event should be brought to the team's attention during debrief. These Team Performance Examples provide one element of an overall Assessment Integration Team Performance Report that will help instructors and teams to analyze 'what happened' during the scenario.

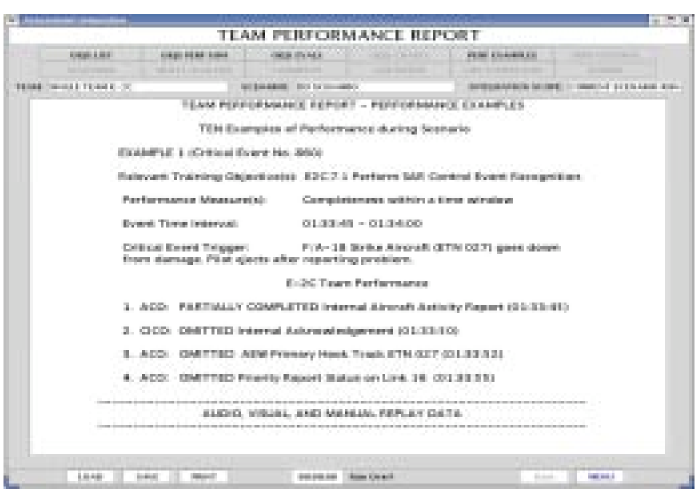

Figure 3. Assessment Integration Team Performance Examples

Interim Evaluation

An interim evaluation of DDSBE was conducted at the end of the first phase (see Johnston, Radtke, Salter & Freeman, this symposium). This evaluation included a pilot test of the DDSBE System involving Naval personnel that participated in nine separate Spiral 1 scenario runs. The Navy personnel operated E-2C consoles and performed as instructors/ evaluators using VCAT semi-automated measurement capabilities during a strike mission scenario. Results indicate that, for virtually all of the critical and key events associated with nine separate scenario runs, the ADCR, APA, Assessment Integration, and other automated assessment components performed accurately, reliably, and stably. The evaluation demonstrated the capability of the ADCR component to capture critical performance data from the HLA environment and local mission computers in real time and convert it to meaningful data for use by APA. For all nine runs, the ADCR component accurately, consistently, and reliably captured and interpreted E-2C Team manual/keystroke data. It also communicated this data efficiently and rapidly to the APA component for further processing. In addition, the overall Spiral 1 Automated Data Collection and Assessment capability was able to differentiate consistently, clearly, and accurately multiple gradations of E-2C individual and team performance (e.g., 'above satisfactory', 'satisfactory, and 'unsatisfactory') at the action, critical/key event, and aggregate scenario levels.

SPIRAL 2 AUTOMATED ASSESSMENT OBJECTIVES

The current development cycle includes the addition of an F/A-18 virtual simulator component. In addition to automated assessment of keystroke-based performance data, the F/A –18 simulator provides opportunities for automated assessment of tactical and flight performance data at the individual, team, and multi-team levels, using aircraft position, kinematic, and weapons release data. The APA software is being expanded to support capabilities to provide automated support for assessment of F/A-18 pilot and team performance, E-2C individual and team performance, and support for within team and across team teamwork performance. Automated assessment can address outcome measures, such as kill ratios and bombs on target, and process measures, such as adherence to briefed contracts, formation flying, and timeline management. Assessment algorithms are in development to identify and assess quantitative measures for various maneuvers. For automated assessment of teamwork behaviors, we are refining the process of flagging potential back-up behaviors for assessment. The next step is to refine the process by identifying when a particular operator may be in need of back up (e.g., through task demand or workload measures) and when an operator may be available for back up (e.g., light workload, capable of performing the required task). We are exploring a combination of approaches that would track the ongoing task and complexity requirements of open events (e.g., Hudgell & Gingell, 2001) and identify the criticality of those tasks (e.g., Bolton, Dorsey & Campbell, 2004).

Spiral 2 Assessment Integration is developing new capabilities for automatic generation of Team Performance Reports and Performance Examples (see Figure 3) associated with: (1) a live F/A-18 ‘Sweep’ Team of air-to-air fighter pilots; and (2) a Community (or Multi-Team) consisting of an E-2C Team and an F/A-18 Sweep Team. For various E-2C, F/A-18, and Community Team Performance Examples, we will include pertinent snippets of audio/visual replay. This replay will consist of audio voice net recordings associated with an event and multi-modal data captured by VCAT evaluators that can involve video capture of operator screens. Assessment Integration will also be developing new capabilities for prioritizing and combining new types of quantitative APA measurements involving aircraft tactical and flight performance data, that may be associated with specific critical/key events or span multiple scenario events. Spiral 2 Assessment Integration will generate new outputs (communicated via the API) to the Diagnosis and Debrief/AAR Preparation, Delivery, and Replay components (shown in Figure 1). After all team debriefs are completed, the Diagnosis and AAR components will provide team performance results and suggestions for improvement to Assessment Integration. This will facilitate future post-exercise reconstruction and analysis activities by instructors and evaluators as well as data archival to Training and Learning Management Systems. It will also enable the fleet to perform data mining activity and to analyze a rich set of automated performance results across multiple scenario runs and operational teams. This will help to efficiently determine the focus, objectives, event types, and degree of difficulty for future individual, team, and multi-team training scenarios, while helping to ensure the appropriate utilization of limited DOD training resources.

ACKNOWLEDGEMENTS

This work is supported with funding from the Office of Naval Research. The ideas expressed herein are those of the authors, and are not necessarily endorsed by the funding agency. This material is approved for public release; distribution is unlimited.

REFERENCES

Bolton, A.E., Dorsey, D.W. & Campbell, G.E., (2004). Application of time sensitive computational models to training within a scenario-based team training system.

Carolan, T.F., & Bilazarian, P. (2004). Automated Data Collection and Assessment for Debriefing Distributed Simulation-Based Exercises (DDSBE). In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Hudgell, A.J. & Gingell, R. M. (2001). Assessing the capacity of novel air traffic management systems. 4th USA/ Europe Air Traffic management R&D Seminar. Santa Fe, NM.

Johnston, J.H., Radtke, P.H., Salter, W. & Freeman, J. (2005). A Human Systems Integration Method for Validating Team Performance Assessment Within a Simulation-Based Training System. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Johnston, J. H., Cannon-Bowers, J. A., & Smith-Jentsch, K. A. (1995). Event-based performance measurement system for shipboard command teams. *Proceedings of the Command and Control Research and Technology Symposium*, Washington, DC, 274-276.

Smith-Jentsch, K. A., Johnston, J. H., & Payne, S. (1990). Measuring team-related expertise in complex environments. In J. Cannon-Bowers & E. Salas (Eds.), Making *decisions under stress: Implications for individual and team training (p61-87).* Washington, *DC*: American Psychological Association.

Zachary, W., Cannon-Bowers, J., Bilazarian, P., Krecker, D., Lardieri, P., & Burns, J. (1999). The Advanced Embedded Training System (AETS): An Intelligent Embedded Tutoring System for Tactical Team Training. *International Journal of Artificial Intelligence in Education*, October 1999, volume 10, pp. 257-277.