Collaborative Performance Evaluation in Multi-Platform Team Training Exercises

Thomas F. Carolan
Micro Analysis & Design Inc.
Okemos, MI 48864
tcarolan@maad.com

Joseph J. Mihal Jr. & Tom Little
Fleet Anti-Submarine Warfare Command
San Diego, CA 92147-5119
joe.mihal@navy.mil
tom.little@navy.mil

Paul H. Radtke
NAVAIR Orlando Training Systems Division
Orlando FL
paul.radtke@navy.mil

ABSTRACT

The Fleet Anti-Submarine Warfare Command (FLTASWCOM) trains Atlantic and Pacific Strike Groups in the area of integrated Anti-Submarine Warfare (ASW). FLTASWCOM training events are flown on both instrumented ranges and in open ocean areas and may last up to 12 hours, and frequently involve up to ten air, surface, and submarine participants. Evaluating these events involves simultaneous real time observation, assessment, and analysis of both individual and multi-platform cooperative missions. The team of evaluators must track, record, assess, and analyze the participants' performance in real time. The evaluators must produce a timely debrief for the participants, provide an in depth analysis for their leaders, and use the evaluation results to tailor and improve the training process.

Tactical Warfare Instructor Support Environment (TacWISE) is an integrated set of training support tools that has been successfully employed to evaluate performance of Navy Strike Groups for over three years. Assessors and mentors have been located at sea, in the air and on shore. Simultaneous evaluations were performed on each individual ship in a STRIKE GROUP, in aircraft participating in the same event and from shore based observation points. Results and lessons learned from applications in Strike Group-level ASW and surface warfare training exercises will be discussed.

ABOUT THE AUTHORS

Thomas F. Carolan is a Senior Research Psychologist with Micro Analysis & Design, Inc. Boulder CO. He received his Ph.D. in Experimental Psychology from the University of Connecticut. He has been involved in research and development related to training systems, performance measurement, and human performance modeling for the past 17 years.

Joseph J. Mihal Jr. is an ASW Operation Specialist with Fleet ASW Command, San Diego CA. He has 20 years of Navy experience including aviation and submarine tours. He has been involved with training and evaluating Navy tactical teams for the past 10 years. He holds a Masters of Science from Naval Postgraduate School in Space System Operations and a Bachelors of Science in Mechanical Engineering from Villanova University.

Tom Little is the Technical Director of Fleet ASW Command, San Diego CA. He has 25 years of experience in government, industry and academia. During his career with the Department of the Navy he was the Deputy & Technical Director of the Center of Maritime Dominance, Technical Director of the Sea-based Weapons & Tactics School, Wargame Coordinator at Fleet ASW Training Center, West Coast Fleet Representative at Commander, Anti-Submarine Warfare Wing, and a Senior Naval Oceanographic Office Scientist at Stennis Space Center. He holds a Bachelors of Science in Oceanographic technology from Florida Institute of Technology.

Paul Radtke is a Research Psychologist with NAVAIR Orlando Training Systems Division. He has had 15 years of experience working on training-related research for the Navy and an additional 13 years working as a trainer and researcher for several professional associations and consulting firms. He holds a Masters degree and has completed coursework for the Ph.D. in Political Science from Northern Illinois University.

Collaborative Performance Evaluation in Multi-Platform Team Training Exercises

Thomas F. Carolan
Micro Analysis & Design Inc.
Okemos, MI 48864
tcarolan@maad.com

Joseph J. Mihal Jr. & Tom Little
Fleet Anti-Submarine Warfare Command
San Diego, CA 92147-5119
joe.mihal@navy.mil
tom.little@navy.mil

Paul H. Radtke
NAVAIR Orlando Training Systems Division
Orlando FL
paul.radtke@navy.mil

INTRODUCTION

In military and other high performance environments, a goal of advanced training is improving and maintaining individual and team performance readiness. Distributed simulation-based training environments continue to play an increasingly important role in warfighter training. However, performance readiness typically is still assessed through multi-platform training events conducted on an instrumented training range or off range.

The effectiveness of the training process is a critical factor in the continuous improvement of individual and team proficiency. Current trends toward workforce reduction increase the need to optimize training effectiveness and efficiency. To focus training to efficiently and effectively attain and maintain performance readiness, a framework is needed to monitor the training process, determine its impact upon combat readiness, and modify the training process as needed to assure continuous improvement.

Figure 1 illustrates a training cycle that supports performance in complex environments. The core training process, represented by the inner panel of Figure 1, involves recording and evaluating observed performance to focus subsequent debrief on observed deficiencies. Focusing the post-exercise debrief on objectively defined measures of performance provides the participants with information to self-correct their performance. The instructor/evaluator typically is the key performance assessment and diagnosis agent in this training and evaluation cycle. As expert instructors observe live training performance, they develop a sense of what is going right or wrong, and intuitively diagnose the reasons why things occur. They then organize their observations into valuable feedback and lessons learned.

The second aspect of the training process, represented by the outer panels in Figure 1, is support for the goal of continuously improving advanced training. This involves feeding the results of the assessment and debrief process into a performance and training analysis process that will drive a feedback loop to training development, and provide the data needed to monitor the impact of training on performance. As training interventions become more effective through the focus on performance, measures of performance readiness also should improve.

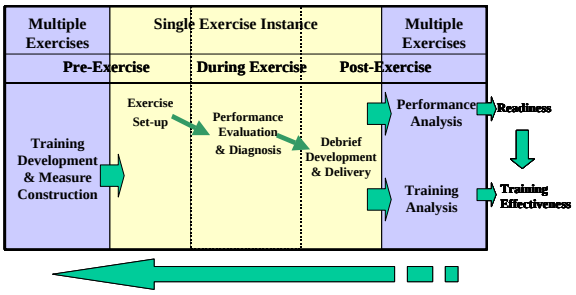


Figure 1. Training Cycle

Approach

An Event-Based Approach to Training (EBAT) has been researched and refined at the NAVAIR Orlando Training Systems Division (e.g., Cannon-Bowers, Salas, & Converse, 1993) over the past decade. In this approach, critical events are identified and related to the tasks or actions that they should trigger. When applied to simulation-based training, scenario events can be designed and constructed to achieve specific training purposes. In live training environments, however, events are not scripted, but dynamically evolve at their own pace. Whether scripted or not, certain key events are still expected to trigger behaviors that encompass specific learning objectives and measurement strategies

Research on the design and evaluation of performance assessment tools for team training over the past decade has emphasized the need for assessing individual and team performance process and for linking performance process to performance outcomes (e.g., Cannon-Bowers & Salas, 1997; Johnston, Smith-Jentsch & Cannon-Bowers, 1997). Individual and team performance in military environments typically has been measured by the results produced. The final

outcome provides a measure of proficiency required for success. Outcome measures provide an important indicator of combat readiness and, to some extent, of training effectiveness. While outcome measures may be useful for "ball parking" a general training need, they typically do not provide sufficient information to adequately diagnose the source of the problem or prescribe an effective training solution. They do not provide information relating to why the participant was successful or unsuccessful. This is especially important as events get more complex.

In military training environments performance readiness for a mission is typically assessed through the completion of a series of Training & Readiness events selected to build competencies on specific mission related tasks. The specific tasks to be exercised in a particular T&R event are defined in terms of Mission Essential Task List (METL) items and, for the Navy, Naval Tactical Task (NTA) requirements. The METL is organized by operational capabilities and mission or warfare areas within these operational areas (e.g. Anti-Submarine Warfare, Strike). Within each mission area, the NTAs are typically organized around mission phases, providing templates that identify the NTA subtasks required for successful accomplishment of each phase of a mission. The METL and T&R framework provides a top level specification of tasks and scenario events relating assessment and diagnosis to readiness and training needs. The mission phase structure provides the conduit from the relatively high-level NTA elements to the more detailed individual and team task and performance methods and requirements as defined in tactical manuals.

The conceptual measurement matrix shown in Figure 2 includes assessment of both performance outcome and performance process for both the team and the individual and is consistent with team training research and models (e.g., Tannenbaum, Beard & Salas, 1992; Dickinson & McIntyre, 1997). These four categories of measurement can be defined as:

- individual competencies,
- teamwork processes,
- individual event outcomes, and
- team event outcomes.

When an observed outcome is not consistent with expectations, a method is needed for diagnosing the source of the problem in terms of deficiencies in the cognitive, behavioral and teamwork processes that generated the outcome. By assessing performance process in addition to performance outcomes, training interventions, such as post-exercise debrief and the design of future training events, can be focused on identified process weaknesses.

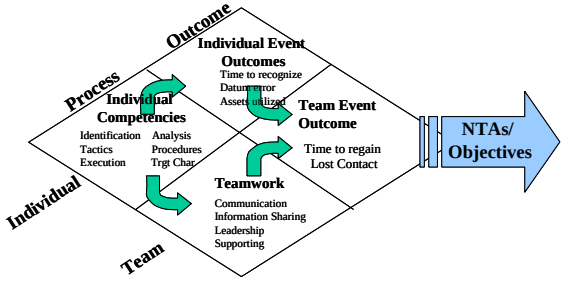


Figure 2. Measurement Matrix

PERFORMANCE EVALUATION PROCESS

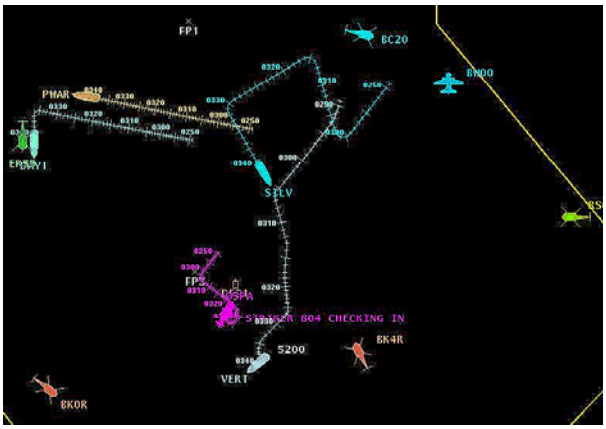
The Fleet Anti-Submarine Warfare Command (FLTASWCOM) trains Atlantic and Pacific Strike Groups in the area of integrated Anti-Submarine Warfare (ASW). FLTASWCOM Integrated ASW Course (IAC) training exercises are flown on both instrumented ranges and in open ocean training areas. They may last up to 12 hours, and involve from 10 to 25 air, surface, and submarine participants. The ASW exercises involve all phases of ASW from search to attack under various conditions. In addition, FLTASWCOM supports additional training exercises, which may cover numerous events over a long period of time or a single event over numerous days. Evaluating these events involves simultaneous real time observation, assessment, and analysis of both individual and multi-platform cooperative missions. The team of evaluators must track, record, assess, and analyze the participants' performance in real time. The evaluators must produce a timely debrief for the participants, provide an in-depth analysis for their leaders, and use the evaluation results to tailor and improve the training process.

Collaborative Evaluation

Effective evaluation of these multi-platform exercises involves evaluators and mentors located at sea, in the air and on shore. Simultaneous evaluations are performed on each individual ship in a STRIKE GROUP, in aircraft participating in the same event and from shore based observation points. Evaluators on ship and in aircraft are focused on individual and team level evaluation The on-board evaluators are subject matter experts (SME) and evaluate team performance through direct observation of system and tactical knowledge, communications, and team performance. The number of evaluators depends on the size of the team being evaluated. Typically, a single evaluator can watch a single aircraft or single watchteam onboard a ship, or staff. The shore-based evaluators are usually

evaluating at the team (i.e., individual platform) and multi-team (e.g., squadron) levels and focus on the tactical picture and how it is evolving. These evaluators observe actions and events, and assess different aspects of performance. These can be different platforms or different performance measures (e.g., teamwork, communications, or tactics).

The shore-based evaluators observe training by viewing range activities on a large screen geoplot or tactical situation display and listening to network communications between team members. The geoplot display is constructed from real-time position data provided by the instrumented range that's updated continuously for each participant. Figure 3 provides an



example geoplot display.

Figure 3. A Typical Geoplot Screen

The evaluation team's primary objectives during the exercise are to track, analyze, and rate performance and annotate and organize performance assessments in preparation for post exercise debrief development. Typically, performance evaluation has involved using a paper and pencil instrument to assign rating scale values on a number of outcome and process variables.

As events occur, it is difficult to predict which will be the key elements that cause problems that may only be recognized later. Therefore, the instructors build event logs that record significant events as they are observed. These logs are instrumental in the evaluation process since they provide the narrative context for performance evaluation. Overall proficiency scores must be supported by specific evidence or examples. Inaccuracies destroy debrief credibility. Therefore, complete and accurate event logs are essential to provide the evidence and context for evaluation. In addition to providing specific evidence or examples to support proficiency comments and evaluations, the event logs from each evaluator are annotated and combined to provide a primary debrief product for Strike Group-level exercises.

Issues

Successive events often occur too fast for each person to observe, record and evaluate them all. This event log development becomes a collaborative process through which the evaluation team communicates to "build the picture" and develop a shared understanding of the evolving tactical situation. It also forms a basis of support for assessment, diagnostic reasoning and making real time modifications to the evolving scenario. However, without a common event log where each evaluator can view and record events as they are observed the data collection process is inefficient, there is overlap in recording observed events, and events can be missed.

Issues with the evaluation process include concerns about standardization, subjectivity, and level of detail. While behavioral anchoring information is provided to encourage a level of standardization in the evaluations across different evaluation teams, different exercises and different difficulty levels, the evaluations still have a degree of subjectivity that promotes variability across instructors and exercise instances. While specific events can be associated with the numeric measures, much of the objective detail that the evaluators use in assessing and diagnosing performance problems is lost when it is collapsed to a single rating value. Enough detail is captured through the event log entries and annotations to provide evidence for evaluations and lessons learned during debrief. However, data to support diagnosis and analysis across exercises has been limited.

Issues involved in debrief preparation include development time, completeness, accuracy, and appropriate content for each level in the hierarchy. Typically debrief occurs at the multi-team and team levels and debrief reports are generated for the management level. At the individual and team level the value of feedback declines as time between the training and the debrief increases. Twenty-four hours after a training event, the trainee's focus necessarily shifts forward. However, it has typically taken at least that long to insure the accuracy and content needed to make the debrief a valuable feedback product for the warfighter. The debrief development process continues the collaboration among the evaluation team members as event logs are merged, additional data sources are integrated, evaluations and comments are reconciled, and the overall themes and individual components of the debrief are developed. Several data products are potentially available for use in debrief development, including event logs from the range

instrumentation, the target submarine's event log, geoplot screen captures, replay segments, and recordings of voice network communications. Organizing and integrating all of these potential sources of information in a timely manner poses a significant challenge to the instructor team.

TOOLS TO SUPPORT THE TRAINING PROCESS

Fleet ASW Command, NAVAIR Orlando Training Systems Division, and Micro Analysis & Design, Inc., have worked together to develop an integrated set of software tools that support the construction, collection, evaluation and analysis of team and individual performance data during live or simulated multi-platform ASW tactical exercises.

Objectives

Objectives of the integrated training support tool included facilitating the performance data collection and analysis process to:

- Reduce the number of evaluators needed;
- Standardize performance measurement by capturing the objective data to support performance measures;
- Automate the rating process as much as possible and desirable; and
- Improve the quality, timeliness, and content of feedback to exercise participants.

Additional objectives included capabilities to:

- Support distributed networked users collaborating to evaluate performance;
- Easily integrate data from stand alone evaluators and from other human and automated sources;
- Analyze and aggregate performance data to rate higher-level categories (e.g., NMETLS, competencies), and provide summary data to support debrief, management reports, and multi exercise trends;
- Support simple to complex performance evaluation requirements; and
- Develop and modify performance measures as requirements change, is a key requirement.

Implementation

The Tactical Warfare Instructor Support Environment (TacWISE) is designed to facilitate performance assessment and analysis tasks to enhance the training process, support post exercise debrief, and improve the user community's ability to link training to performance readiness.

A spiral design approach has been used to develop TacWISE. Interim versions of the tool are delivered periodically to the user community, providing the best means for instructors to become familiar with the software, use it in the training and evaluation context, effectively evaluate its functionality and usability, and provide feedback. The schedule was designed to coincide with major training events. This has provided the evaluation team with the opportunity to use new functions during their most intensive exercise evaluations. This process worked well, resulting in a development process that included instructors as an integral part of the development team.

The integrated capabilities or tools that comprise TacWISE are designed around the five segments of the training process illustrated as Figure 1. Expected events and tasks and associated metrics for a mission type are constructed during the training development process to populate a database for that mission type that can be applied to multiple exercises. Capabilities support assessment and debrief preparation before, during and after the exercise. Analysis capabilities apply to both single and multiple exercises.

Performance Assessment

The assessment module is a single screen interface organized around functions to support multiple evaluators collaborating to record, annotate and integrate performance data, evaluate performance, and develop debrief at multiple levels. The assessment interface is shown as Figure 4. The interface is divided into four components that support the primary functions of collaborative data collection and assessment. These are the event/task selection tree, the event log, performance evaluation tabs, and participant and status lists. The Graphic Users Interface (GUI) can be reconfigured by hiding or resizing interface objects.

The data collection process is organized around recording observed performance events in a timeline based event log. Since it is live training, there are no or few pre-scripted and pre-scheduled events, and there are multiple participants and targets, so events are entered as they are observed using a simple point-and-click process to select from the predefined event tree developed for the particular mission exercise type. The organization of the event/task tree is critical to ease of use and rapid access. One strategy is to use an EBAT approach to organize events under phases and expected tasks under trigger events. The hierarchy shown in Figure 4 is organized by task type, so all reports are

grouped together. A preferences tab allows users to select a subset of the events or tasks that correspond to the areas they will be evaluating. Logged events are time stamped and linked to specific performance agents or reference units (e.g., individuals, teams, or platforms). When evaluators are networked, they share a common event log and can access the events logged by other evaluators to evaluate performance, add comments or make changes.

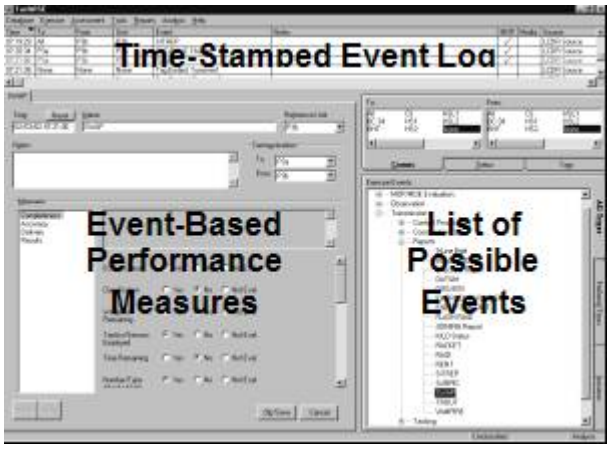


Figure 4. Assessment Interface

The tabs on the status window allow the evaluator to quickly record the source and intended recipient of a communication and to mark time dependent aspects of the exercise such as changes in the prosecution phase, the scene of action commander or other events for a particular target. This provides an easy way to integrate assessments that pertain to a particular phase of the mission for example.

When an event or task is logged, an evaluation tab is opened for the task. The evaluation tab provides fields for comments at the event or task level and at the level of each performance measure. The tab provides access to the measures created for that task. The evaluation tabs contain the predefined measures and any performance criteria or rating scale anchors created for those measures. Any number of measures can be associated with a single event, from a single measure (e.g., check on a checklist, rating on a rating scale), to multiple measures or an entire measurement instrument. This feature provides the flexibility to support a full range of performance evaluation needs, from multi-platform Strike Group exercises to simulator based training and "over the shoulder" evaluation of individual task performance. Evaluators can enter the evaluations whenever they have the opportunity.

Multiple networked evaluators collaborating to develop a common event log allows the team to divide the workload in different ways and gives any evaluator the capability to annotate, amplify and assess any logged event. In addition to the synchronous collaboration that involves discussion of evidence and information seeking to support assessments, the event log promotes asynchronous collaboration where logged events are annotated by different evaluators over time and the record of previous actions supports diagnosis of observed outcome deficiencies.

Data from non-networked evaluators are integrated into the single common database and event log to facilitate reporting and debrief development. Event data captured from digital sources, such as simulator or instrumented range files, also can be merged into the event log and integrated with events that are captured manually.

Constructing Performance Measures

The user community develops the mission task/event tree and performance measures to support evaluation of the specific mission area, training exercise, or operational task. Since different communities and situations may require measurement at various levels of detail, the measure construction process should support assessing performance at a level of analysis sufficient to provide diagnostic assessment and feedback. The user constructs task and event hierarchies at any level of detail and organized in a way that will facilitate the distribution of evaluation tasks among evaluators and promote timely access to frequent and important events. Once an evaluation instrument is created, it is easily tailored to new mission areas.

Performance measures, performance standards, and weighting factors such as importance and difficulty are then constructed for each task to be evaluated. Multiple performance measurement types are supported. For example, report completion can be measured at a number of different levels of detail depending on the measurement objectives. A rating scale can be used to evaluate the completeness of an individual report provide or used as a measure of report completion over the exercise. At a more detailed level, a simple yes or no measure can be constructed for each report to indicate completeness. The most objective information to support assessment, diagnosis and feedback is a checklist of the elements required for each report.

Constructing a new measure involves identifying the type of data and range of values required, selecting the GUI type required to capture the measure data, quantifying the values for analysis, and formatting the measure as it is to appear to the assessor. The process

for creating the individual performance measure is illustrated as Figure 5. Using the sequence of dialogue boxes displayed the author constructs a measure by selecting the appropriate GUI type, creating labels for each possible value, assigning a quantitative value to each label, and formatting the measure using the preview function. In Figure 5, the measure is one required element in a particular report. The measure value labels are Yes, No, and Not Evaluated. The GUI object type is Radio Button. The values are 1, 0 and –1 respectively. GUI types include: check boxes, radio buttons, text boxes, formatted text boxes, pick lists, spinners, and a complex type to compute duration.

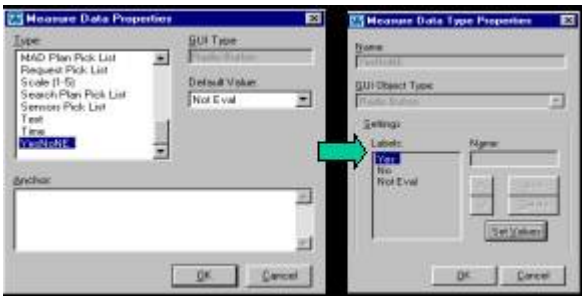


Figure 5. A Portion of the Process for Creating an Individual Performance Measure

Automated assessment of high-level tasks such as the Navy Mission Essential Task List and the Navy Tactical Tasks (NTAs) can be achieved using the *Grouping* function when constructing measures. After defining the specific types of events/tasks to be evaluated, the associated measures, and the standards for converting the measures to ratings, the developer defines Groups that correspond to the NTA tasks, and then assigns each task to one or more of the appropriate NTA task groups. In Analysis Mode, to get ratings for the NTAs, the user creates an analysis definition and in the group filter checks the NTA groups to be included in the evaluation. Overall evaluation and breakdown by groups is selected as the output options. The output will be the ratings for the high level task and for each of the NTAs

Events and measure values also can be populated by extraction from performance data files to provide an automated assessment capability. Figure 6 illustrates functions and user interface to define a parsing procedure for a specific file type. Once the parsing definition is created, performance data can be extracted from this type of file after each exercise, and then used to log events and provide values for performance measures.

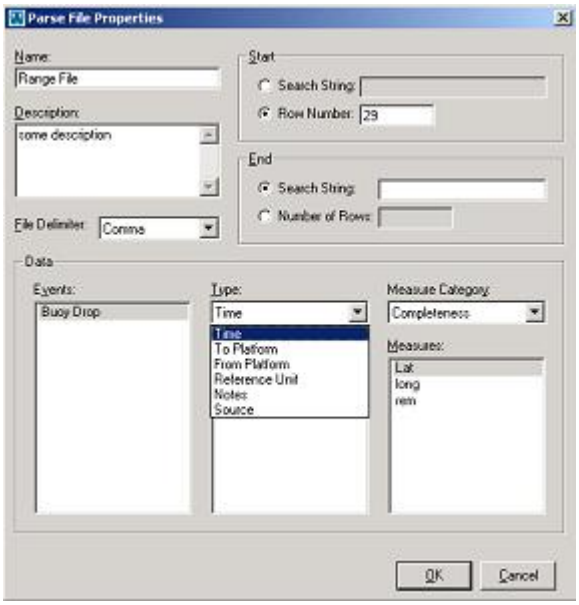


Figure 6. Interface To Define Data Extraction Function For A Specific Data File Type.

Exercise Set-up

Each training exercise has specific information requirements that should be defined before commencement of the exercise. These exercise properties include: the exercise participants and evaluators, the difficulty factors (i.e., conditions of performance) and level of difficulty for each factor, and any predefined occurrences or prompts to the evaluator. In these training events, the participants are not always known in advance and participants may enter or leave while the exercise is ongoing. A means of accurately and efficiently identifying new participants during the scenario, and linking them to their representation in the master database, is required. Figure 7 illustrates a portion of the interface for defining participants.

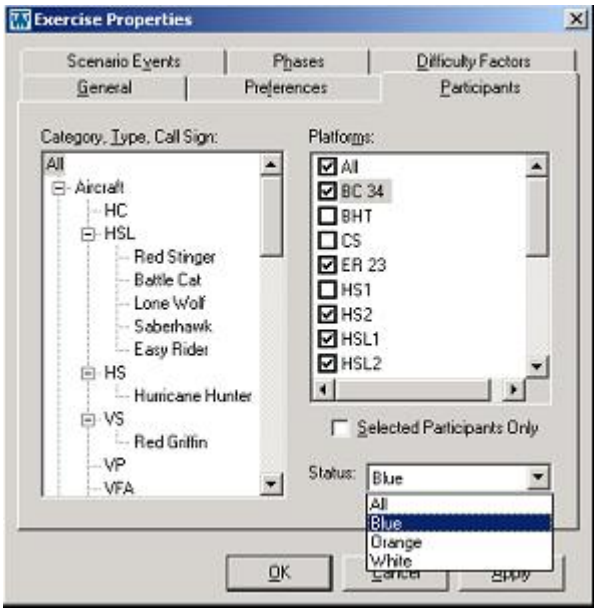


Figure 7. A Portion of the User Interface for Defining Participants

Debrief Preparation

The primary post-exercise activities involve merging data files from non-networked participants and other sources of performance data and collaborating to construct a set of participant debriefs and management reports. A key product for the ASW training community is an annotated event log report. This report highlights the key events that occurred through out the exercise and provides evaluative comments. Software to support an annotated event log report includes event log filtering and sorting functions to develop an event timeline-based debrief product with time and reference unit indexed events that can be annotated and augmented with media files to illustrate performance context. A comprehensive capability to automatically link media files, such as, exercise replay, augmented screen capture from tactical displays, and voice communication files, to time-stamped event entries support the generation of rapid, reliable post-exercise feedback.

Performance Analysis

After the exercise, the goal is to rapidly generate a set of standard analyses that summarize aspects of performance, compare to performance on previous exercises, and perhaps, to develop a custom report for a specific situation. Analysis functions provide tools to support the analysis of performance within and across training exercises. To speed up and standardize the process, summary performance measures are predefined and then generated immediately following an exercise to provide debrief and training or performance management information. Performance analysis tools, such as filters, focus analysis on specific event types, competency groups, participants, sources, engagement phases, targets, and/or time periods. The filters are used to determine the exercise performance data to be included in the analysis. Figure 8 displays a user interface for defining performance analyses for single exercises. A similar user interface provides a capability to display trends, averages and comparisons over multiple exercises.

Two general types of analyses are used: standardized rating scale analyses, and summary analyses. Standardized rating scales are used to develop automated ratings based on objective measures and predefined standards. These support the goal of providing measures at the level of NMETLs, while continuing to provide assessments at a finer-grained level of analysis. The standardized rating scale analysis converts all measures to a user community defined standardized rating scale (for example a 1-5 scale or a 4-1 scale as in the current NMETL approach) using the performance standards created during the measure construction process. Ratings are aggregated up from measures to tasks, events and high-level category groups such as Navy Tactical Tasks or NMETL tasks. For a rating scale analysis, importance and/or difficulty weighting can be included in the computations. Weighting can be turned on or off at each level and weights can be modified based on the relative importance of each measure in the specific analysis.

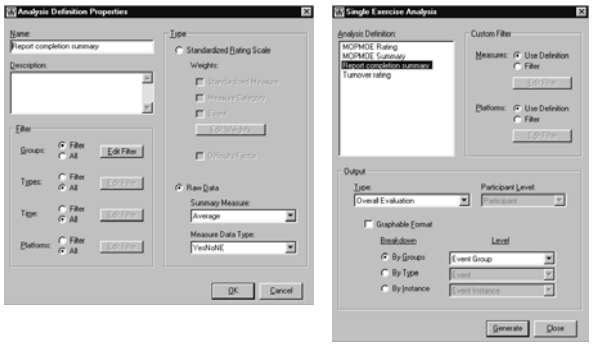


Figure 8. Single exercise analysis functions

The second type of analysis involves the development of summary data. Summary analyses can be quickly and easily generated immediately following an event or exercise based on analysis definitions predefined by the user community. Summary analyses are measures defined over single measure data types, such as checklists, counters, rating scales, etc. Summary measure types include average, proportion

(percentage), sum, and frequency count. Once an analysis is defined, it is stored and can be repeatedly generated after each exercise, or combined with different report output formats to generate different reports. Overall, Group, Event and Participant output types can be generated with the output type representing the highest level for which scores will be computed and displayed. For graphical reports, two levels of detail can be displayed. A number of analyses and report formats can be defined and organized using these analysis tools. An entire set of analyses can then be generated with a single input action after each exercise. The analyses are then previewed and automatically exported to Microsoft Excel© or Word©, or HTML versions. In excel, the predefined graphs are automatically generated. In addition, any predefined analysis can be customized for a particular report without changing the original definition for future reports.

RESULTS AND CURRENT DEVELOPMENT

TacWISE has been instrumental in the successful evaluation of both qualitative and quantitative measures of Navy Carrier Strike Groups and Expeditionary Strike Groups for over three years on both the east and west coasts. TacWISE provides Strike Group Commanders and Sea Combat Commanders (SCC) rapid feedback on their teams' performance during high-tempo, multi-warfare exercises. The independent observations of evaluators located at sea, in the air, and on shore were merged, and feedback was provided to underway participants within four to six hours, and within two hours for shore-based exercises. The rapid feedback received has been critical to Warfare Commanders. After receiving debriefs based on data collected, war-fighting guidance was rewritten, and then tried again the following day. This ability to rapidly test and modify real world tactical policy is critical to mission success.

The training value involves more than just providing a numerical output to indicate how well something was done but also how and why it was done. Numeric (NMETL based) grades are given based on observed and measured parameters. These parameters are very rigid and do not allow for a lot of evaluator discretion. During an exercise the SCC has to make tactical choices. These choices may require him to sacrifice the "numerical grade" in one area for the accomplishment of a separate objective in a different mission area. This been observed time and again and the numerical grade suffers because of it. However, through the collection of not only hard numerical data but also through SME comments, tagged to specific measures the disparity between grades and outcomes are quantified.

Evaluation of staff performance must be viewed from both direct observation of the staff and through the actions and performance of subordinate units. This is an area in which the collaboration among and integration of data from multiple evaluators and sources using TacWISE has excelled. Through the use of numerous evaluators spread throughout a Strike Group a complete picture of team performance is obtained. Common metrics are used to evaluate various groups of people throughout the force. These diverse sets of data are then merged to give a complete picture of the overall team's performance and can be used to identify weak areas or weak links in a team. Another valuable feature has been the ability to adapt the type of data that are collected. As training and assessment evolve, it is easy to adapt and modify measures or if needed the entire measurement architecture to meet the evolving needs.

In order for a tool to be effective, it has to be easily trained and more importantly easily used. TacWISE training was conducted in numerous ways over the past three years. It has occurred in a structured classroom setting, in an office, and underway enroute to an exercise area. It has even been conducted over a phone while both the instructor and the evaluator were at sea on different classes of ships. Average time to train an evaluator in the use of TacWISE was 40 minutes. Evaluator training included the entering of both graded and non-graded events, editing the timeline, adding new units during an exercise and the saving and zipping of databases. Personnel of all rates and ranks, from E-5 through O-6, and all Navy warfare communities have been trained. The trainability and usability of TacWISE for assessment has been demonstrated through its continued effectiveness over a number of years. Figure 9 depicts networked instructors recording observations and assessing performance during a live ASW exercise.

Figure 9. Networked Instructors Record Observations and Assess Performance During a Live ASW Exercise.

Current Developments

Two developments in the Fleet ASW training framework have spurred additional need for enhancements to the integrated training support tools. As the distribution of evaluators has expanded to include collaborators on shipboard and aircraft, the need to have evaluators either connected to the TacWISE server and collaborating with other evaluators over the Secret Internet Protocol Router NETwork (SIPRNET) or be operating as stand alone evaluators and uploading their assessments via SIPRNET to be merged with the common event log and assessment database. This has led to the development of a browser-based version of the TacWISE assessment module that connects to a server via the SIPRNET.

The second development is the emergence of a simulation-based ASW training environment that will eventually allow operational and virtual systems to train over a common network. This has led to the need for a DIS/HLA capable version of TacWISE with automated data collection and assessment functionality that will support and contribute to the collaborative evaluation process.

ACKNOWLEDGEMENTS

This work is supported with funding from the Fleet ASW Command and the NAVAIR Small Business Innovative Research program. The ideas expressed herein are those of the authors, and are not necessarily endorsed by the funding agencies or reflect the official position of the organizations with which they are affiliated. This material is approved for public release; distribution is unlimited.

REFERENCES

Cannon-Bowers, J. A. & Salas, E. (1997). A framework for developing team performance measures in training. In M. T. Brannick, E. Salas, & E. Prince (Eds). Team performance assessment and measurement: Theory, methods, and applications (pp. 45-77). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert decision making teams. In N. J. Castellan, Jr. (Ed.), Current issues in individual and group decision making (pp. 221-246). Hillsdale, NJ: Erlbaum.

Dickinson, T. L. & McIntyre, R. M. (1997). A conceptual framework of teamwork measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp.331-356). Mahwah, NJ: Erlbaum.

Johnston, J. H., Smith-Jentsch, K. A., & Cannon-Bowers, J. A. (1997). Performance measurement tools for enhancing team decision-making training. In M. T. Brannick, E. Salas, & E. Prince (Eds). *Team performance assessment and measurement: Theory, methods, and applications* (pp. 311-327). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Tannenbaum, S. I., Beard, R. L., & Salas, E. (1992). Team building and its influence on team effectiveness: An examination of conceptual and empirical developments. In K. Kelley (Ed.), *Issues, theory, and research in industrial/organizational psychology* (pp. 117-153). Amsterdam: Elsevier.