

MEASURING HUMAN PERFORMANCE IN MULTINATIONAL DISTRIBUTED EVENTS: LESSONS LEARNED FROM THE FIRST WARFIGHTER ALLIANCE IN A VIRTUAL ENVIRONMENT (EXERCISE FIRSTWAVE)

Antoinette M Portrey
Loren Keck
Brian Schreiber
Winston Bennett, Jr.
Air Force Research Laboratory
6030 S. Kent Street
Mesa, AZ 85212 USA

E-mails: Antoinette.portrey@mesa.afmc.af.mil, loren.keck@mesa.afmc.af.mil, brian.schreiber@mesa.afmc.af.mil,
Winston.bennett@mesa.afmc.af.mil

KEYWORDS

Military, Behavioral science, Decision Support Systems, Statistical analysis, Performance analysis.

ABSTRACT

The Air Force Research Laboratory (AFRL), Warfighter Readiness Research Division, is conducting research to measure and track Warfighter performance of knowledge and skills from an individual level to the Command and Control (C2) level, within both high fidelity distributed simulation environments and live training environments. One critical development is a performance measurement system, the Performance Evaluation Tracking System (PETS), which captures data necessary for competency-based assessment and evaluations, end-user performance feedback, simulator technology developer validation, and for researcher and program manager evaluation of training techniques and technologies. PETS is comprised of various components that allow integration into focused benchmark studies, large scale distributed coalition operations, and live-fly training scenarios. This paper will describe the application of PETS during the multi-national Warfighter Alliance in a Virtual Environment (FirstWAVE) coalition simulation event; discuss the lessons learned, the impact of different levels of abstraction and representational levels of correlation in the data, and the challenges facing both researchers and operational personnel for both producing and using performance data and methods.

INTRODUCTION

The dramatic transformation of America's strategic environment has a significant impact on today's Warfighter and how we prepare for combat operations. Emphasis remains on shifting from deliberate to adaptive war planning, and from permanent organizations and large hierarchies to smaller, highly-distributed joint and combined forces (United States. Office of the Under Secretary of Defense (Personnel and Readiness), 2004). A distributed training environment that constitutes global, multi-national networks of constructive computer simulations, man-in-the-loop virtual simulators, and live forces at instrumented ranges is key to achieving military

performance objectives and meeting current and future Warfighter demands (United States. Office of the Under Secretary of Defense (Personnel and Readiness), 2004).

As the necessity of military commitments steadily increases in today's world, joint and coalition training continues to play a critical role in Warfighter preparation. Currently, Warfighter forces regularly participate in large scale "live-fly" joint exercises such as Red Flag and Maple Flag to meet training objectives. This type of high-visibility training event often involves coordination of multinational forces and typically occurs several times a year, helping to keep the Warfighter combat ready.

Distributed Mission Operations (DMO) in joint and coalition simulation environments

As the number of "live-fly" events increase, it has caused strains on the use of equipment, increased maintenance costs, and has demanded a significant growth in the amount support personnel required. These logistical and economic challenges have stimulated a significant need for simulation-based training. The Distributed Mission Operations (DMO) environment is currently used to prepare the Warfighter and augment live-fly events by providing the ability to practice or learn skills and techniques that can be validated and sharpened through live training and real world use. DMO based training fills the need to "train the way we intend to fight," and has become a critical requirement in Warfighter preparation, enabling the Warfighter to obtain and maintain combat readiness through joint and coalition mission rehearsal in operationally realistic environments.

Simulation technology today allows Warfighters to participate in a continuous training cycle and maintain a high state of combat readiness by using cost-effective simulation alternatives in conjunction with live-fly operations and training missions. The rapid advancement of networking technologies, increase in bandwidth capabilities, and continued improvement of protocol standards/architectures such as Distributed Interactive Simulation (DIS), High Level Architecture (HLA) and the Test and Training ENabling Architecture (TENA) have all

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 24 OCT 2005	2. REPORT TYPE Proceedings	3. DATES COVERED 01-01-2003 to 30-09-2005			
4. TITLE AND SUBTITLE Measuring Human Performance in Multinational Distributed Events: Lessons learned from the First Warfighter Alliance in a Virtual Environment (Exercise First WAVE)		5a. CONTRACT NUMBER F41624-97-D-5000			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER 62205F			
6. AUTHOR(S) Antoinette Portrey; Loren Keck; Brian Schreiber; Winston Jr Bennett		5d. PROJECT NUMBER 1123			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) L-3 Communications,6030 South Kent Street,Mesa,AZ,85212-6061		8. PERFORMING ORGANIZATION REPORT NUMBER ; AFRL-RH-AZ-PR-2005-0001			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory/RHA, Warfighter Readiness Research Division, 6030 South Kent Street, Mesa, AZ, 85212-6061		10. SPONSOR/MONITOR'S ACRONYM(S) AFRL; AFRL/RHA			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RH-AZ-PR-2005-0001			
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES In Proceedings of the 2005 European Simulation and Modeling Conference (pp 354-366), held 24-26 October 2005, in Porto, Portugal; also includes briefing charts					
14. ABSTRACT The Air Force Research Laboratory (AFRL), Warfighter Readiness Research Division, is conducting research to measure and track warfighter performance of knowledge and skills from and individual level to the command and control (C2) level, within both high-fidelity distributed simulation environments and live training environments. One critical development is a performance measurement system, the Performance Evaluation Tracking System (PETS), which captures data necessary for competency-based assessment and evaluations, end-user performance feedback, simulator technology developer validation, and for researcher and program manager evaluation of training techniques and technologies. PETS is comprised of various components that allow integration into focused benchmark studies, large-scale distributed coalition operations, and live-fly training scenarios. This paper describes the application of PETS during the multinational Warfighter Alliance in a Virtual Environment (First WAVE) coalition simulation event; discuss the lessons learned, the impact of different levels of abstraction and representational levels of correlation in the data, and the challenges facing both researchers and operational personnel for both producing and using performance data and methods.					
15. SUBJECT TERMS Lessons learned; Exercise First WAVE; Human performance; Multinational; Distributed events; Virtual environment; Military; Behavioral science; Decision support systems; Statistical analysis; Performance analysis					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Public Release	18. NUMBER OF PAGES 18	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

contributed to an environment where large-scale, multi-force DMO joint/coalition exercises have become a reality.

Current development and integration of live, virtual, and constructive systems for training, mission rehearsal, and operations support has emphasized the need for more complete simulation network interoperability among joint and coalition forces. Specifically, improvements in protocol standardization and acceptance of comprehensive performance measurement systems stimulate interoperability and help to validate the return on investment that joint operations provide. With the change in training scope from traditional to transitional training, including environments such as special operation forces, urban operations, and joint/coalition, measuring human performance gains in these environments is critical if we are to understand our Warfighters' readiness levels.

THE NEED FOR ROBUST PERFORMANCE MEASUREMENT SYSTEMS

The Air Force Research Laboratory in Mesa, AZ (AFRL/Mesa), is a research organization chartered to implement and/or evaluate methods to improve Warfighter readiness, has been commissioned to do human performance assessment research with respect to USAF Mission Essential Competencies. One of the primary goals of this research was to investigate the ability to assess Warfighter performance in a DMO environment. This capability, if done properly, would allow the research division to collect data on any number of diverse studies.

That initial line of research resulted in a DIS compatible tool capable of collecting objective data on the DMO assets within the local area network (LAN) site at that division. As an evaluation tool, the proof-of-concept Performance Effectiveness Tracking System (PETS), see references (Schreiber et al. 2003, Watz et al. 2003, and Watz et al. 2004), emphasized tracking human performance data in an empirical (i.e., scientific) manner for evaluating training techniques and technologies.

Performance Effectiveness Tracking System in controlled Human Performance Studies

The original PETS software component, developed at AFRL/Mesa in 2001-2002, was designed as a proof-of-concept human performance measurement tool that could collect over 80-100 'core' air-to-air and air-to-ground combat performance measures in real-time from a distributed network using the DIS protocol. This original PETS software architecture is currently used by AFRL/Mesa to provide a number of previously unavailable objective measures which significantly increase the effectiveness and quality of DMO training and research.

The early successes of this prototype system allowed AFRL/Mesa to automatically capture and store kill ratios, weapons employment, and other skill-related metrics on over 400 fighter pilots participating in multi-player DMO exercises at AFRL/Mesa's site since January 2002. The

usefulness of this empirical data facilitated a number of recent human performance DMO studies, see references (Gehr et al. 2004, Schreiber et al. 2003b, Stock et al. 2004, Portrey 2004, and Schreiber et al. 2005), and data collaboration and study efforts from both industry and academia.

Due to the a) usefulness of PETS-generated data for studies, b) piqued warfighter interest in its potential for feedback, and c) AFRL's desire to expand the scope of DMO exercises (e.g., joint and coalition), the PETS performance measurement capability was then sought by AFRL/Mesa for DMO exercises involving assets *outside* the local simulation network—essentially an expanded scope of the original human performance assessment research from the individual warfighter towards the Command and Control (C2) level. These broadened requirements increase the emphasis to collect data on any entity involved in a DMO network and report performance metrics as feedback (i.e., an increased emphasis on a new "observational" measurement capability) to Warfighters and their instructors. Furthermore, it provides the ability to assess performance from a group or "team" perspective. To fulfill these new, larger objectives, additional components under the PETS architecture was required.

PETS in Coalition Environments

As part of overall Performance Effectiveness Tracking System, the beta-version "PETS²" component was developed in 2003-2004. This component was designed to address the flexibility and configurability issues inherent in the original proof-of-concept version (Schreiber 2005). More specifically, the need for an architecture that went beyond the previously "empirical-only" LAN architecture to a more robust and flexible wide-area "observational" architecture. The PETS² software functionality was also driven by requirements for increased warfighter performance feedback and higher-level, aggregate measurement capabilities. A high-level view of the PETS² design is illustrated in Figure 1.

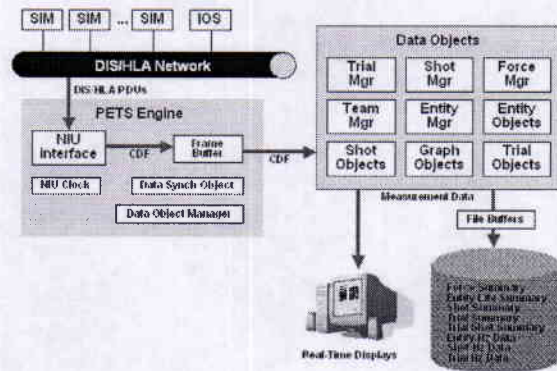


Figure 1: PETS² Architecture

The PETS² project was designed as a modular, multi-threaded application, capable of robustly handling high volumes of networked entities (Portrey et al. 2005 and

Watz et al. 2003). The preliminary version of this system is capable of handling several hundred entities within a DMO environment. It provides interfaces used to customize real-time informational tabular/graphical displays, such as a force demographic summary. This system also employs a multi-tiered lookup system for additional user supplied internal state data that may be unavailable on the network, thus making it more interoperable to any networked entity by removing the dependency on custom data requests.

PETS² development has the ability to calculate measurements (see Table 1) at the team, inter-team (package), and teams-of-teams (force) levels, which can theoretically extend the potential measurement capabilities of PETS² up to, and including, objective data at the Command and Control (C²) level. PETS² is able to evaluate overall mission performance on all entities including both man-in-the-loop and of Computer Generated Forces (CGF), allowing the trainer to assess the entire picture from both the friendly and enemy perspective.

Table 1: Types of DMO objective training effectiveness data

Kill ratios
Missile hit ratios / Targets destroyed
Distance of misses
Time spent within Minimum Abort Range
Clear Avenue of Fire measures
Missile measures, such as launch altitude, range, loft angle, mach, g-loading
Bombing (absolute error, left/right and long/short)

PETS² was designed from the ground up to support DIS standard data (currently DIS 2.04) and HLA via the MAK Gateway. Although this may limit the amount of special (i.e., non-standard) data that a particular simulator emits, it allows PETS² to work with any simulator that conforms to the DIS 2.04 standards. However, in order to support more complex measurements, PETS² has an array of user input screens that allow configuration of non-standard data such as weapons load, and various initial entity state conditions. PETS² is designed as a “horizontally integrated” application that provides as much vertical depth as the protocol and user-provided input allow. For example, this could allow participants in a Joint or Coalition exercise to capture additional data on certain entities on a site-by-site, entity type, or force affiliation basis.

To help take advantage of non-standard network data, PETS² has the capability to add “plug-in” entity modules that can handle custom information packets “passed through” the network via the DIS protocol (such as DIS Set Data PDUs, etc.). These custom entity modules are physically separate from the PETS² core and can be added or removed as the functional need or security requirements permit. For example, if a particular site has a new, cutting edge, flight simulator operating at a higher classification

level, that site can create a custom entity module that is kept and implemented only at their site for collection of special measurements while standard measurements are collected at all other locations on the network.

A REVIEW OF PERFORMANCE MEASUREMENT AT FIRSTWAVE

Joint and Coalition exercise environments usually consist of diverse but limited simulation systems (concentrating mainly on individual aircrew training) and a wide array of internal and external networks. This diversity limits the effectiveness of a performance measurement system in providing information to the units and command staff. Performance outcomes measured are inversely proportionate to the number of simulation systems and the complexity of the teams within the exercise environment (see Figure 2).

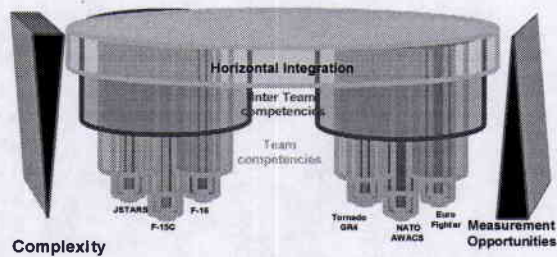


Figure 2: Performance Measurement Complexity

The First Warfighter Alliance in a Virtual Environment (FirstWAVE) was a multi-national technology demonstration that showcased the promise and capability of implementing HLA in a large scale, wide-area training event. This event was the first of its kind and never in the history of military simulation has such a large-scale, global event been attempted. As such, FirstWAVE provided the ability to determine the serviceability of the simulation protocols and to evaluate the possibility of performance measurement on high-magnitude, internationally coordinated DMO training exercises.

The primary objective of FirstWAVE was to establish, demonstrate, and document the potential of distributed simulation to enhance training for NATO combined air operations. The Mission Training through Distributed Simulation (MTDS) demonstration took the form of an air coalition training exercise involving front-line aircrew, with a multi-national White Force, (a team of military personnel with expertise in NATO air operations) who supervised the event. The MTDS exercise participants consisted of Canada, France, Germany, Italy, Netherlands, and the United Kingdom (see Table 2); with US Air Force personnel supporting the engineering development, connectivity and testing, scenario planning and management, and training effectiveness data collection.

Table 2: Exercise FirstWAVE Entity List

Blue Assets (France, Germany, Italy, Netherlands, UK)
Tornado GR4 2-ship (CGF)
(2) F-15C 4-ship (CGF)
F-16CJ 4-ship (CGF), F-16C 4-ship (CGF)
F-16 MLU 3-ship and 1-ship
FAC and E3C
Mirage 2000C 2-ship
Eurofighter Typhoon 1-ship
(2) CF-18A singletons and (1) two-ship
Red Assets (Canada and UK)
(2) Red C2/ Red GBAD
MiG 29 2-ship
Red GCI/Red Air CGF

Observation of FirstWAVE Issues

The FirstWAVE network traffic was recorded for the purpose of mission support, playback/debrief and automated performance measurement. The PETS architecture used the recorded log files as raw data input to generate assessment metrics. While the FirstWAVE event was deemed highly successful and an invaluable training demonstration overall, the recorded data log files, unfortunately, contained instances of incomplete or inaccurate data. Many data recording errors were identified and attributed to issues such as long-haul network instability, simulation errors, etc. However, as there are several potential causes for many of these types of errors (e.g., excessive network latency or simulator specific deficiencies) it is impossible to isolate the cause for all of these data errors without extensive testing that is beyond the scope of this effort. Ignoring the network related errors, there were several sources of errors introduced by the simulation systems themselves. For example, errors encountered in the data log files included:

- Missing fire or detonate information (e.g., targeted entity) normally included upon munitions release and munitions detonate.
- Missing entity identification when transmitting communication
- Invalid entity identification
- Incorrect entity status
- Unrealistic entity state changes

The problems associated with the simulation data recorded during FirstWAVE, made it impossible to accurately validate all resulting performance metrics (in order to provide accurate and valid performance measurement results, pre-event emphasis *must* be placed on verifying the accuracy, standardization and protocol compliance of all simulation systems that participate in the training event). However, the PETS² architecture was used to analyse the data and illustrate the types of performance metrics that can be obtained automatically in an MTDS-

type training event and to test the robustness of PETS² in coalition environments.

Deficiencies Found Through Raw Data Analysis

The nature and complexity of simulation systems and distributed environments, means that missing or incomplete network data are likely to plague such large scale events. Compounding the problem, many simulation environment developers only include or emit data that is significant to their own site or purposes, without regard on how this affects their overall visibility in a large scale DMO exercise. The following sections provide specific examples and the errors and impact from problems observed in FirstWAVE objective data collection.

Our first example data are missing fire and detonate information (e.g., targeted entity) normally included upon munitions release and munitions detonate. For example, many simulators do not emit shooter/target information for bombs. This causes a problem when determining performance at time of launch, measuring data during the fly-out, or doing real-time performance comparisons with other simulators on the network. These omissions can cause critical air-to-ground performance measurements to be incomplete or non-existent. For example, this information may be translated by the PETS² system into an unknown entity that released the weapon and impact a great number of metrics (who fired, what speed/altitude fired, who fired at, subsequent calculation of kill ratios, etc.). Though logic code within PETS² attempts to account for these types of missing network information, absent data still poses serious problems—most notably is increasing the probability of producing erroneous data as output.

A second example is the inconsistency of entity identification. There must be clear standards regarding DIS header information and enumerations that should be adhered to by all simulations in the exercise. Accurate identification of platform models is essential to utilizable performance measurement. Entity marking string usage was inconsistent among the simulation systems adding complexity to entity identification within the performance measurement system. The complexity significantly increases the time to analyze the data for entity performance and outcome measures. As another missing data example, some entities did not attach ID information when transmitting communication. As a direct consequence, the PETS² system could not perform its normal calculations for determining how often participants from each force were “stepping” on one another’s communication.

A third example is related to inconsistent timeout values; another major issue that can affect outcome measures. For example, many sites will have different timeout values for air and ground entities. This common change done to improve the performance of the network can hinder performance measurements by impacting the fidelity and latency of the output. Due to the timeout differences, single

entities could show up as multiple entities within the same exercise confusing the output results.

In addition to the lack of "standard" data and identification, simulators also vary in the consistency of their programmed models and platforms. This was also observed in the FirstWAVE analysis. Observed inconsistencies included flight models such as missile and aircraft performance and shots with either mission or invalid identification. With respect to various models, it was clear that some of the models had significantly unrealistic ranges or maneuvers. One aircraft, for example, repeatedly performed turns in excess of 35G, spiking as high as 100G. Another aircraft shot two missiles with a neutral force ID as the identifying force ID (should only be red or blue—in this case, blue). In violation of interoperability standards, there were numerous instances where the fire and detonate event IDs for a given weapon did not match, causing the PETS system to create duplicate munitions.

There were also examples of shots with either missing or invalid target IDs (e.g., 142.0.0), which can create a misreporting of the shot result. As yet another example, entities were indicating that they were destroyed when indeed they were not. This last example, likely attributable to network delays tripping the inactive bits in the entity state PDU, causes complications for any software system to automatically count how many unique entities actually participated in the scenario, which again impacts the automatic calculation for important statistics such as force size or kill ratios. This inconsistency can significantly affect performance measurements such as shot/kill ratios. All of these inaccuracies or inconsistencies serve to lessen the capability to extract meaningful training performance data from issues further push down the quality of the raw data log file far from that necessary for automatic analysis.

Exercise Protocols and Rules of Engagement

As with any study involving operational personnel, additional data discrepancies were observed during mission execution. For example, humans intervened to reposition entities, override simulated kills, and entities were regenerated during simulation. Protocol issues such as the use of shields and regeneration rules need be set, agreed upon, and followed consistently to ensure that common measurements such as shot/kill ratios are recorded accurately. It is extremely difficult to assess overall mission performance when the "fair-fight" exercise doctrine is compromised. For example, a particular site determined that they would use shields on their aircraft in order to maximize time in the simulation event, bypassing protocol. Because they were perceived as invincible by both themselves and other players, this caused the pilots and other participants to not respond in a realistic manner, thereby compromising the training quality of the data. Worse yet, any shots taken and detonated on an entity with shields severely (and negatively) impacts the accurate assessment of "kills" – a key performance criterion in combat simulation. In order to accurately train and assess

performance at the C² level, all players in an exercise must adhere to common rules that map directly to real world situations.

Objective Performance Measurement: Process and Analysis

Specific to exercise FirstWAVE, data screening was accomplished and problems were identified and resolved to the best of our ability and the log files were submitted to the beta version of the PETS² component (the "observational" components of the PETS architecture) for analysis. A number of the issues described above necessitated some patches to be written in an attempt to output accurate objective data. Since a considerable portion of log file data were missing, the beta version of the PETS² component revealed its need for continued development to compensate for MTDS exercise log file. As a direct result of variability in incomplete fire/detonate PDUs, the PETS² component duplicated some shots in the data (e.g., a missile without a munitions ID in the fire PDU would mismatch with entity state PDU during fly-out; as a result, two missiles were tracked). Also, as part of the first generation PETS system (which operates on only 24 or fewer airborne entities), substantial logic was programmed to account for missing data and other network anomalies.

As one example, if a fire PDU went missing, the original PETS² architecture would do proximity calculations on nearby aircraft to the just-appeared missile entity (on its first appearance entity state PDU) and determine/assign the most appropriate firing entity. By applying this type of logic to a variety of missing, non-standard and unusual circumstance data, virtually all assessment data could be performed and output accordingly.

In the beta PETS² component used for FirstWAVE, only a fraction of extra logic had been incorporated. This resulted in the first major problem with the PETS² analysis; the PETS² component treated much of the missing raw log file data as missing and therefore reported a high percentage of missing values in the output. Additionally, we discovered that the NIU used within the PETS² component to read the log files may be dropping/missing network packets under high load when rerunning the recorded log files, thereby exacerbating the first problem. As an overall result, a considerable proportion of end output descriptive data points resulted in missing values.

During the last two days of the FirstWAVE exercise, the beta-version PETS² component was barely robust enough to collect and calculate data. Most metrics normally captured and reported by the PETS² beta software simply could not be calculated, at least with any degree of accuracy. Given the issues mentioned previously in the paper, the analysis became more of a human endeavor than an automated one. Writing additional software specifically to pull out and analyze certain types of PDUs (e.g., fires, detonates, entity states) was accomplished to aid in validation and to obtain a degree of confidence in the limited data reported here. Finally, with a "true" sample size of just two scenarios,

there was insufficient power for any type of inferential statistics to be used as part of an effectiveness evaluation.

After processing the log files, we first counted Blue and Red force size/employment. This was done by examining the total unique participating entities that occurred during days 2 and 3 of FirstWAVE. A unique participation is defined as a unique life or entity for the Red or Blue force (e.g., an aircraft or SAM site). If an aircraft came onto the network and then left the network, this was counted as one "life or entity." Most frequently, this on/off network pairing is a natural consequence of initialization and death or end of scenario circumstance. As stated previously, for accurate assessment, we advocate for not using shields or regeneration during a scenario. Since protocol deviated from this recommendation during FirstWAVE, we counted regeneration or a hit entity with shields "on" as a new valid life/entity (i.e., as if a "replacement" had been called in during combat); this was determined as the most accurate method of calculating force size/employment as it would relate to actual combat.

Ordinarily the valid force number would be equal to or slightly greater than the number of entities revealed in the log file number (number of regenerations/shield hits typically constituting the upside discrepancy; e.g., Blue force, day 3). For day 2, these results were not found; upon further review of day 2, eighteen red entities were created, but then removed from the network prior to the actual push, which was recorded by PETS². The PETS² beta software consistently recorded more lives/entities than the true valid number predominantly due to excessive latencies creating time-outs. When these entities re-emerged, PETS² recognized them as a new, unique life/entity; while the true "intent" of the exercise was that they never should have left and were the "same" entity. Due to this and a high proportion of missing target information in the log files, we could not rely on any of our kill and kill ratio statistics.

Once the intensive human-assisted checking and filtering was accomplished, to our positive surprise, however, the assessment system and methodology did produce *some* usable data and valid measurements, despite the log file abnormalities (NATO 2005). This limited success with automated objective analysis reveals promise for automatically assessing future MTDS exercises. However, considerable quality improvements must be made if reliable, valid, automatic data are to be expected from future exercises. Specifically, network data needs to be complete, accurate, and timely, and exercise directors must adhere to protocols of realism. As of today, only smaller, more localized distributed simulation exercises present network data of sufficient quality to afford completely automated objective analysis. Larger exercises require further maturation for automated objective analyses—a level of robustness we are planning over the next several years of research and demonstrations.

THE NEED FOR STANDARDS FROM SIMULATION COMMUNITY

To effectively study individual, team, and force performance within the DMO environment, simulation systems need to ensure that standards are adhered to. Although the protocols are standard, due to real world requirements, each simulation system might be different in the variations of data packets that are emitted or handled. For example, interoperability is frequently adversely affected by missing or incomplete DIS PDUs, incorrect DIS enumerations, or non-standard data such as proprietary voice data implementations, or use of unsupported tactical data links. Better use of network protocols and improved interoperability awareness is needed to present a clear picture for performance measurement of all participants within the exercise. Most simulations systems output some degree of non-standard or incomplete data during an exercise; while standards exist, installations do not strictly adhere to them. Non-adherence simply should be resolved across all DMO installations, not just for performance measurement, but also for improved large-scale exercise interoperability.

In the process of following current network protocols, the extension of standardized data by using existing and new capabilities of the DIS/HLA standards would be beneficial in establishing a more concise picture of performance for individuals, teams, joint and coalition exercises. The main purpose of this extension would be to provide additional data needed to perform C²-level Warfighter performance assessments and training, which is a significant reason distributed simulation environments were created in the first place.

A significant portion of the extended data would be in the form of "internal" state data, examples of which include, but are not limited to, switch positions, radar modes, radar tracking lists, weapons load, throttle position, targeting information for bombs, and others, all of which are not currently part of the DIS standard or base HLA RPR Federation Object Model (FOM). This type of information is needed to provide a performance analysis system such as PETS² insight into what the simulation operator (pilot, weapons officer, etc.) is doing. Having internal state data as part of an established protocol-level standard would enable performance measurement systems such as PETS² to effectively gather C²-level data and analyze performance across multiple DMO sites. Currently, standardized measurement tools such as PETS² are limited in collecting performance measures from multiple sites, as the enumerations or data structures used to pass such internal state data are not consistent from site to site (Watz et al. 2003).

One simple over-arching rule must govern all simulation exercise protocol—realism. That is, if it is not possible in war, the console operator(s), mission directors and simulation operators should not be allowed to do it in a simulation (e.g., use shields, reload weapons mid-air, "freeze" or "static" fuel burn, etc.). If realism is

emphasized when performance is calculated for an exercise, the results have much higher credibility and relationship to that of an actual battle—after all, the original impetus behind training in DMO was to translate performance gains into actual battle! The paradox is that the DMO community continually strives for realism in all technology aspects (e.g., visuals, flight models, missile models, CGFs), but appears to largely disregard exercise realism from the IOS. Of course, occasional uses of shields or other functions may be highly desirable during early stages of training, but those non-realistic scenarios must be the exception, not the norm (Portrey et al. 2005).

CONCLUSION

There are a number of valuable lessons learned which we have taken from the Exercise FirstWAVE event. All of the lessons learned provide valuable insight into how future large scale events can and should be structured, managed, and evaluated. To respond quickly to the dynamic challenges of today's environment, training, rehearsal and exercise approaches and doctrine must remain flexible, operationally focused, mission effective, and integrated with real-time, globally distributed mission rehearsal capabilities. To do this, available networks for mission rehearsal, simulation, and just-in-time training must be continuously modernized and utilized; and performance metrics need to be systematic and complete to quantitatively demonstrate improve operational effectiveness, both individually and collectively.

This global training model emphasizes the necessity for following strict standards such as DIS, HLA (with a common FOM such as RPR) or Test and Training Enabling Network Architecture (TENA), and the need to establish new standardized data to provide better performance feedback to the training and operational communities. There are many DMO installations throughout the Modeling and Simulations (M&S) community, and currently most of these sites have no method of objectively assessing the degree or amount of knowledge transfer that takes place when Warfighters train in these virtual environments. With the training community developing performance assessment systems such as PETS² to address joint and coalition training issues, there is an increase need to broadcast internal state information on the simulation network for acquisition and analysis. Providing a dynamic, capabilities-based training for the Warfighter must be a joint effort between the simulation, operational, and training communities if it is to succeed in today's environment.

REFERENCES

Gehr, S.E., Schreiber, B.T., & Bennett, W. Jr. (2004) Within-Simulator Training Effectiveness Evaluation. In *2004 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

NATO Research & Technology Organization: Mission Training via Distributed Simulation (MTDS) Task Group. (2005). *Mission Training via Distributed Simulation and First WAVE: Final Report (SAS-034/MSG-001)*: Brussels: Author.

Portrey, A. M., (2005) *The Escape-G Metric: A Concise Measure for Air Combat Maneuvering Performance*. Unpublished master's thesis, Arizona State University, Tempe, Arizona.

Portrey, A. M., Keck, L. B., & Schreiber, B. T. (2005). Challenges in Developing a Performance Measurement System for the Global Virtual Environment. Paper presented at the 2005 Spring Simulation Interoperability Workshop. [Electronic Version] Paper available as 05-SIW-100.doc.

Schreiber, B.T. (under review). *Transforming Training: A Perspective on the Need and Payoffs from Common Standards*. Manuscript submitted to *Journal of Military Psychology* for publication.

Schreiber, B.T., Bennett, W. Jr., Schurig, M., & Gehr, S.E. (in preparation). *Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study*. Manuscript in preparation.

Schreiber, B. T., Watz, E., Bennett, W. Jr., & Portrey, A. (2003). Development of a Distributed Mission Training Automated Performance Tracking System. In *2003 Proceedings of the Behavioral Representations in Modeling and Simulation (BRIMS) Conference*. Scottsdale, AZ.

Schreiber, B. T., Watz, E. A., & Bennett, W. Jr. (2003). Objective Human Performance Measurement in a Distributed Environment: Tomorrow's Needs. In *2003 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Stock, W., Schreiber, B.T., Symons, S., Portrey, A., & Bennett, W. (in preparation). *Can Performance of Air Combat Teams in Distributed Mission Operations Training Exercises be Described by Signal Detection Parameters?* Manuscript in preparation.

United States. Office of the Under Secretary of Defense (Personnel and Readiness). (2004). *Department of Defense Training Transformation Implementation Plan June 2004*. Washington D.C.: Office of the Under Secretary of Defense for Personnel and Readiness. <http://www.oft.osd.mil/library/library_files/document_214_Training%20Transformation%20Implementation%20Plan.pdf>

Watz, E. A., Schreiber, B. T., Keck, L. B., McCall, J. M., & Bennett, W. Jr. (2003). *Performance Measurement Challenges in Distributed Mission Operations*

Environments. [Electronic Version] Simulation Technology Newsletter. 03F-SIW-022.

Watz, E.A., Keck, L. B., & Schreiber, B.T. (2004). *Using PETS software to capture complex objective measurement data from Distributed Mission Operations (DMO) environments*. Paper presented at the 2004 Spring Simulation Interoperability Workshop. [Electronic Version] Paper available as 04S-SIW-143.doc.

AUTHOR BIOGRAPHIES

ANTOINETTE M. PORTREY is a Research Engineer with Lockheed Martin Systems Management at the Air Force Research Laboratory, Warfighter Readiness Research Division in Mesa AZ. She completed her M.S. in Human Factors from Arizona State University in 2005. She is the project lead for the automated PETS 1.X and PETS² systems for Distributed Mission Operations.

LOREN KECK is a Lead Software Engineer with Lockheed Martin Systems Management at the Air Force Research Laboratory, Warfighter Readiness Research Division in Mesa, Arizona. He received a B.S. in Computer Science from Arizona State University in 2001, and has several years of experience in the real-time data acquisition and analysis, voice communications, and human operational training fields. He is responsible for the program architecture and implementation of the automated (PETS²) system for Distributed Mission Operations.

BRIAN T. SCHREIBER is a Staff Scientist with Lockheed Martin Systems Management at the Air Force Research Laboratory, Warfighter Readiness Research Division, in Mesa, AZ. He completed his M.S. in Human Factors Engineering at the University of Illinois at Champaign-Urbana in 1995. He is the conceptual designer and advisor for both PETS 1.X and PETS².

DR WINSTON BENNETT, JR. is a Senior Research Psychologist with the Air Force Research Laboratory Human Effectiveness Directorate, Warfighter Readiness Research Division, located at Williams-Gateway Airport, Mesa AZ. He is the team leader for training systems technology and performance assessment research and development. He received his Ph.D. in Industrial Organizational Psychology from Texas A&M University in 1995.



Measuring Human Performance in Multi-National Distributed Events: Lessons Learned from the First Warfighter Alliance in a Virtual Environment

(Exercise FirstWAVE)



Antoinette M. Portrey
Loren B. Keck
Brian T. Schreiber
Winston Bennett, Jr.

Air Force Research Laboratory
Warfighter Readiness Research Division

This presentation is UNCLASSIFIED



Outline

- Why military DMO simulation environments exist (improved human performance).
- Illustrate growing desire/need for human performance assessment.
- Overview how we assess human performance within DMO environments.
- Example results.
- Challenges in expanding to large-scale, highly distributed exercises.



Why Do DMO Simulation Environments Exist?

- "Live" exercises such as Red Flag are costly, time-consuming, pose unrealistic constraints, and cannot afford nearly the training repetition that simulation can.
- Regarding military environments, to produce a more competent warfighter! Allowing Warfighters to participate in a continuous training cycle and maintain a high state of combat readiness.



Let's Not Forget What's Important

- Interoperability standards enable the distributed training for individuals, teams, and teams of teams.
- Visuals, databases, platform/missile models, modeling threat behaviors, etc. exist primarily to produce human proficiency that translates to combat success!
- If the warfighter does not benefit—that is, become more competent from these systems, then the systems will cease to exist.
- “It’s all about pixel count, the quality of the visuals...”
It is?!



So How Do We Assess Improvement?

- Measure kill ratios, bomb errors, time & degree of exposure to weapons envelopes, etc.
- Goal is to see human performance improvements in metrics like those above!



Outline

- Why military DMO simulation environments exist (improved human performance).
- Illustrate growing desire/need for human performance assessment.
- Overview how we assess human performance within DMO environments.
- Example results.
- Challenges in expanding to large-scale, highly distributed exercises.



Desire for Human Assessment in DMO: The Beginning

➤4/2000: Stated/funded desire for a LAN proof-of-concept.

➤Early '02: Numerous human performance assessment studies & data reqs generated: 350+ pilots for WS learning, approx. 100 pilots for decay/transfer, Bayesian modeling, data fusion, SDT research, event prediction research, etc. All purely designed to assess/understand human performance in a DMO environment.



Desire for Human Assessment: Momentum Building (cont.)

➤2003: Interest from operational pilots in seeing data for feedback purposes.

➤2003: Interest/funding from AFRL to expand capability to other LAN environments and WAN exercises.

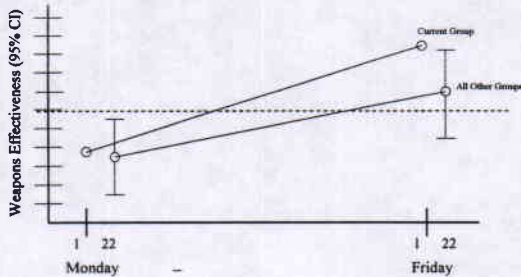
➤2003: Interest from several international communities in collaborating/leveraging such capability (e.g., UK, Netherlands, Canada, others...)

➤2004: Participation in the First Warfighter Alliance in a Virtual Environment

➤2005 and beyond:



Competency-Based Training: Assessing MECs



Benchmark Session



Outline

- > Why military DMO simulation environments exist (improved human performance).
- > Illustrate growing desire/need for human performance assessment.
- > Overview how we assess human performance within DMO environments.
- > Example results.
- > Challenges in expanding to large-scale, highly distributed exercises.



19 Teams, Mon-Fri Benchmarks

Experience

5 days, 35 scenarios against an avg of 293 threats, employing 483 shots

Outcome Metrics

63% fewer enemy strikers to target

68% fewer F-16 mortalities

24% more enemy fighter kills

7% increase in F-16 missiles resulting in a kill

62% reduction in threat missiles resulting in a kill

Ex. MEC Skill Metric

63% less time allowing hostiles into MAR (maps to ctrl's intercept geometry)



Outline

- > Why military DMO simulation environments exist (improved human performance).
- > Illustrate growing desire/need for human performance assessment.
- > Overview how we assess human performance within DMO environments.
- > Example results.
- > Challenges in expanding to large-scale, highly distributed exercises.

ESM
2005

Potential

Opportunities for measuring and evaluating successful mission employment exist at the individual, team, and inter-team levels. Performance outcomes measured are inversely proportionate to the number of simulation systems and the complexity of the teams within the exercise environment. A goal of the PETS² software is to develop and standardize a software tool to enable this multi-platform, multi-level measurement ability.

ESM
2005

First Warfighter Alliance in a Virtual Environment

➤ FirstWAVE was a multi-national technology demonstration that showcased the capability of implementing HLA in a large-scale, wide-area training event.

➤ Provided the ability to determine the serviceability of the simulation protocols and to evaluate the possibility of performance measurement on high-magnitude, internationally coordinated DMO training exercises.

ESM
2005

Primary Objective of FirstWAVE

➤ To establish, demonstrate, and document the potential of distributed simulation to enhance training for NATO combined air operations.

➤ The Mission Training through Distributed Simulation (MTDS) demonstration took the form of an air coalition training exercise participated by Canada, France, Germany, Italy, Netherlands, and the United Kingdom; with US Air Force personnel supporting the engineering development, connectivity and testing, scenario planning and management, and training effectiveness data collection.



Observations of FirstWAVE Issues

- FirstWAVE network traffic was recorded then used in the PETS2 architecture as raw data input to generate assessment metrics.
- Recorded log files contained instances of incomplete or inaccurate data.
- Many data recording errors were identified and attributed to issues such as long-haul network instability, simulations errors, etc...
- Challenges existed due to issues within and across DMO simulation environments.



Issue #1: IOS Operator

- No standard for Instructor Operator Station (IOS) personnel.
- Paradox—DMO community continually strives for improving realism (e.g., models, databases, flight dynamics), but scenario realism is often disregarded (e.g., shields, freezing fuel burn, deleting entities, etc.)
- If we want to train the way we intend to fight, we can't use shields! Similarly, if we want to assess performance and have it actually relate to the "real world", we must create a realistic scenario!



Issue #1 (cont.): IOS Proposed Rules...

- General theme: "If it can't be done in war, it can't be done from the IOS once a scenario starts."
- No shields (i.e., real-time kill removal)
 - No deleting weapons/entities
 - No mid-air weapons reloading
 - No freezing fuel burn
 - No mid-air refueling w/o visiting tanker
 - No relocating entities
 - New entities (including) regeneration from a specific point OK (i.e., pre-planned reserve forces called upon when reinforcements necessary)



Issue #2: Incomplete DIS/HLA Data

➤Entities in various simulation environments often successfully interoperate, but it does not mean that complete DIS/HLA information is sent across the network.

➤Examples: Missing shooter/target information for munitions, missing munition velocity, or using blank munition ID (0.0.0) for tracked munitions.



Issue #3: Inaccurate DIS/HLA Data

➤Entities in various simulation environments often successfully interoperate, but it does not mean that accurate DIS/HLA information is sent across the network.

➤Examples: Using "unspecified" munition ID when shooter was blue (or red), incrementing event ID for fire AND detonate on same munition (should use same event ID), or not sending entity state PDUs every 5 seconds.



Issue #4: Unrealistic DIS/HLA Data

➤Entities in various simulation environments often successfully interoperate, but it does not mean that realistic DIS/HLA information is sent across the network.

➤Examples: AIM-9 (heat-seeking missiles) with unrealistically long ranges, aircraft performing turns in excess of 30G, aircraft flying at unrealistically high Mach, extremely high pk rates on missiles, etc.



Issue #5: Customized Implementations

➤Entities in various simulation environments often successfully interoperate, but it does not mean that completely standard DIS/HLA information is sent across the network.

➤Examples: Proprietary voice data, unsupported tactical data links, or weapons load variations in data PDU.



Issue #6: Standards Currently Insufficient (or utilization not standard)

➤Entities in various simulation environments often successfully interoperate, but it does not mean that sufficient DIS/HLA information is sent across the network.

➤Examples: IOS operator, entity marking field in the entity state PDU, timeout values, more internal state data (weapons load, display modes & radar tracking lists, fuel status, selected switch settings, etc.)



Summary

To respond quickly to the dynamic challenges of today's environment, training, rehearsal, and exercise approaches and doctrine must remain flexible, operationally focused, mission effective and integrated with real-time, globally distributed mission rehearsal capabilities.

➤Issues exist within and across DMO environments that pose great challenges, but those challenges can be overcome.

➤Providing a dynamic, capabilities-based training for the Wafighter must be a joint effort between the simulation, operational and training communities if it is to succeed in today's environment.



Acknowledgements

Special thanks to our direct funding agency:

Air Force Research Laboratory, Warfighter Training
Research Readiness Division in Mesa, AZ (Dr. Winston
Bennett, PI)

And to our supporters/advocates at Air Combat Command
(Chuck Colegrove and Bob Chapman)



Questions?

