Automated Human Performance Measurement: Data Availability and Standards

Emily Wiese[1], Beth F. Atkinson[2], Matthew Roberts[1], Jeanine Ayers[1], Delonna M. Ramoutar[3]
[1]Aptima, Inc., Woburn, MA; ewiese, jayers, mroberts@aptima.com
[2]Naval Air Warfare Center Training Systems Division, Orlando, FL; beth.atkinson@navy.mil
[3]Kaegan Corporation, Orlando, FL, delonna.ramoutar.ctr@navy.mil

ABSTRACT

As Human Performance Measurement (HPM) continues to advance, developers are facing unique challenges in emerging environments due to a lack of availability to the right data—information specific to measuring and assessing human performance. Additionally, while simulation-based training utilizes a number of standards and protocols, there is limited guidance for making HPM data available and consistent across devices. This paper discusses on-going efforts that demonstrate the impacts of data availability on trainee assessment capabilities, thereby presenting a data driven case for the importance of HPM data. Additionally, the authors present lessons learned for increasing HPM data availability, including defining performance measures early in the training system development lifecycle and establishing HPM standards.

ABOUT THE AUTHORS

Emily Wiese is a Principal Human Factors Engineer at Aptima and Associate Division Director for the Human Performance Division. She leads Aptima's efforts developing performance measures for training in various domains (including various fighter jet platforms, command and control platforms, ISR platforms, and ground operations) and is the product manager for Aptima's suite of performance measurement technologies. She has an M.S. in Industrial Engineering from the University of Iowa.

Beth F. Atkinson is a Research Psychologist at NAWCTSD. She manages several research and development efforts devoted to investigating capability enhancements for training and operational environments. Her research efforts include automated performance measurement technologies, instructional tools and simulation environments for hypoxia and spatial disorientation training, performance trend assessment and post-mission reporting, and Human Computer Interaction design and evaluations. She holds an M.A. in General Psychology from the University of West Florida.

Matthew Roberts is a Principal Software Engineer and Team Lead for the Performance Assessment Systems group. He leads teams of software engineers developing performance measurement and assessment systems and computer based training systems. Mr. Roberts has extensive experience architecting and developing automated system based measurement solutions and leads the integration of Aptima's measurement technologies (e.g., PM Engine™) with DoD distributed simulation environments. Mr. Roberts holds a B.S. in Computer and Systems Engineering from Rensselaer Polytechnic Institute.

Jeanine Ayers is the Associate Division Director for Aptima's Software Engineering division. She leads teams of software engineers in development efforts focused on architecting and building data-driven software applications in support of multi-agent systems, computer based training systems and performance measurement and assessment systems. Ms. Ayers has extensive experience with High Level Architecture (HLA) and Distributed Interactive Simulation (DIS) simulators in DoD environments, including work for the Army, Navy, and Air Force. She serves on the TRAIN SISO committee which promotes the use of simulations to support the acquisition of knowledge and skills through education, training, and performance support. Ms. Ayers has an M.B.A from Boston University and a B.S. from Carnegie Mellon University in Industrial Management/Information Systems.

Delonna M. Ramoutar is a Psychology Research Assistant with Kaegan Corporation providing research and development support at NAWCTSD. She has a B.S. in Psychology from the University of Central Florida and is looking to pursue a M.S. in Industrial/Organizational Psychology. Her research interests include human cognition, usability, personnel selection, and performance measurement.

Automated Human Performance Measurement: Data Availability and Standards

Emily Wiese[1], Beth F. Atkinson[2], Matthew Roberts[1], Jeanine Ayers[1], Delonna M. Ramoutar[3]
[1]Aptima, Inc., Woburn, MA; ewiese, jayers, mroberts@aptima.com
[2]Naval Air Warfare Center Training Systems Division, Orlando, FL; beth.atkinson@navy.mil
[3]Kaegan Corporation, Orlando, FL, delonna.ramoutar.ctr@navy.mil

INTRODUCTION

Human Performance Measurement (HPM) is used to assess the structured use of simulation-based training and to provide feedback to ensure an effective and efficient training event (Stacy, Freeman, Lackey, & Merket, 2004). HPM has made great strides over the last 25 years, yet the next generation of Automated Performance Measures (APMs) and assessment capabilities are facing unique challenges. Through the adoption of simulation data protocol standards such as High Level Architecture (HLA) and Distributed Interactive Simulation (DIS), simulation environments developed in the past decade have largely overcome challenges in publishing large volumes of raw data (Stacy, Ayers, Freeman, & Haimson, 2006). In fact, an increased focus on distributed simulation (Dwyer, Oser, Salas, & Fowlkes, 1999; Portrey, Keck, & Schreiber, 2006) and Live, Virtual and Constructive (LVC) training environments has resulted in an inverse problem. While massive volumes of data can now be published over standard communication protocols, these emerging environments present developers with a shortfall in publishing the right data for HPM (Portrey et al., 2006). This results in a lack of information specific to measuring and assessing human performance.

Additionally, while simulation-based training utilizes a number of standards and protocols such as HLA and DIS, there has been limited attention to defining standards for making data specific to HPM available. Even with the definition of HPM data standards, simulation environments must consistently publish this data in order to support consistent APM calculations across devices and platforms (Portrey et al., 2006).

This paper will discuss on-going efforts that demonstrate the impacts of data availability on trainee assessment capabilities, thereby presenting a data driven case for the importance of HPM data in simulation-based training. Further, in support of future efforts, the authors will describe lessons learned for increasing HPM data availability, including defining performance measures early in the training system development lifecycle and establishing HPM standards (i.e., HLA Base Object Model).

HUMAN PERFORMANCE MEASUREMENT

In assessing trainee individual and collective performance, it is important to first identify those aspects of performance that are most critical to determining warfighter readiness. In today's complex platforms, readiness is not simply about dropping bombs or shooting the right target. Today's warfighters must employ a sophisticated set of tools and interact with other warfighters to conduct missions. In the Navy's P-8A Poseidon (a new maritime patrol and reconnaissance aircraft), for example, the entire crew works together to conduct long-range anti-submarine warfare; anti-surface warfare; and intelligence, surveillance, and reconnaissance missions (Boeing, 2012). To be successful, members of the crew must be skilled at both their individual task work and at the coordination and collaboration skills necessary to achieve the mission goals. For instance, the collective employment of different sensors by each crewmember may result in a more robust method of identifying an entity than each crewmember using a singular method in isolation.

To assess trainee performance, HPM data must be collected. Typical methods for data collection include self-report (i.e., directly from trainees), observational (by instructor-observers), or through automated means (i.e., using simulator data). Developing self-report or observational performance measures is relatively straightforward, yet finding the instructor-observers to administer or rate surveys can present manning constraints (MacMillan, Entin, Morley, & Bennett, In press; Wiese, Nungesser, Marceau, Puglisi, & Frost, 2007). In addition, care must be taken to ensure that observational and self-report measures are collected consistently across instructor-observers.

Automatically collecting performance measurement data directly from the simulator is often preferred, for a number of reasons. First, APMs provide a consistent, reliable way of collecting the same performance data across trainees, training sessions, and training scenarios. Second, this data can be stored in a database for immediate use or subsequent analysis. Finally, APMs can significantly reduce workload, allowing instructors to spend their time managing the simulated training scenario and instructing, rather than collecting performance data.

To truly achieve these benefits, however, the training simulator must publish the HPM data necessary to develop instructive, detailed APMs. Specifically, HPM data captures the trainee interactions with key mission systems, the simulated world, and other players (real or constructive), as well as the resulting performance measurements. Naturally, the precise nature of these data vary somewhat by platform and should be identified during a Front End Analysis (FEA) in which specific training objectives and associated performance measures are defined (Duke & Shook, 2010). Utilizing the previous example, assessing a P-8A crew's ability to effectively employ sensors requires HPM data including information about the sensors employed (e.g., which sensors and modes, who employed them and when) and the entities being investigated. While some of this type of data is published from many simulators, it is not universally the case.

SIMULATION STANDARDS & PROTOCOLS

Current simulation communication protocols exist and are being actively used and developed, with DIS and HLA being two of the most common standards used by the Department of Defense. The Naval Aviation community, in particular, has been supporting the adoption and use of the Naval Aviation Simulator Master Plan (NASMP) version of HLA. The NASMP Federation is an actively managed HLA federation, with additions and changes to the standard occurring on a regular basis. As new simulator platforms or upgrades are developed, changes to different versions of NASMP are introduced to the NASMP Federation Working Group, discussed, and (generally) accepted. These changes typically involve the addition of objects and interactions required to enable the effective portrayal of new platform capabilities to other entities on the simulation network.

In recent years, the engineering community has identified the need to supplement existing standards with protocols to specifically address instructional capabilities. For example, a Simulation Interoperability Standards Organization (SISO) study group published a final report for distributed debrief control (SISO DDCP Study Group, 2011). While this protocol addresses some capabilities applicable to observer-based HPM (i.e., bookmarking, annotation tools), the goal of this protocol is to support the *control* of distributed debriefing leaving the specifics of sharing HPM data unaddressed.

HPM Data Requirements

Data currently communicated via HLA and DIS standards includes information about the entity state, such as location, altitude, speed, weapons events, damage assessment. Information specific to platform subsystems (e.g., sensors used) is typically not published at a level of detail that is useful for comprehensive assessment. When that type of information is published, it is generally very basic (i.e., a sensor is on or off). For example, recent efforts to define APMs for the P-8A Antisubmarine Warfare mission (ASW) resulted in full implementation of only 27% of measures. Achieving even this level of implementation required some changes to the data published on the NASMP Federation. Because of the unique mission of the P-8A, many of the measures that are deemed of high importance by the P-8A FIT are substantially different from those seen in other platforms (e.g., tactical performance of fighter jets). Assessing the Techniques, Tactics, and Procedures (TTPs) associated with the employment of the P-8A sensors is, in this situation, much more relevant than assessing flight performance. Even after making some modifications to NASMP, there are still major deficiencies in data availability across a number of areas, to include:

- Datalink messages (Link 16/11)
- Radio communications
- Trainee's systems inputs that reflect their perceived ground truth (e.g., track placement)
- Platform subsystem information and trainee inputs (e.g., sensor modes used, track designations)
- Expected crew standards, such as EMCOM, that vary by mission
- Performance characteristics about Semi-Automated Forces (SAFs)

As more platforms with intelligence and surveillance focused missions adopt simulation-based training as their primary means of training, APMs will also increasingly become a requirement. Thus, the need for this type of data will only increase in the coming years.

Existing simulation-based training devices are more likely to face this issue today. For example, only 34% of the measures identified to support APM during Strike Weapons training events at the Fleet

Replacement Squadron (FRS) at NAS Lemoore could be implemented due to data availability issues (Seibert, Amodeo, Jungemann, Keeney, Estock, Entin, & Thoreson, 2012). Even fewer APMs (2% of those identified) were implemented in a recent study conducted with the MRT3, an MH-60S desktop simulator, also due to data availability issues (Wiese, Dean, Schnell, & Anderson, 2012).

Requirements evaluations for meeting user expectations for debriefing a simulated training event reveal the need to play back recorded data (e.g., simulation logs, audio, and video as described in Distributed Synchronized Playback Protocol and Implementation; McDermott et al., 2010). While this type of playback is beneficial to the instructor and trainee, it does not capture all necessary HPM data to allow for an effective and efficient training exercise. Furthermore, in order to conduct effective distributed debriefing sessions for distributed simulation-based training events, HPM data must be captured and communicated back to all participants (Wiese, Freeman, Salter, Stelzer, & Jackson, 2009).

Reasons for Current State of the Practice

Much of the data described above hasn't been published for a variety of logical reasons. First, it's not required in order for other SAFs or simulators to operate realistically in a simulation event nor has there been a formal requirement for simulators to publish HPM data. Historically, much of the focus in developing and/or integrating simulators has been in ensuring that the simulator behaves in a realistic manner during scenario execution.

Second, in the case of the communications data, although some DIS environments capture some amount of radio communications, it can still be bandwidth heavy. Additionally, methods for automatically analyzing text or voice communications data are still nascent and not providing an external push to obtain this data. Finally, for data such as SAF performance characteristics, it may be sufficient for each SAF modeling environment to provide references to this information rather than publish it over the network.

BENEFITS AND IMPACTS OF HPM DATA AVAILABILITY

Availability of the correct HPM data can impact both trainee learning and instructor activities during a training event. Research has shown that providing the right feedback to trainees at the right time increases learning by 25%-50% (Thalheimer, 2002). Furthermore, current evidence suggests an optimal window of time to provide feedback to trainees ranges from 20 minutes to one day (Thalheimer, 2008). In simulation-based training, this feedback is traditionally provided in the context of a debrief that occurs after the training event. Current simulation-based training debriefs consist largely of playing back recorded data, rather than providing specific feedback to trainees on areas of poor performance. While playback is a valuable component of providing feedback, it rarely actively supports the instructor in diagnosing trainee performance, understanding trainee performance in the context of the simulated training scenario, and summarizing that performance in a way that supports retention (Wiese et al., 2009). Explicitly including human performance measurement data in a debrief can result in improved retention of key concepts, thus making subsequent training more effective and efficient. Automating the process of calculating performance measures can ensure that the feedback occurs in a timely manner.

HPM data availability, along with resulting APMs, can also decrease instructor workload associated with monitoring performance in real-time and evaluating performance during debriefing. Some preliminary feedback on how using automated performance measurement technology reduces workload has been collected during standard testing events. Discussions regarding workload levels after APM implementation suggest a 25-50% reduction in workload associated with monitoring trainee performance during a simulated training scenario and preparing for the subsequent debrief (P-8 FIT Subject Matter Experts, personal communication, July 29, 2011). While this provides a preliminary understanding of the baseline data benefits associated with implementing APMs, there are limited sources on workload levels to fully understand the implications on return on investment.

HUMAN PERFORMANCE MEASUREMENT BASE OBJECT MODEL (HPM BOM)

As described above, the simulation world has not yet developed standard mechanisms to collect APMs of trainee behavior. Efforts focused on defining HPM data requirements resulted in the development of the HPM Base Object Model (BOM) for use with the Navy Aviation Simulation Master Plan (NASMP) Federation Object Model (FOM), which defines standards to support the specification and sharing of measures of human performance. The intent of the HPM BOM is to provide a mechanism for communicating performance measurements and assessments to interested federates. Making trainee performance data available to others while a training exercise is ongoing can positively impact simulation-based training in a variety of ways.

First, that data can be consumed by real-time displays of trainee performance, reducing instructor workload and enabling more timely instruction and accurate feedback to trainees. Second, trainee performance data can be used to dynamically adapt training scenarios to support the evolving instructional needs of trainees (e.g., Dean, Stacy, Keeney, Day, Terry, & Alicia, 2011; Stacy, Picciano, Sullivan, & Sidman, 2010). Finally, trainee performance data can be used to identify trends in performance across time, and to adjust the training curriculum and/or instruction methodologies to address performance gaps.

Importantly, the HPM BOM was not intended to fully represent all data required to calculate measures of trainee performance. Rather, its intent is to provide a mechanism for directly representing concepts associated with trainee learning in the simulated world. The HPM BOM provides the class structure of the HLA objects and interactions needed to define information required by simulation federates to describe the performance of human trainees against specific training objectives while performing a given mission. The HPM BOM Class 1 level object classes include the *HPMPerformanceMeasurementData* and *HPMTeamData* and one new Class 2 level object class, *HPMDetailData*. The *HPMPerformanceMeasurementData* object class is designed to provide a base set of data describing a performance measurement object that relates to a specific trainee or a team of trainees within the simulation environment. Attributes include:

- *Author*
- *Description*
- *DisplayValue*
- *MeasureName*
- *TeamObjectIdentifier*

The *HPMTeamData* object class is designed to provide an object to describe and group a number of human trainees in the simulation environment as a team and to associate performance measures to the team. Attributes include:

- *Description*
- *TeamName*
- *TraineeList*

The *HPMDetailData* object class is designed to provide additional attributes for use in describing a performance measurement object that relates to a specific trainee within the simulation environment. Attributes include:

- *Assessment*
- *AssessmentContext*
- *Location*
- *TrainingObjective*

Existing Use Cases

The HPM BOM was formally accepted by the NASMP Federation Working Group (FWG) in 2010. The first operational simulator to use the HPM BOM to share human performance data will be the P-8A Tactical Operations Flight Trainer (P-8A TOFT). The P-8A TOFT will be used to train crews for ASW, Anti-Surface Warfare (ASuW), and Intelligence, Surveillance, Reconnaissance and Targeting (ISR&T) missions. In the requirements for this trainer, the Navy called for automated performance measurement. This requirement was met by leveraging the combination of Performance Measurement Engine (PM Engine) and the HPM BOM.

In order to enable a simulator to automatically capture performance measures, effort must first be spent identifying reliable and operationally valid measures of trainee performance. In the case of the P-8A, analysis of the three mission spaces (i.e., ASW, ASuW, ISR&T) resulted in definition of over 150 performance measures. The set of measures defined spans all phases of the three missions and each measure assesses the crew as a whole.

Many considerations need to be made when integrating automated performance measures with a simulator. Some are fairly obvious such as: *What mechanisms will be used to get the data? What data will be available?* and *How should the results be presented?* Others are not so obvious such as: *How will the measurement system be configured?* and *How will the instructors and students learn to trust the results being produced*?

Because the P-8A adheres to the NASMP HLA standards, the answers to some of these questions were relatively straight-forward. The HLA standard and the Navy requirement to function with a specific Run Time Environment (RTE) defined the mechanism to subscribe and receive the data, while the NASMP FOM and Guidance Rationale and Interoperability Manuals (GRIM) clearly defined the data that would be readily and immediately available. For data that was required for trainee assessment but not already planned to be available, conversations with PMA-290 and the simulator manufacturer helped determine what and when changes to the NASMP FOM could be made to support comprehensive APM.

The questions regarding how to configure the measures and display the results took more effort to answer and resulted in the development of two plug-in applications which reside within the Instructor Operator Station (IOS) of the P-8A simulator. The first plug-in allows for the configuration of the performance measures, the association of configured measures with scenarios, and the communication of configured measures to the PM Engine. The second plug-in displays the performance measurement results on a timeline at run-time and during debrief.

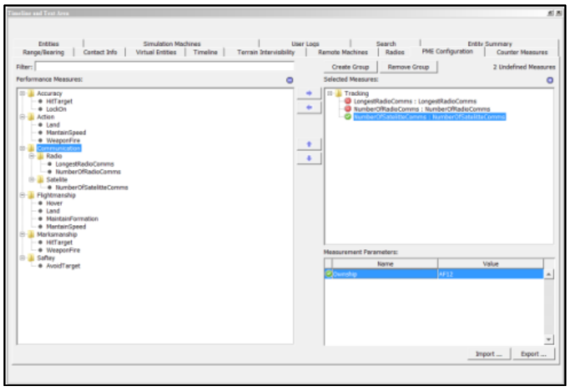

Figure 1. Measurement Configuration Plug-in

The *Measurement Configuration Plug-in*, as seen in Figure 1, enables instructors to select and configure measures at both scenario authoring time and at run-time. When configuring measures during scenario authoring, the resulting measurement configuration is associated with and saved along-side the scenario file. When a scenario is loaded, the plug-in sends the configuration information to the PM Engine over HLA via an additional BOM (i.e., PM Engine Configuration (PMEC) BOM) created during the development of the P-8A TOFT. It has been ratified by the NASMP Federation Working Group and can be used to communicate this information within any trainer.

The body of the PM Engine configuration messages contain Human Performance Markup Language (HPML), which details the measurement definitions and how they've been configured for the current scenario (Stacy et al., 2004; Stacy, Merket, Freeman, Wiese, & Jackson, 2005). The PM Engine interprets the HPML contained within these messages and instantiates measures which subscribe to and subsequently processes various pieces of simulator data. Based on the rules defined in the HPML, the measures will produce measurement and assessment results. These results are then published over the HLA using the HPM BOM. More specifically, the PM Engine creates and updates both *HPMPerformanceMeasurementData* and *HPMDetailData* objects. As described above, these objects represent performance measurement and assessment results and their relationships to each other.

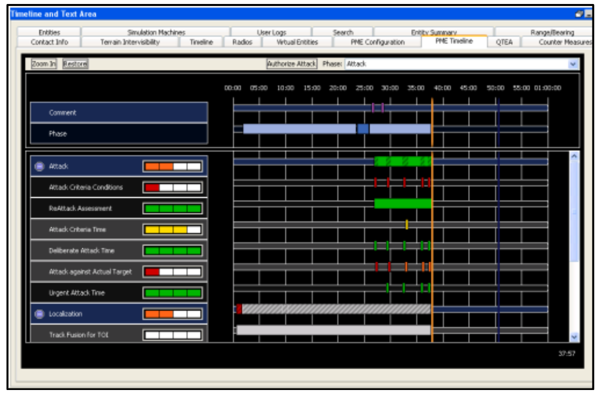

Figure 2. Measurement Result Timeline Plug-in

On the other end of the performance measurement process, the *Measurement Result Timeline Plug-in* (Figure 2) receives all of the performance measurement objects in real time over HLA via the HPM BOM. The *Measurement Result Timeline* plug-in maintains a history of the objects and displays them in an organized manner on a stacked timeline. The calculated assessment categories are represented by a color scale ranging from purple for extremely poor to green for expert performance. Hovering over a measurement in the timeline results in the display of a tool tip. The tool tip contains the details of the underlying measures and data. These details help instructors and trainees understand why a certain assessment was given. This is important for two reasons. First, it provides rich objective feedback for trainees. Second, it facilitates increased trust of the APMs by offering a view into the measurement system and the means by which they were calculated.

By employing these standards to communicate performance measurement definitions and results within the P8-A simulation environment, we have left the door open for other federates to both produce and/or consume this data in the future. The components used to configure, calculate, and display performance measurement data can be easily migrated to other NASMP compliant environments. Finally, this concept can be readily integrated into other HLA Federations or simulator standards.

CONCLUSIONS

In order to make simulators effective trainers, simulators must make human performance measurement data more available and support the transfer of performance measurement instructions and results. As both live and simulated training technologies continue to advance, the feasibility of collecting automated performance measurement and assessment data is increasing. This paper reviewed the progress and challenges associated with integrating APM technologies into the P-8A trainer. Similar efforts are underway for other Navy, Marine Corp and Air Force platforms, including:

- Additional Navy platforms investigating integration of APM capabilities include the MH-60R and E-2D. An analysis is underway to identify available data within the MH-60R simulator and aircraft data recordings that support HPM calculations. For the E-2D Hawkeye Integrated Training System – Aircrew, initial performance measurement and debrief technologies are being fielded on early increments of the devices, with plans to integrate more comprehensive APM and debriefing capabilities.
- The Marine Corps Combat Hunter program provides an example of leveraging HPM technologies to increase observation and perceptual skills within trainees (DeVore, 2007).
- Air Force investments in the Performance Evaluation and Tracking System (PETS) has resulted in technologies being integrated within F-15, F-16, F-22, Joint Terminal Air Controller (JTAC) and Predator platforms, as well as within Air Force Live, Virtual and Constructive events (Neubauer, 2009; Neubauer & Watz, 2011).

Additionally, an increased reliance on distributed and LVC environments due to fiscal constraints will continue to increase the need for effective and efficient training through inclusion of HPM capabilities. However, unless the proper standards and protocols are put in place now, the current challenges associated with the lack of availability of HPM data will continue.

The benefits of making HPM data available in these environments include increased fleet readiness and reduction of instructor workload. By leveraging HPM data, APMs provide instructors with the information necessary to provide diagnostic feedback in debriefing and post-mission readiness reports. Research suggests a 10-20% improvement in training effectiveness and enhancement of fleet readiness due to providing diagnostic feedback (Astwood, Van Buskirk, Cornejo, & Dalton, 2008; Azevedo, & Bernard, 1995). Further, the results of APMs aid in the detection of performance deficiencies and support providing targeted remediation to students/crews. Using APMs to provide trends in performance supports the identification of critical training shortfalls that require modification of training curriculum or scenarios (e.g., students not becoming proficient at a skill based on existing training, students 'gaming' the system through a priori knowledge of training scenario events). Additionally, APM results can support streamlining the training cycle by providing a performance driven training readiness curriculum. For example, current training syllabi are often based on a specific number of training iterations. Use of APM data would allow trainees who have achieved operational readiness in one domain or skill set (e.g., tactical skills) to progress rather than continue to focus on the same training.

In addition to enhancing training, making data available for APMs to calculate measures of performance reduces instructor workload. First, by assessing trainee performance for tasks that can be understood by automated systems, instructors' attention can be better directed at other required tasking. Second, making APM results immediately available post-mission facilitates a shorter turnaround to the debrief session, thereby allowing instructors more time to focus on issues or to increase student throughput. Finally, whether done in live or simulated platforms, qualification training typically requires instructors or support staff to document mission summaries and to complete post-mission readiness reporting. Automatically collecting and sharing HPM data can support automated reporting capabilities and reduce instructors' time to complete post-mission summaries.

The HPM BOM described in this paper provides the first step towards addressing current shortfalls of access to the right data. A standard mechanism for sharing performance data such as this can increase the consistency across systems in the performance being measured, thereby increasing the standardization of feedback provided to trainees.

ACKNOWLEDGEMENTS

The views expressed herein are those of the authors and do not necessarily reflect the official position of the organizations with which they are affiliated. The authors would like to thank members of the P-8A FIT for their invaluable input in the efforts described in this paper. We would also like to thank Rob Abbott and colleagues at Sandia National Labs for their inputs to the HPM BOM.

REFERENCES

Astwood, R. S., Van Buskirk, W. L., Cornejo, J. M., & Dalton, J. (2008). The impact of different feedback types on decision-making in simulation based training environments. *Proceedings of the Human*

Factors and Ergonomics Society 52nd Annual Meeting 52(26), 2062-2066.

Azevedo R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research 13*(2), 111-127.

Boeing International. (2012). *P-8A Poseidon*. Available: http://www.boeing.com/defense-space/military/p8a/index.html

Dean, C., Stacey, W., Keeney, M., Day, E., Terry, R., & Alicia, T. (2011). Item response theory adapts training to disparately skilled trainees. *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference*, Orlando, FL.

DeVore, L. (2011). Release of observation and perceptual skills training system. *Military Training Technology, 16* (6). Retrieved from http://www.kmimediagroup.com/files/ MT2_16-6_final(2).pdf.

Duke, D., & Shook, W. (2010). Using Department of Defense Architectural Framework (DODAF) to identify initial training system requirements. *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference*, Orlando, FL.

Dwyer, D. J., Oser, R. L., Salas, E., & Fowlkes, J. E. (1999). Performance measurement in distributed environments: Initial results and implications for training. *Military Psychology, 11*(2), 189-215.

MacMillan, J., Entin, E. B., Morley, R., & Bennett, W. (In press). Measuring team performance in complex and dynamic military environments: The SPOTLITE method. Submitted to *Military Pyschology*.

Neubauer, P. (2009). Improving simulation data quality for automated human performance assessment. *Proceedings from Fall Simulation Interoperability Workshop 2009.* Red Hook, NY: Curran Associates.

Neubauer, P. & Watz, E. (2011). Network protocol extensions for automated human performance assessment in distributed training simulation. *Proceedings from Spring Simulation Interoperability Workshop 2011.* Red Hook, NY: Curran Associates.

Portrey, A. M., Keck, L. B., & Schreiber, B. T. (2006). *Challenges in developing a performance measurement system for the global virtual environment* (Report No. AFRL-HE-AZ-TR-2006-0022). Air Force Research Laboratory, Warfighter Readiness Research Division, Mesa, AZ: Department of Defense.

Seibert, M. K., Amodeo, A. M., Jungemann, J., Keeney, M. J., Estock, J. L., Entin, E. E., & Thoreson, J. A. (2012). Fidelity in Learning Environments for the Effectiveness of Training SBIR , Phase II (FLEET II). Patuxent River, MD: Naval Air Warfare Center, Training Systems Division.

Simulation Interoperability Standards Organization (SISO) Distributed Debrief Control Protocol (DDCP) Study Group. (2011, April). *Final report for: The distributed debrief control protocol study group* (technical report SISO-REF-028-2011). SISO.

Stacy, W., Ayers, J., Freeman, J., & Haimson, C. (2006). Representing human performance with human performance measurement language. *Proceedings of the Fall Simulation Interoperability Workshop*, Orlando, FL.

Stacy, W., Freeman, J., Lackey, S., & Merket, D. (2004). Enhancing simulation-based training with performance measurement objects. *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference*, Orlando, FL.

Stacy, W., Merket, D., Freeman, J., Wiese, E., & Jackson, C. (2005). A language for rapidly creating performance measures in simulators. *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference*, Orlando, FL.

Stacy, W., Picciano, P., Sullivan, K., & Sidman, J. (2010). From flight logs to scenarios: Flying simulated mishaps. *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference*, Orlando, FL.

Thalheimer, W. (2002). *Percentage improvements that are likely when key learning factors are utilized*. Retrieved from Work-Learning Research, Inc. website: www.work-learning.com

Thalheimer, W. (2008, May). *Providing Learners with Feedback—Part 2: Peer-reviewed research compiled for training, education, and e-learning*. Retrieved November 31, 2008, from http://www.work-learning.com/catalog/

Wiese, E. E., Nungesser, R., Marceau, R., Puglisi, M., & Frost, B. (2007). Assessing trainee performance in field and simulation-based training: Development and pilot study results. *Proceedings of the Interservice/Industry Training, Simulation & Education Conference* , Orlando, FL.

Wiese, E. E., Freeman, J., Salter, W. J., Stelzer, E. M., & Jackson, C. (2009). Distributed After Action Review for Simulation-based Training. In D. A. Vincenzi, J. A. Wise, M. Mouloua, & P. A. Hancock (Eds.), *Human Factors in Simulation and Training* (pp. 287-299). Boca Raton, FL: Taylor & Francis Group.

Wiese, E., Dean, C., Schnell, T., & Anderson, N. (2012). Knowledge Optimized Displays of Information in Human Computer Interaction II (ORCID). Orlando, FL: Naval Air Warfare Center Training Systems Division.