Automated Human Performance Measurement: Standardizing Lifelong Learning Training Data

Mitchell Tindall, Beth F. Wheeler Atkinson
Naval Air Warfare Center Training Systems Division
Orlando, FL
mitchell.j.tindall.civ@us.navy.mil
beth.f.atkinson.civ@us.navy.mil

LT Sarah Beadle
Naval Air Warfare Center Aircraft Division
Patuxent River, MD
sarah.c.beadle3.mil@us.navy.mil

ABSTRACT

Data-driven decision-making and big data have become ubiquitous in the Department of Defense (DoD) as there is widespread acknowledgement of the potential for both to advance warfighter performance (DoD Data Strategy, 2020). Advances in artificial intelligence (AI), machine learning (ML), and cloud-based data storage and processing have created opportunities to conduct previously impossible analyses. Advances in wearable and non-invasive physiological sensors, along with eye tracking technology, can be utilized with AI/ML to objectively capture important human performance measurement (HPM) that currently require human observation. While these capabilities hold significant promise for advancing DoD data analytic tools, substantial groundwork must be performed to ensure the reliability and validity of the underlying data. Fortunately, the DoD has made significant strides in this space by investing in data standards (e.g., Experience Application Programming Interface, Human Performance Modeling Language) that delineate the types and format of system-based data required to better understand warfighter learning and performance. This paper is intended to revisit the progress made on learning and HPM standards as an essential capability for data strategy and reframe their function as part of the larger DoD wide data strategy guiding principles. The ultimate goal is to ensure learning and HPM data standards include language to capture and utilize reliable, valid, and transparent data from the beginning of training and throughout the performance of their duties. Now is also an opportune time to accommodate emerging technologies and the unique potential they offer to close data gaps, and meaningful visualizations for a variety of stakeholders. By applying *standards* and an *architecture* as a framework to a beginning stage Naval Aviator training and advanced Naval Flight Officer use cases, the authors will conceptualize extensions for advanced learners and provide initial recommendations for changes to the standards that account for emerging technologies.

ABOUT THE AUTHORS

Mitch J. Tindall, PhD, is a Senior Research Psychologist at Naval Air Warfare Center Training Systems Division (NAWCTSD) in the Basic & Applied Training & Technologies for Learning & Evaluation (BATTLE) Laboratory. During his graduate work, he performed research and consulting in human performance assessment and productivity enhancement. His work for the Navy includes several areas such as psychometric development and validation, human performance measurement, human computer interaction, data management and analytics, and training systems enhancement and validation. His research interests include human performance measurement, physiological episode training enhancement, and data science. His Ph.D. is in Industrial-Organization (I-O) Psychology from University of Central Florida (UCF).

Ms. Beth F. Wheeler Atkinson, is a Senior Research Psychologist at NAWCTSD and a Naval Air Warfare Center Aircraft Division (NAWCAD) Fellow. She has led several research and development efforts devoted to investigating capability enhancements for training and operational environments and has successfully transitioned a post mission reporting and trend analysis tool that leverages automated performance measurement technology. As the lead of the BATTLE Lab, she supports a portfolio focused on naval aviation training for survival, pilots, aircrews, and maintainers through advancing human performance analytics, automation and decision support technology, instructional tools and methods, synthetic environments enhancements, and extended reality. She holds an M.A. in Psychology, Applied Experimental Concentration, from the University of West Florida.

LT Sarah C. Beadle, PhD, is an Aerospace Experimental Psychologist and Project Officer at NAWCAD in the Human Systems Engineering Department. LT Beadle has a PhD and M.S. in Human Factors Psychology from Clemson University. Her prior work is primarily in simulator sickness, extended reality for training, and human systems integration with a variety of Naval Aviation platforms. LT Beadle is a 2019 I/ITSEC Scholarship alumna.

Automated Human Performance Measurement: Standardizing Lifelong Learning Training Data

Mitchell Tindall, Beth F. Wheeler Atkinson
Naval Air Warfare Center Training Systems Division
Orlando, FL
mitchell.j.tindall.civ@us.navy.mil
beth.f.atkinson.civ@us.navy.mil

LT Sarah Beadle
Naval Air Warfare Center Aircraft Division
Patuxent River, MD
sarah.c.beadle3.mil@us.navy.mil

INTRODUCTION

The 2020 Department of Defense (DoD) Data Strategy highlights a motivation to become a data centric organization that leverages enterprise-level management to ensure valid and critical data is widely available to all relevant stakeholders. Furthermore, that data should be usable, delivered in real-time, secure, and connected to other relevant information sources (DoD, 2020). Lastly, the DoD recognizes that software and hardware systems must be designed, procured, tested, upgraded, and sustained with data interoperability as a key requirement. To realize the noted goal of improving proficiency, it is critical to emphasize the criticality of interoperability to facilitate use of data within and across systems for understanding human performance. While the DoD Data Strategy identified four *Essential Capabilities* necessary for achieving the above aim, a review of two of these–*Standards* and *Architecture*–will provide a framework to successfully implement human performance measurement (HPM) throughout the lifetime of naval learners. As such, the overall goal of this paper is to propose for expansion of *standards* that includes HPM data derived from advancing technologies (e.g., physiological and eye-tracking sensors, natural language processing) and leverage an *architecture* that supports defining the application of these solutions within modular systems enabled by the emerging capabilities of artificial intelligence and machine learning.

Learning and Human Performance Standards

Standards are defined as "published documents that establish technical specifications and procedures designed to maximize the reliability of the materials, products, methods and/or services people use everyday" (IEEE SA, 2021). One of the most important features of standards is that they establish protocols that fuel compatibility and interoperability (IEEE SA, 2021). There are two standards related to or associated with HPM in use by government and industry: Experience Application Programming Interface (xAPI) and Human Performance Modeling Language (HPML). For transparency, HPML remains a draft standard as it was never formalized by the Simulation Interoperability Standards Organization (SISO). Broadly, these standards help facilitate the tracking of learning and performance over time and across modes of education (Nouira, Cheniti-Belcadhi & Braham, 2018). The original development of these standards resulted from specific challenges tracking learning and performance in various environments (e.g., eLearning, classrooms, simulation, on-the-job). To start this review, we will provide examples of available learning and HPM standards and their application to support sound systems engineering that facilitate interoperability and data capture in military contexts that advance human performance and training systems.

Learning Standard: Experience Application Programming Interface (xAPI)

xAPI is an Institute of Electrical and Electronics Engineers (IEEE) approved standard (IEEE 9274.1.1-2023) that expanded on SCORM's eLearning focus by enabling stakeholders to collect and exchange information about learner's experiences in any environment (Torrance & Jackson, 2020). xAPI is the broadest and most widely adopted of the standards that uses Learning Record Stores (LRSs) to track experiences associated with learning (Torrance & Wiggins, 2016). The xAPI standard relies on statements that include information on who the learner is, what the learning did, and what the event focus was (see Figure 1) to track learning experiences (Poeppelman, Long, Amburn, Hruska & Bink, 2013). The learning experiences can range from completing an action, like reading material, to how the learner performed on a multiple-choice test.

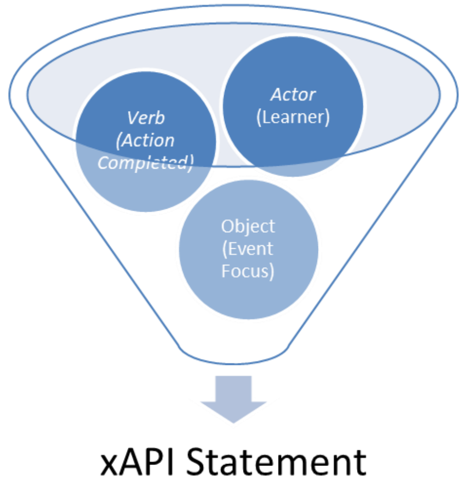

Figure 1. The components that are part of an xAPI statement as outlined in IEEE 9274.1.1-2023

Currently, xAPI is primarily supported for learning technologies such as videos, eLearning, Learning Experience Platforms, and LRSs (Torrance & Jackson, 2020). Learning experiences within such technologies provides a valuable method in military training contexts for tracking the "crawl" phase when the goal is primarily the acquisition of declarative knowledge. However, learning experiences may fall short of the data required for assessing detailed proficiency with procedural knowledge during live or simulated operations. Specifically, while an xAPI statement may document that a trainee completed a specific simulation event with the outcome score, the xAPI statement may lack important details about the underlying reasons for the training outcome. The risk of not capturing more granular context and performance data by employing only the xAPI standard is the inability to share real-time data associated with the multiple facets of trainee performance necessary to inform instruction related to process feedback or root cause analyses.

HPM Standard: Human Performance Modeling Language (HPML)

HPML was a draft Simulation Interoperability Standards Organization (SISO) standard specifically focused on the unique challenge of defining what data (i.e., computations, measures, assessments, results, and instances/periods) are required for automated HPM in a variety of DoD services and domains (Walker, Tolland & Stacy, 2015). This detailed data during simulated operations is primarily focused on the "walk" and "run" phases of training when procedural knowledge is acquired. HPML is also designed to ingest data that is relevant to or influences performance—like environmental, systems, or adversarial data that would impact mission difficulty. The added context can provide an instructor with contextual information to consider when providing feedback and assessments. While simulators produce large volumes of data, not all data is relevant to performance nor needed for comprehensive performance assessment. Lack of inclusion of the right data despite large available data may result in measurement challenges. For example, early efforts to leverage HPML for automated measurement resulted in the validation of 27% of all measures defined due to lack of critical data within the system (Wiese, Atkinson, Roberts, Ayers, Ramoutar, 2012).

Future research and development efforts will improve our ability to use simulator data to measure performance, however, outcome data will only ever be able to account for a portion of HPM. If we are to continue to expand our ability to accurately and comprehensively measure HPM, future efforts will have to extend beyond only simulator or aircraft data. Continual advances in eye tracking and physiological sensing technology, offer component capabilities that may assess unique variance in performance. While performance processes are difficult to assess without humans in the loop, novel video and audio capture technologies combined with emerging AI/ML analytic capabilities may offer advanced means to automatically capture both cognitive and behavioral processes that lead to performance outcomes. Such detailed process information is imperative for expediting learning by providing trainees the specific information they need to show improvement. If the DoD is to meet its aim as a data centric organization, standards will need to be adopted, iterated on, and utilized to inform requirements.

However, standards alone are insufficient for developing, transitioning, and maintaining enterprise level HPM systems in inherently complex DoD communities. Even when standards exist and are adhered to, advancing HPM systems for DoD communities and platforms is still complex. The average DoD community performs multiple mission sets, is multi-crewed, and employs training in classrooms, desktop computers, part-task training simulators, whole-crew training simulators, and during live flights (see Figure 2). Each crewmember is responsible for operating highly sophisticated systems. Important information related to performance can occur before and after operations during the planning, pre-brief, and debrief portions of their duties. Moreover, instructors, trainees, leadership, program managers, and many others all have a stake in this data. Given this inherent complexity, the development and transition of enterprise level HPM systems is a challenge. While monumental, this challenge is not unfamiliar to systems engineers who are specialists in implementing methodical multi-disciplinary approaches for designing, managing, and operating highly complex technical systems (NASA Systems Engineering Handbook, 2019). To facilitate successful integration of HPM standards, one approach is to concurrently consider an architecture such as model based systems engineering (MBSE) and consider the role it can play in defining and documenting requirements for HPM systems intended to assess multi-level performance throughout the training pipeline.

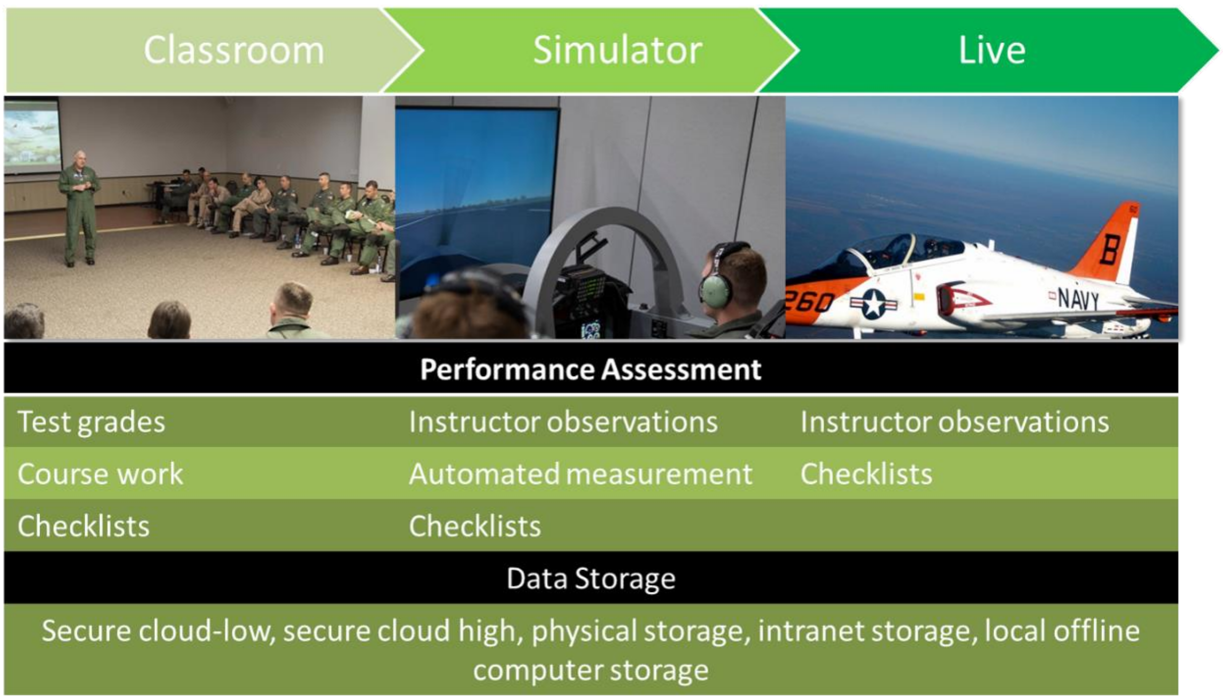

Figure 2. Navy aviation training typically follows a “crawl-walk-run" approach, where training begins in a classroom, before complexity is gradually increased across exposures offered in various part or full task simulation environments, and then ultimately offering live flight training. Each of these environments offer different opportunities for data types and sources related to trainee performance. DVIDS Images for aviation classroom, simulation and T-45 training (https://www.dvidshub.net/).

Architecture: Model Based Systems Engineering (MBSE)

One methodology used by systems engineers to support the lifecycle of complex systems is MBSE (Shevchenko, 2020). MBSE is not a new approach to systems engineering but one that has proven immensely beneficial as the complexity of systems grows. Not only is MBSE useful when establishing the requirements of future and current systems, it is equally useful for tracking system deficiencies and gaps that can be addressed with research and development investment in emerging technologies. An MBSE model is an ideal starting point to establish where connections need to be made within existing performance measurement systems to perform multi-level trend analyses that are of interest to a variety of stakeholders. Once established, these models would support all future updates and changes to paradigms to support resilient and adaptable systems design, combatting technology obsolescence. This last point is important to emphasize as military contexts are inherently dynamic necessitating frequent changes to individual performance measures and measurement systems as a whole. The sustainment and maintenance of performance measurement systems will require constant and consistent updating to ensure validity and relevancy of measures.

CONSIDERATIONS FOR STANDARDS AND ARCHITECHURE

As we conceptualize the next generation of training and operational systems, we must consider the vital role both learning and HPM data standards, as well as an architecture such as MBSE, will play in achieving expanded capabilities that address the goals of the DoD data strategy and improve proficiency. Specifically, xAPI and HPML provide the data points needed to deliver usable, real-time, and interoperable data, while an MBSE architecture provides a strategy for connecting this data in a meaningful way to facilitate valid and data centric decisions related to proficiency. However, emerging technologies considered as part of these standards and architecture will also play a role in advancing our ability to accurately and comprehensively assess human performance. In the following sections, we provide an overview focused on physiological monitoring, eye tracking sensors, and speech processing as a starting point for enhancing performance measurement. As technologies continue to mature, a broader look at alternative monitoring approaches should be considered.

Physiological Monitoring

The use of physiological sensors is not a novel concept. Sensors that monitor individual physiology such as heart rate, breath rate, and blood pressure have provided a wealth of data for supporting monitoring of individuals in medical or training situations to ensure safety. As such, a broad range of physiological sensors and derived measurements have been leveraged in research to better understand human performance in various contexts. Some of the more frequently used metrics include heart rate and heart rate variability, respiration rate, core or skin temperature, electrodermal activity (EDA), galvanic skin response (GSR), skin conductance level (SCL), and Electroencephalography (EEG)(Baig & Kavakli, 2019; Putz, Mertens, Chuang, & Nitsch, 2024). However, advances in technology to include reliability of sensors and ability to embed sensors in deployable configurations that are easier to use without significant training or calibration are creating an increased consideration for these technologies in training and/or operational environments where tracking and understanding human performance are important.

Physiological sensors have been leveraged in past research to provide objective methods for understanding changes to an individual's physiological state. For example, cognitive load or the amount of mental effort being used in working memory (Barrouillet, Bernanrdin, Portrat, Vergauwe, & Camos, 2007) has been correlated in previous research to heart rate variability and EEG (van Weelden, Alimardini, Wiltshire, & Louwerse, 2022). Physiological responses such as cortisol levels, GSR, and heart rate have also been correlated with changes in stress levels (Driskell & Salas, 2013). In extreme operations, fatigue is often a concern and research has previously correlated electrooculography and EEG to this construct. One challenge that remains for research in this area is how to best validate constructs given the typical overlap in physiological responses with a number of psychological constructs that may impact human performance. For example, while the relationship between heart rate variability and stress is substantiated in the aviation domain (Sekiguchi, et al., 1978), fluctuations in heart rate during flight operations may be as much the result of physical exertion as it is stress associated with operating in a threat environment. Parsing these two potentialities to better understand the underlying relationship with human performance is a complex analytic challenge that will likely require advanced analytic strategies such as machine learning.

Eye Tracking Sensors

Similar to physiological monitoring sensors, eye tracking capabilities have been on the market for some time. These sensors provide a method to collect and understand ocular metrics such as blink rate, gaze patterns, fixation time, fixations, saccades, and pupillometry. Similar to physiological sensors, technology advances in recent years have decreased the size and increased the data outputs of these sensors. Available products include screen-mounted eye tracking sensors or wearable glasses. Screen-mounted eye tracking sensors typically provide higher sample rate capabilities that increase the metrics available. Additionally, these solutions are coupled with a hardware screen setup that remains stationary, and if configured within the same operating system architecture, can support ease of analysis through synchronization of display features and the individual's ocular metrics. Alternatively, eye tracking glasses provide a wearable solution that is minimally intrusive and allows the individual to move freely in an environment.

This flexibility increases the use cases in which eye tracking can be employed but there remain challenges associated with data processing and analysis in complex and/or dynamic environments.

The benefit of considering eye tracking sensors as part of future human performance monitoring systems lies in the ability to leverage ocular metrics as a proxy to an individual's cognitive process or state. For example, fixations or a longer duration spent looking at a specific area of interest has been hypothesized to indicate deeper processing or increased cognitive load (Just & Carpenter, 1976). Research into saccade length may allow for understanding how shorter distances covered during rapid eye movements may indicate more efficient visual processing, similar to an aviators scan pattern. The more nuanced and detailed metrics associated with pupillometry may offer similar insights to other physiological sensors; for example, pupil dilation or changes in pupil size may indicate changes in cognitive load, workload, or engagement (Fehringer, 2021).

Speech Processing

Automatic speech recognition solutions to support dictation or other monotonous tasking have been evolving for decades. Early solutions required a significant amount of initial engineering investment for domain specific uses that included fully documenting a corpus of language and creating grammar structures that increased the reliability of recognition rates. These hard coded and rigid applications of speech recognition solutions typically required an operator in the loop to address reliability issues or system failures and required significant maintenance support to maintain systems to evolving vocabulary or grammars. Recent advances in deep learning and other techniques have exponentially matured automated speech recognition solutions, and with the wealth of data at the disposal of commercial systems for training and refinement (e.g., Apple's Siri, Amazon's Alexa, Google Gemini) have increased user adoption and application of the technology. Military domains offer some unique challenges outside these general use cases to include jargon or brevity terms that would not be recognized by commercial systems trained on generalized speech corpus, lack of available data for training, noisy and/or encrypted audio files, and a mix of protocol-based and natural language interactions. However, advances of artificial intelligence and data synthesis techniques to help increase the availability of training data may increase the ability to successfully employ these technologies in these environments.

In addition to applications of natural language processing for increasing the availability of contextual data related to human performance, speech processing algorithms provide an ability to provide a more comprehensive understanding of a training or operational environment. For example, variations in the cadence or pitch of speech may be indicative of cognitive processing or workload (MacPherson, Abur, & Stepp, 2017). Alternatively, consideration of integration of disfluency analysis or increased pauses in responses may indicate a lack of confidence in responses or altered physiological state like fatigue and sleepiness (de Vasconcelos, Vieira, Kecklund, & Yehia, 2019; Gao et al., 2022).

Recommended Standard Changes

In practice, it is recommended that a section of the HPML manual should be added to describe the considerations outlined above. The HPML manual is the appropriate venue for adding definitions of these terms and their considerations for use. This would be applied with a more narrative approach. By defining additional terms in HPML, they can be leveraged in the xAPI structure to track those concepts in a learning interaction. A few recommended definitions to describe in the standard are time of measurement, the type of measurement to be recorded (i.e., heart rate), the recording tool (i.e., wearable device name), and the unit of measurement (i.e., beats per minute). This logic can be expanded to behavioral and eye tracking measures, not just physiology. By recording data in this structure and level of depth with consistency across systems, there is a capability to collate and model data across events and trainees. In addition, describing the sampling rate and introducing recommendations for documenting data synchronization would also make this data more usable for MBSE applications.

The benefit of expanding the HPML manual and xAPI scripts to include physiological measurement-specific information is that it can be used as a consistent reference for those acquiring new training systems. One of the major goals of these standards is interoperability—with use cases to follow highlighting challenges with trying to model language without consistent collection across the lifelong learning experience.

NAVAL AVIATION USE CASES

As an example of how this standard update could be applied, we can use the example of a Student Naval Aviator in the undergraduate training pipeline. The undergraduate or primary syllabus has well-defined training events in simulation and flights and is quantified at the individual level. Recently, extended reality (XR) devices have been adopted into the syllabus allowing for additional human performance quantification via embedded eye tracking capabilities and interaction with natural language processing software for communications training (see Figure 3). There are advantages to capturing performance both within a given training event and learning over weeks in training. From an MBSE and HPM perspective this is equivalent to the "walk" stage of the "crawl, walk, run" analogy often used in training.

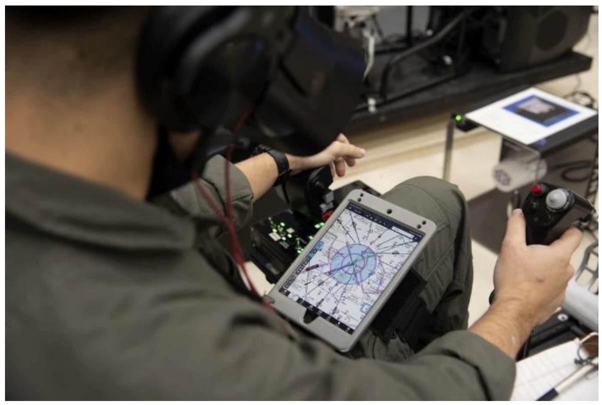

Figure 3. Example of a Student Naval Aviator completing a primary simulator event using an XR device. *Naval Aviation Training Next - Project Avenger ground school, by LCDR Michelle Tucker,* https://www.dvidshub.net/image/6461467/naval-aviation-training-next-project-avenger-ground-school

The updates to the HPML language can be incorporated by defining the eye tracking structure to follow—with an example of a measurement being eye tracking/gaze behavior, recorded via embedded XR device eye tracker, with the measure of interest being fixation duration in milliseconds, and the goal is to quantify development of a sight picture for the pilot. The same information can be stored for voice data—with an embedded microphone, measuring speech rate, with words per minute rate for speed produced, capturing communications completion rate and potential confidence or fatigue during practice of a given skill. This would then be used in xAPI statements to be included as an LRS for this given training event. By doing so, we could ensure consistent storage and labeling of the eye tracking measurement to see how behavior changes as the Student Aviator advances in their training and compare that to other students completing the same event. In order to do this, an architecture is needed to retain the context of the event in a consistent measurement framework for all the trainee and event-level details.

After pilot proficiency is established, there is a challenge to capture and model how this performance can feed into and be leveraged for complex mission scenarios. One example we can use for this is understanding the role of the Naval Flight Officer and other aircrew members as they start to work together as a team (see Figure 4). When we add this layer of complexity, the environment gets more challenging as people are often operating as a group and in a more adaptable environment over a longer period of time. The syllabus for assessing learning and performance changes from one that is very repeatable and geared toward assessing one student at the undergraduate pilot training level. That does not necessarily mean that the same data architecture and data standard should not apply to the more complex scenario, but recommended changes to human performance modeling should also account for what data is needed beyond the individual and early trainee. Once we expand to multiple crew stations and more variable mission scenarios, the fidelity of our human performance modeling extends into the "run" phase of "crawl, walk, run".

Figure 4. Example of P-8A Poseidon crew working collaboratively on a mission with a team-based environment. *Navy P-8 Poseidon crews enhance maritime partnerships during rotational detachments to Singapore, by PO2 Joshua Fulton,* https://www.dvidshub.net/news/204022/navy-p-8-poseidon-crews-enhance-maritime-partnerships-during-rotational-detachments-singapore

There has been a growing interest in modeling team coordination and team situational awareness with psychophysiological measures, which has also advanced significantly with data analytic methods and improvements to wearable physiological monitoring tools. In order to gather team information, the data architecture and core measurement principles from measuring individuals must be first met. The concept of "*the whole is greater than the sum of its parts*" applies to the world of mission-oriented data standardization for teams as well. In order to optimize team performance and training, it is necessary to first understand the individual in that context and then the team at large's "interpersonal autonomic physiology"—the relationship between their physiology with interacting together (Palumbo et al., 2017). Gathering autonomic measures during a team activity and examining team state during a shared exercise has been utilized as an indicator of cognitive readiness and situational awareness (Walker, Muth, Switzer, & Rosopa, 2013). While gathering this data, it is important to leverage subject matter experts (SMEs) to identify the variables in the mission that may influence both team proficiency and psychophysiological response. One of these examples is identifying and retaining information about the interdependent dyads that exist in the system—a tactical coordinator and acoustics operator in a P-8 (shown in Figure 4). They rely on each other to complete the overall mission tasking with shared data, heavy communication, and repeated tasking together. Psychophysiological measures such as heart rate variability/inter-beat interval, skin conductance, and respiration rate have been examined to capture dyad synchronization (Palumbo et al., 2017).

In order to scale HPM standardization to a more complex level, lessons from software can be leveraged to support basic system architecture and establish an MBSE approach to human performance. One example of this is creating HPM tools that follow a modular open systems approach (MOSA) or containerization model, such that they are resilient to additional sensors or algorithms being added (Department of Defense, 2019). There is also the capability to retain meta-data and time synchronize the outputs of each sensor to tie the complex system of sensors and performance data to a usable format for future analysis. The advantage of using a MOSA or containerization approach is that the individual person and individual sensor output can be preserved for better understanding of the contributions of each. One way this could be integrated into HPML and xAPI standards would be to include ties to the existing acquisition language on how to ensure MBSE and MOSA language is included in novel system design- particularly at the Analysis of Alternatives review for acquisition, specified:

> *Requirement to address modular open system approach in program capabilities development and acquisition weapon system design. (c) ACQUISITION STRATEGY.—In the case of a major defense acquisition program that uses a modular open system approach, the acquisition strategy required under section 2431a of this title shall— ''(1) clearly describe the modular open system approach to be used for the program; ''(2) differentiate between the major system platform and major system components being developed under the program, as well as major system components developed outside the program that will be integrated into the major defense acquisition program; ''(3) clearly describe the evolution of major system components that are anticipated to be added, removed, or replaced in subsequent increments; ''(4) identify additional major system components that may be added later in the life cycle of the major system platform; ''(5) clearly describe how intellectual property and related issues, such as technical data deliverables, that are necessary to support a modular open system approach, will be addressed; and ''(6) clearly describe the approach to systems integration and systems-level configuration management to ensure mission and information assurance. (Department of Defense Acquisition Agility, 2020)*

While this section was written toward major weapon systems acquisition, the capability can and should be considered in training systems design as well. The concepts of MOSA link to the overall goal toward MBSE approaches and the general structure of the LRS framework from xAPI. The Office of the Under Secretary of Defense for Research and Engineering also published MOSA assessment criteria—one of which is to reference open standards published by recognized organizations to encourage system developers to reference for supporting more robust systems development (Geier, 2022). The work conducted on MBSE and MOSA development in other acquisition systems can be expanded upon in the human performance domain to take advantage of other software engineering approaches to network development to address gaps for scenario complexity in the advanced training domain.

A major challenge to adding psychophysiological measurement and more emphasis on advanced learning to HPM standards is the potential of new approaches or tools leading to the obsolescence of a given approach. MOSA and containerization are examples of MBSE software approaches that allow for resilience of the system to changing of the subsystem components (Morgan, Holzer, & Everleigh, 2021). While this has traditionally been a best practice for using products from multiple vendors and allowing programs to adapt to add new software features over time (Zimmerman, Ofori, Barrett, Soler, & Harriman, 2019), the same logic applies to an advanced training scenario. The challenge applied in the multi-crew advanced training is both the number of people who may be working in different team compositions and the tie of data on performance from multiple different means of collection (traditional LRS-types of data collection, physiological measurement, simulator context and performance). As these systems improve over time, a modular architecture will allow stakeholders to add or remove the unnecessary pieces without interruption to the trainee or data loss. With future analytic methods and real-time simulation adaptability tailored to the learner, methods like MBSE and MOSA get at the requirements for interoperability needed to advance this technological capability.

RECOMMENDATIONS

In order to support the increase in training and human performance data being gathered, standardization updates are required to meet the demand of the new data environment. The benefit of modifying these standards allows for clear consensus and traceable language to use in the acquisition of new HPM tools and training systems. Additionally, by standardizing the collection of this information there is potential to create more robust models of human performance using AI/ML approaches and MBSE. There have been considerable advances to psychophysiological recording tools as well as overall systems design that warrant an update to our standards to enable robust acquisition of tools for capturing human performance in operational settings.

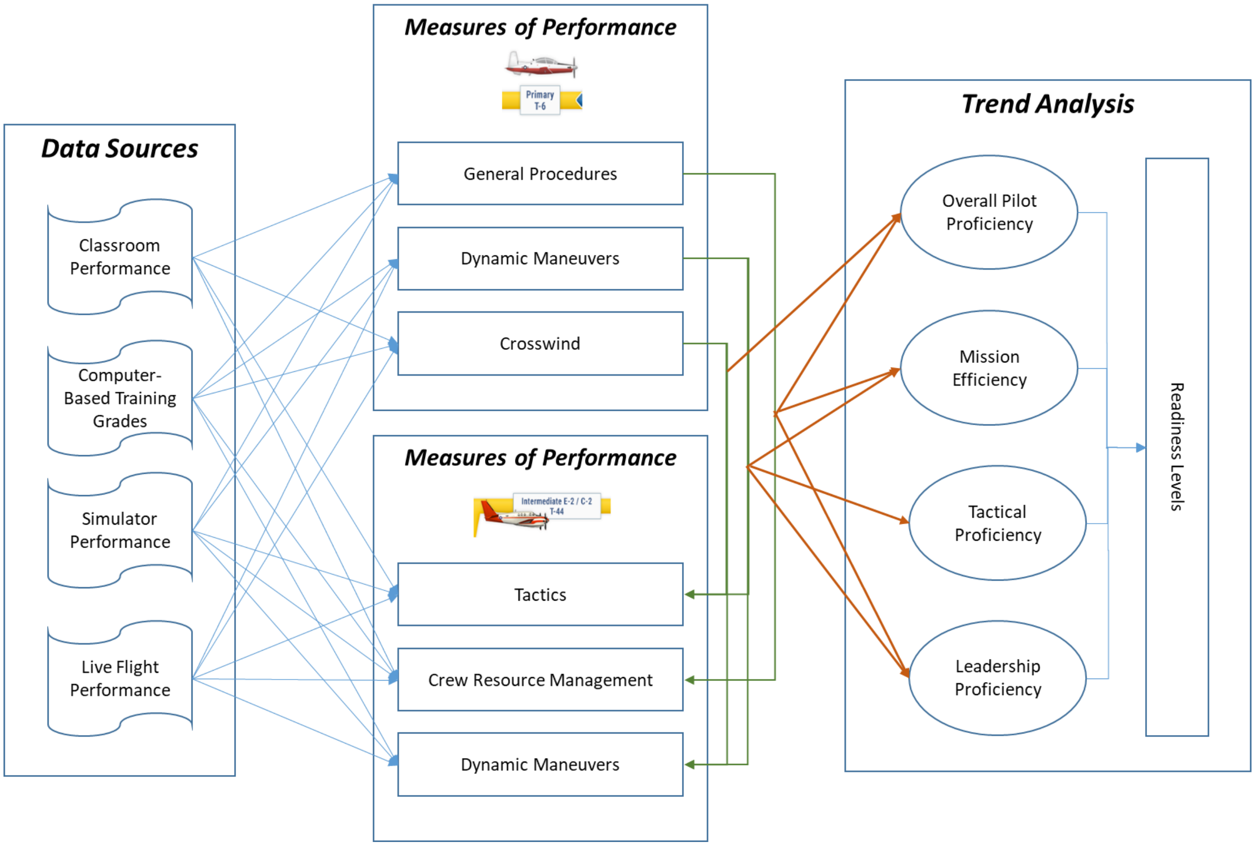

Figure 5. Combining data via standards supports capture of reliable and interoperable data across a variety of sources at multiple points in the training progression. Through the addition of an MBSE architecture that maps data sources to measures of performance, comparison of related building blocks in performance can be achieved, which ultimately provides greater insight via trend analysis that is built on an enterprise level data management solution.

In summary, the xAPI and HPML standards would benefit from language specific to psychophysiological recording, specifying up-front the minimums for certain sensors and the addition of these measures from a data architecture perspective. Modern approaches to MBSE, with methods like MOSA, have enabled more resilient software infrastructure to enable data-driven solutions. In further pursuit of building in these software concepts and enabling MBSE in the future, there has been tri-service agreement on pursuing MOSA to enable resilience in our systems acquisition. This memo presents the call to action, "MOSA supporting standards should be included in all requirements, programming, and development activities... to the maximum extent possible" (Department of Defense, 2019). To accelerate development of scalable human performance solutions, MOSA-specific MBSE language should also be included in the HPML and xAPI standards as we scale these concepts into larger, more complex training environments. We have provided the basis for the language to be updated and expanded upon in the future. The overall goal is to align the goals of HPM across the training and operational environment to capturing life-long learning as well as enable more data-driven decisions to be made with robust systems design.

APPROVAL FOR RELEASE

The views expressed in this paper are solely those of the authors, and do not necessarily reflect the opinions of the Naval Air Warfare Center Aircraft Division, or any other Department of Defense agency, unless stated in official directives. This work has been sponsored by the NAVAIR Naval Innovative Science and Engineering (NISE) program. NAVAIR Public Release 24-0443 Distribution Statement A – Approved for public release; distribution is unlimited.

REFERENCES

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 570.

Department of Defense (2019, January 7). Modular Open Systems Approaches for our Weapon Systems is a Warfighting Imperative. https://www.dsp.dla.mil/Programs/MOSA/

Department of Defense (2020, September 30). Department of Defense Data Strategy. https://media.defense.gov/2020/Oct/08/2002514180/-1/-1/0/DOD-DATA-STRATEGY.PDF

Department of Defense Acquisition Agility (2020). Code 10. Sec 805. Ch. 144B. § 2446a. Requirement for Modular Open System Approach. https://www.cto.mil/sea/mosa/

de Vasconcelos, C. A., Vieira, M. N., Kecklund, G., & Yehia, H. C. (2019). Speech analysis for fatigue and sleepiness detection of a pilot. Aerospace medicine and human performance, 90(4), 415-418.

Fehringer, B. C. (2021). Optimizing the usage of pupillary based indicators for cognitive workload. *Journal of Eye Movement Research, 14*(2).

Gao, X., Ma, K., Yang, H., Wang, K., Fu, B., Zhu, Y., ... & Cui, B. (2022). A rapid, non-invasive method for fatigue detection based on voice information. *Frontiers in Cell and Developmental Biology, 10,* 994001.

Geier, N. (May, 2022). TECHNICAL REPORT: OUSD(R&E) MOSA Assessment Criteria. https://www.cto.mil/wp-content/uploads/2023/06/MOSA-Assess-2022.pdf

IEEE Standards Association. (2021). What are standards? Why are they important? https://standards.ieee.org/beyond-standards/what-are-standards-why-are-they-important.

IEEE 9274.1.1. (2023). IEEE Standard for Learning Technology—JavaScript Object Notation (JSON) DataModel Format and Representational State Transfer (RESTful) Web Service for Learner Experience Data Tracking and Access. IEEE SA.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology, 8*(4), 441-480.

MacPherson, M. K., Abur, D., & Stepp, C. E. (2017). Acoustic measures of voice and physiologic measures of autonomic arousal during speech as a function of cognitive load. *Journal of Voice, 31*(4), 504-e1.

Morgan, M., Holzer, T., & Eveleigh, T. (2021). Synergizing model-based systems engineering, modularity, and software container concepts to manage obsolescence. *Systems Engineering, 24*(5), 369-380.

Nouira, A., Cheniti-Belcadhi, L., & Braham, R. (2018). An enhanced xAPI data model supporting assessment analytics. *Procedia Computer Science*, *126*, 566-575.

Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., & Goodwin, M. S. (2017). Interpersonal autonomic physiology: A systematic review of the literature. *Personality and social psychology review*, *21*(2), 99-141.

Poeppelman, T. R., Hruska, M., Ayers, J., Long, R., Amburn, C., & Bink, M. (2013, December). Interoperable performance assessment using the Experience API. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference*.

Pütz, S., Mertens, A., Chuang, L., & Nitsch, V. (2024). Physiological measures of operators' mental state in supervisory process control tasks: a scoping review. *Ergonomics*, *67*(6), 801-830.

Sekiguchi, C., Handa, Y., Gotoh, M., Kurihara, Y., Nagasawa, A., & Kuroda, I. (1978). Evaluation method of mental workload under flight conditions. *Aviation, space, and environmental medicine, 49*(7), 920-925.

Shevchenko, N. (2020). An introduction to model-based systems engineering (MBES). SEI Blog. https://insights.sei.cmu.edu/blog/introduction-model-based-systems-engineering-mbse/.

Torrance, M. & Jackson, J. (2020). Get a piece of the xAPI pie: You may have heard about the benefits of xAPI; here's how to put it into practice. Association for Talent Development. https://www.td.org/magazines/td-magazine/get-a-piece-of-the-xapi-pie.

Torrance, M. & Wiggins, C. (2016). What is x API? Association for Talen Development https://www.td.org/magazines/what-is-xapi.

van Weelden, E., Alimardani, M., Wiltshire, T. J., & Louwerse, M. M. (2022). Aviation and neurophysiology: A systematic review. *Applied ergonomics*, *105*, 103838.

Walker, A. D., Muth, E. R., Switzer III, F. S., & Rosopa, P. J. (2013). Predicting team performance in a dynamic environment: A team psychophysiological approach to measuring cognitive readiness. Journal of cognitive engineering and decision making, 7(1), 69-82.

Walker, A., Tolland, M., & Stacy, W. (2015). Using a human performance markup language for simulator-based training. In *Simulation Interoperability Standards Organization (SISO) Simulation Interoperability Workshop (SIW) Meeting. SISO-Simulation Interoperability Standards Organization, Orlando*.

Wiese, E., Atkinson, B., Roberts, M., Ayers, J., & Ramoutar, D. (2012). Automated human performance measurement: Data availability and standards. In *Proceedings of the 2012 Interservice/Industry Training, Education, and Simulation Conference, Orlando, FL*.

Zimmerman, P., Ofori, M., Barrett, D., Soler, J., & Harriman, A. (2019). Considerations and examples of a modular open systems approach in defense systems. *The Journal of Defense Modeling and Simulation*, *16(*4), 373-388.