Performance Measurement in LVC Distributed Simulations: Lessons from OBW

Jaclyn Hoke
Rockwell Collins, Inc.
Cedar Rapids, IA
Jaclyn.Hoke@rockwellcollins.com

Lisa Townsend, Sam Giambarberee
Naval Air Warfare Center Training Systems Division (NAWCTSD)
Orlando, FL
lisa.townsend@navy.mil
sam.giambarberee@navy.mil

Sae Schatz
Advanced Distributed Learning (ADL) Initiative
Washington, D.C.
sae.schatz@adlnet.gov

ABSTRACT

Operation Blended Warrior (OBW) 2016 marked the second year of a three-year effort to document lessons learned and understand barriers to implementing Live, Virtual, Constructive (LVC) distributed training. In the first year of the event, LVC focus areas included connectivity, interoperability, data standards, after-action review, and cyber security. Year two introduced additional focus areas: multi-level security, cross domain solutions, long-haul feeds, and performance measurement. This paper focuses on this latter area—defining and collecting performance measures.

Performance measurement in simulation-based training faces formidable obstacles, including the identification of individual and collective performance dimensions, how these dimensions relate to training goals, and how training transfers to operational readiness. Blending of LVC elements introduces additional complexity, not only for human performance assessment but also for evaluating the effectiveness and efficiency of the technical system.

In this paper, we present the measures defined and collected during OBW in four primary areas: 1) cost analysis, 2) network performance, 3) trainee performance, and 4) whether OBW met the expectations of participating organizations. We also discuss three categories of Measures of Effectiveness (MoEs) and Measures of Performance (MoPs) established by the OBW Strategic Integrated Product Team: Programmatic, Technological, and Learning. These MoEs and MoPs will facilitate annual comparisons of performance measurement at OBW and encourage use of the event as a sandbox to design and validate LVC performance measurement tools. Finally, we present the goals and measures established for Performance Measurement during OBW 2017 and recommendations for future events.

ABOUT THE AUTHORS

Jaclyn Hoke, Ph.D., is a Principal Engineer in the Rockwell Collins Advanced Technology Center with 10 years of experience in research and development. She has a doctorate in Computer Engineering. She is the technology lead for Rockwell Collins research initiatives in Cognitive State Assessment and Training Effectiveness, and served as the OBW Operational IPT Lead for Performance Measurement.

Ms. Lisa Townsend is a Research Psychologist at the Naval Air Warfare Center Training Systems Division (NAWCTSD), Human Integration and Performance Division, Applied Training Research branch in Orlando, FL. She has 22 years of military research and Instructional System Design (ISD) experience and an MSc in Industrial/Organizational (I/O) Psychology from the University of Central Florida (UCF), specializing in the areas of team performance and team training.

Mr. Sam Giambarberee is a Computer Engineer at NAWCTSD, Advanced Simulation Division, Technology Development and Integration branch in Orlando, FL. He has an MSc in Computer Science from UCF, and specializes in networking and developing distributed simulation systems. He is a lead software and project engineer, with 25 years of experience supporting distributed simulation projects for the Navy and Marine Corps.

Sae Schatz, Ph.D., serves as the director of the Advanced Distributed Learning (ADL) Initiative. She is also an applied human–systems scientist, with an emphasis on human cognition and learning, instructional technologies, adaptive systems, human performance assessment, and modeling and simulation. She holds a doctorate in Modeling and Simulation from UCF.

Performance Measurement in LVC Distributed Simulations: Lessons from OBW

Jaclyn Hoke
Rockwell Collins, Inc.
Cedar Rapids, IA
Jaclyn.Hoke@rockwellcollins.com

Lisa Townsend, Sam Giambarberee
Naval Air Warfare Center Training Systems Division (NAWCTSD)
Orlando, FL
lisa.townsend@navy.mil, sam.giambarberee@navy.mil

Sae Schatz
Advanced Distributed Learning (ADL) Initiative
Orlando, FL
sae.schatz@adlnet.gov

INTRODUCTION

Effectively training our warfighters has become increasingly challenging in the current budgetary environment. Furthermore, as the capabilities of our adversaries continue to grow and operational contexts expand into "contested environments," this introduces new sets of challenges for military training. It is difficult, if not impossible, to replicate these operational contexts within a purely live training environment, and range spaces have become too constrained to support the advanced capabilities of our forces and advanced threats of our adversaries. These trends are driving considerable interest throughout all branches and levels of our military, and across industry, in maturing Live-Virtual-Constructive (LVC) distributed training capabilities. The blending of LVC elements promises increased training fidelity and density at a reduced cost, but it can also introduce complexity in many areas.

Establishing an integrated LVC environment has been the subject of considerable research and development for several decades, yet we continue to face many of the same technical challenges today. These issues motivated the I/ITSEC community to institute Operation Blended Warrior (OBW) as part of the conference. OBW is a multi-year representative LVC event that provides a collaborative environment for government, industry, and academia to document lessons learned and better understand the barriers to implementing LVC distributed training (National Training and Simulation Association, 2016). By utilizing a networked architecture on the exhibit hall floor, organizations involved in OBW could explore challenges in several focus areas, push the boundaries of the state-of-the-art in LVC, and showcase their capabilities to senior leadership in the Department of Defense (DoD).

OBW 2015 marked the first year of the event. Approximately thirty participating organizations in twenty-two booths on the show floor examined focus areas such as connectivity, interoperability, data standards, after-action review, and cybersecurity during five 90-minute exercise "blocks." The storyline unfolded throughout the event, and it spanned land, air, maritime, and cyber domains. In 2016, the event grew to nearly fifty participating organizations in approximately forty booths on the show floor. The second year of the event followed a similar format as the previous year, and it added additional focus areas: multi-level security, cross-domain solutions, long-haul feeds to remote (off show floor) participants, and performance measurement.

This paper focuses on the challenges and lessons learned observed during OBW in defining and collecting performance measures. The paper begins by describing the initial findings from measures collected in OBW 2015 and 2016. Then it describes a framework for comprehensively defining what performance measures could be included in LVC, in general. Finally, it summarizes how these results will influence performance measurement in OBW 2017.

OBW 2016 PERFORMANCE MEASUREMENT FOCUS AREAS

Simulation-based training emulates aspects of the "real world" environment, providing a safe and cost-effective environment to develop core competencies. However, one of the most important and most challenging aspects of simulation is defining and measuring performance to ensure trainees acquired the *correct* competencies (Salas, E., Rosen, M., Held, J., & Weissmuller, J., 2009). Effective performance measurement faces many obstacles, including identifying what dimensions to measure, how to measure these dimensions, and how to assess training transfer for operational readiness.

Distributed LVC simulation exercises, such as OBW, present additional measurement challenges assessing the technological effectiveness and efficiency of the system. To help address these, a collaborative team of industry and DoD subject-matter experts collected performance measures during the OBW 2015 and 2016 events on cost analysis, network performance, trainee performance, and exercise planning.

Cost Analysis

Clearly, increased use of simulation-based training extends the life of the corresponding live platforms, reduces their maintenance costs, has a lower cost-per-hour of operation than the live assets, and may offer many other cost-associated benefits. Unfortunately, simulation-based training also incurs costs for maintaining the simulation devices and for providing necessary upgrades to support effective training for a broader range of skills and mission tasks. Furthermore, sizable near-term investments will likely be required to extend the capabilities of the live platforms, simulation devices, and infrastructures to support LVC distributed training exercises. Consequently, accurate cost forecasts for LVC training are critical. However, characterizing the cost savings and return on investment for LVC training is not straightforward, and as we continue to struggle to quantify these cost savings, DoD acquisition authorities continue to lack the necessary justification for these critical investments.

Thus, one of the performance measurement goals for OBW 2016 was to calculate the cost of conducting similar training in a "live" setting and comparing those figures to the costs associated with LVC exercises. This cost analysis represents a stepping stone for characterizing the return on investment associated with LVC training. Two participating companies, Calytrix Technologies and PricewaterhouseCoopers LLP, attempted to perform cost analyses for OBW 2016. Relevant data on operating costs and costs-per-item, such a munitions, were collected and used to generate the models for calculating the comparative costs.

Calytrix Technologies used their LVC Cost Counter tool, which calculates costs in real time by monitoring the Distributed Interactive Simulation (DIS) and High-Level Architecture (HLA) network traffic. The tool stored the information on munitions cost, platform operating costs by hour and/or distance traveled, daily maintenance rates, platform damage costs, daily salary costs, and similar factors in a relational database. When specific activities occurred in the simulation, their cost data were retrieved and added to the overall event total. PricewaterhouseCoopers utilized a similar costing model approach but performed their analysis *post hoc*, after the conference. Unfortunately, the two analyses generated vastly different overall cost estimates for the OBW event.

The cost analysis effort was hampered by several key factors: obtaining appropriate cost data for all relevant items, unanticipated entity types on the network, and an unclear timeline within the scenario. An enumerations list of entity types was established during the OBW planning process; however, it was finalized only a few weeks prior to the event, which did not leave sufficient time to collect and incorporate all of the cost data into the models. Additionally, cost data were not readily available for all entity types involved in the event, meaning the cost models developed could not represent the entire picture. Furthermore, entity types not included in the enumerations list were still observed on the network, potentially affecting cost accuracy. Finally, no firm correlation between the "simulation time" and an elapsed real-world time was established during the planning process. The OBW scenario unfolded over seven and a half hours (five 90-minute blocks) during the conference, but in an operational context, those events would have unfolded over several days or weeks. While a cost analysis should relate the OBW timeline to that of an all-live training context, as opposed to the operational context, no relationships were ever established between any of these timelines.

OBW organizers can address the many issues encountered in the cost analysis as part of future planning processes. For instance, establishing the entity list early in the planning processes, and preventing additions after the deadline, will provide the necessary time to collect the cost data and generate the associated models. Industry partners can also increase collaboration with service representatives to access the required cost data. Unfortunately, short of filtering all network traffic, it is unlikely we can guarantee that entity types not identified in the enumeration list will not be present during the exercise. Finally, the relationship between elapsed simulation time and time in the live training context should be established as part of the scenario definition and planning discussions.

Network Performance

Effective training in distributed simulation environments is founded upon an expectation of consistency in the environment as observed by the participants. A stable and reliable computer network is the cornerstone of providing this consistency in LVC training exercises, creating the illusion trainees are interacting in the same world. The network provides real-time communication of the DIS Protocol Data Units (PDUs) that represent the position and status of each entity and the interactions between them. Communication of these PDUs creates multiple synchronized copies of the training environment for participants, no matter where their training systems are located. Degraded network performance can negatively impact the entire training event. Bandwidth limitations, latency, and malformed PDUs affect the completeness and synchronicity of the training environment (to the degree that it can be observed by the trainees) and may impede achievement of training objectives.

The 2015 and 2016 OBW events used dedicated "government-like" networks on the I/ITSEC show floor, which replicated a typical DoD training network. The OBW network consisted of separate unclassified Virtual Local Area Networks (VLANs) for HTTP traffic, video feeds, DIS high and low side simulation traffic (to support participants demonstrating cross-domain solutions), HLA, cyber, and several others. This separation isolated network traffic, improved reliability, and supported rapid assessment and intervention when issues arose.

Networking guidelines were established to support interoperability and ensure the success of OBW events. OBW organizers performed network analysis during two pre-conference staging activities (integration testing) and during the OBW exercises to verify each participant adhered to the guidelines. Unfortunately, several participants did not attend the staging activities, which resulted in last minute challenges during the OBW event. To resolve these integration challenges, the NAWCTSD networking team used network analysis tools to determine the source of the issues and then either to correct the issue or, as a last resort, isolate the participant from the network.

During all five blocks of OBW 2016, the networking team monitored performance using the Wireshark network analysis toolset. Packets were analyzed for bandwidth, throughput, content, and errors. Network analysis was briefly performed on each VLAN, but most had few anomalies and minimal bandwidth usage (a few hundred packets/second) noted before and during the exercise. The majority of the packets were generated on the DIS high side network, so monitoring efforts were concentrated on this VLAN more than the others.

During the first block of the OBW 2016 event, shown in Figure 1, over 1 million packets were analyzed in about 9 minutes, resulting in an average packet rate of 2007 packets/second. The majority of these packets (61%) were non-DIS packets, meaning only 39% of the traffic on the DIS VLAN was valid. The valid DIS traffic consisted mostly of Entity State and Signal/Transmitter PDUs, with a small percentage (<3%) of Event; identification, friend or foe (IFF); and other PDUs. The non-DIS traffic consisted of packets using the non-approved protocols, ports, or IP addresses for the DIS VLAN. The largest segment of non-DIS traffic took up 32% of the packet total; it consisted of unknown data sent on port 3010 to a non-approved IP address (192.168.1.255). The next largest segment of non-DIS traffic took up 21% of the packet total, and was unknown data sent on port 3494 to a non-approved multicast address (225.6.7.8). Due to the high packet rate, these additional packets proved troublesome to some participants and resulted in miscues and system problems.

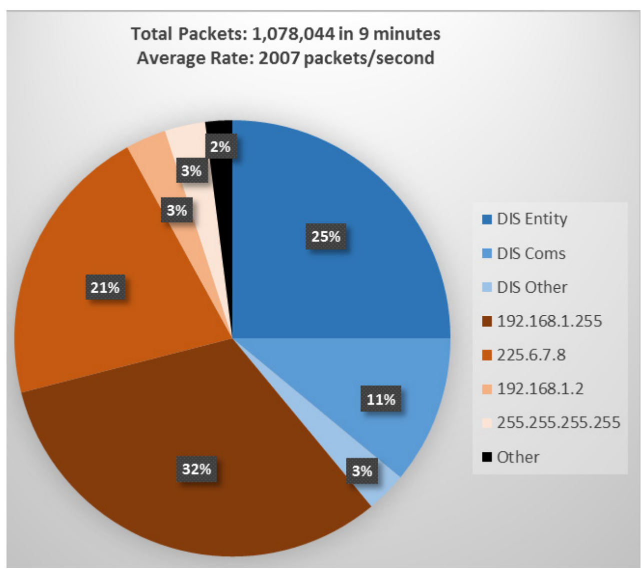
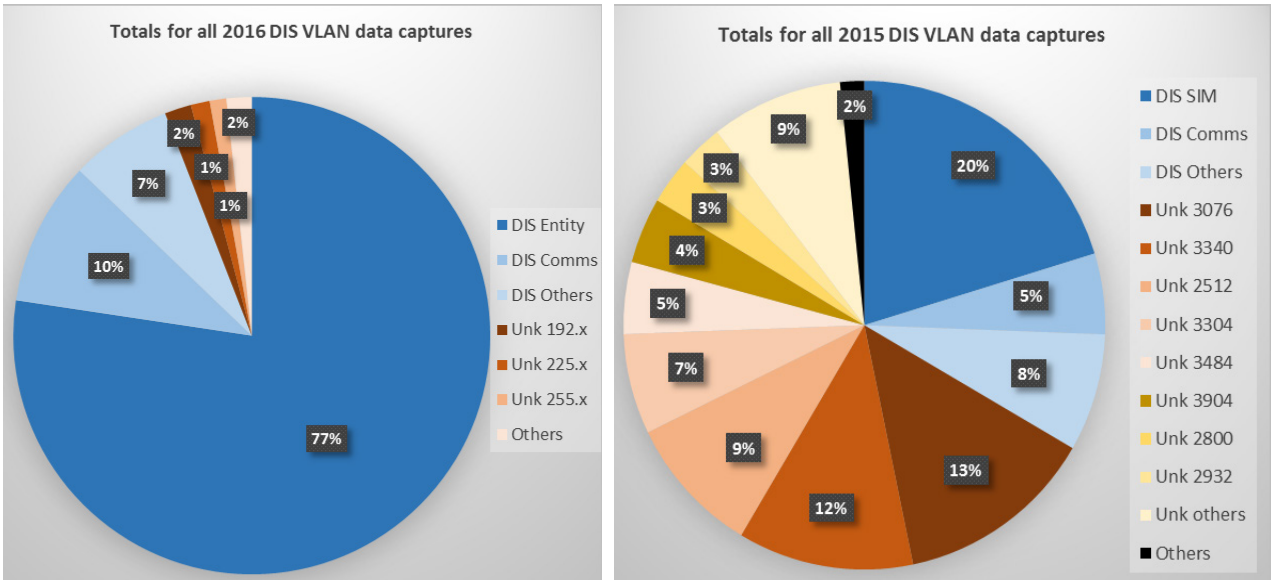


Figure 1. Analysis of DIS VLAN Data Captured During Block 1 of OBW 2016

As the event progressed throughout all five exercise blocks of OBW 2016, the non-approved traffic reduced significantly. As the weeklong totals for the DIS VLAN (Figure 2a) demonstrate, by the end of the week more than

94% of the network traffic consisted of valid DIS packets, mostly Entity State PDUs (77%) and Signal/Transmitter PDUs (10%), with a small percentage (7%) of other valid PDUs. The non-DIS traffic was virtually non-existent by the last exercise block. This is an immense improvement from the weeklong totals of the 2015 OBW event, shown in Figure 2b, where only 33% of the network traffic contained valid DIS packets and 67% of the DIS VLAN network traffic consisted of invalid packets using non-approved ports, protocols, and IP addresses.



(a) (b)

Figure 2. Side-by-Side Comparison of the Weeklong Totals for DIS VLAN Data Captured During (a) OBW 2016 and (b) OBW 2015

Trainee Performance

The human performance assessment includes evaluation of behavioral outcomes that indicate how well a trainee performs a set of tasks in the training environment (Salas et al, 2003). Human performance assessment results directly inform the success of training. Unfortunately, many current measures of training success rely upon subjective assessment by instructors, and personnel readiness is often determined in terms of time (e.g., flight hours) and qualification tasks. However, significant improvements are being made in this area.

Several participating organizations collected outcome-based measures of trainee proficiency during OBW 2016. Saab Defense and Security USA used a distributed observation and analysis tool, called AKKA, to monitor the DIS network and collect subjective and objective performance metrics. In particular, AKKA collected measures of performance for Joint Terminal Attack Controllers participants. The Air Force Research Laboratory also employed the Performance Evaluation Tracking System tool to passively observe network traffic to track the performance parameters of airborne platforms including fast jets, helicopters, and Unmanned Aerial Systems (UAS). Finally, Aptima, Inc. used their PM-Engine tool to collect observer measures during air-to-air and close air support engagements, collecting data on factors such as force preservation (evading enemy fire) and efficiency of fires (weapons employment).

While critically useful, all of these performance measures are scenario-dependent and only target specific types of trainees. More universal trainee performance measures that remain relevant throughout the entire training exercise are also desired. Aptima collected two such measures: training rule violations and rules of engagement (RoEs) violations.

Training rule violations represent a unique area of interest for LVC, where unique safety challenges arise when all three elements are present. Live-only training exercises strictly adhere to a set of training rules intended to maintain safety, but participants often ignore these rules during virtual-only exercises. In an LVC environment, all entities should obey the training rules to maintain the safety of the live participants. Developing a set of real-time measures for training rule violations not only provides feedback to trainees, but also provides assistance to Range Controllers and Range Safety Officers. Specific training rules monitored during OBW included violations of physical separation requirements and incidents of supersonic flight over land when not performing a tactical maneuver. So called “sim-isms,” such as “warping” in and out of the battlespace, were also monitored, as they can result in both negative training and safety hazards.

During OBW 2016, there were 555 violations of the two primary training rules (489 physical separation violations and 66 incidents of supersonic over land), with the majority of each performed by a small number of operators. Just three operators performed 50% of the physical separation violations, and 77% of the supersonic flights over land were performed by 2 operators. These results indicate that most participants adhered to the training rules. However, the 10,857 instances of inappropriate warping were performed by 287 operators, with the majority of individual operators performing 20 or more violations each. Further analysis is required to fully understand the underlying causes of so many violations.

RoE violations in training may be attributed to the “online disinhibition effect” (Suler, 2004), which describes how online environments create a perceived sense of safety that may result in behaviors not usually performed in face-to-face interactions. Parallels to this phenomenon occur in LVC training, e.g, violating RoEs by performing a risky munition delivery despite (virtual) collateral damage. While trainees rarely violate RoEs willfully, LVC environments provide an enhanced opportunity to practice RoEs. LVC also presents opportunities to train battle commanders in the associated decision-making processes, including escalating and de-escalating hostilities. For instance, during OBW 2015, the scenarios quickly escalated into all-out “kinetic” warfare. The addition of the battle commander role to OBW 2016, with associated performance measures, maintained a scenario evolution more representative of real-world operations. Specific measures of RoE violations that were tracked during OBW 2016 included unprovoked hostilities and collateral damage inflicted during engagements with military targets.

There were no instances of unprovoked hostilities observed during OBW 2016. This outcome suggests that the military (active, reservists, and retired personnel), civilian, and contractor participants maintained awareness of, and adhered to, the RoEs as the scenarios unfolded. Nonetheless, 85% of the tactical engagements during OBW 2016 still resulted in collateral damage. Further analysis is required to fully understand the underlying causes for this high percentage of collateral damage during the engagements with military targets.

Exercise Planning Metrics

A key lesson from OBW 2015 was the need to establish better measures of success for future events, particularly industry partners’ perception of its planning and execution. These data were collected through two surveys administered at the culmination of OBW 2016 (and planned for the culmination of OBW 2017). Organizations that helped develop scenario vignettes for OBW 2016 completed surveys on their expectations and overall reactions to the event. The list of expectations in the survey originated from initial meetings with OBW 2016 industry partners. Many expected outcomes were related to steps in the planning process, both leading up to, and during, I/ITSEC.

Expectations Met Survey

The Expectations Met Survey contained 30 questions. The first two pertained to participation in OBW 2015, and the remaining 28 questions focused on how well specific expectations were met during OBW 2016. The survey was partially, or wholly, completed by a total of 34 respondents, of which approximately 88% were from industry and almost 12% were government. Within industry, organizational roles were categorized by management (senior management or management) or technical (senior technical staff or technical staff). The majority of respondents overall were in a management role (29%), but those in one of the two technical roles responded to more questions than those in both management roles (47.1% compared to 41.2%). This makes sense, as most of the expectations were technologically focused. The majority of respondents (73.5%) had participated in OBW the previous year. This means that most respondents had baseline OBW knowledge and experience that may have contributed to more perceptive ratings of whether expectations were met or not. Questions that were focused on the expectations related to planning aspects of OBW 2016 were included from the full survey and are summarized in Table 1. The remaining questions, not pertaining to planning of aspects of OBW 2016, were excluded. These included questions such as, “To what degree was the expectation of ‘Building relationship for future collaboration’ met?”

Table 1. Select Results from the Expectations Met Survey

Scale				
Much Less Than Expected	*Less Than Expected*	*Met Expectations*	*Exceeded Expectations*	*Greatly Exceeded Expectations*
1	2	3	4	5

To what degree were the following OBW expectations met?		
Expectation	*Mean*	*N*
1. Obtaining technical information regarding participation in a LVC event of this magnitude.	3.20	30
2. Access to technical support for network and enterprise issues.	3.22	27
3. Better understanding of technical challenges associated with integration of diverse systems.	3.60	30
4. Satisfactory resolution of technical problems during the integration and testing periods prior to OBW I/ITSEC events.	3.00	28
5. Adequate time and ability to test and rehearse before I/ITSEC.	2.63	30
6. Sufficient time to participate during vignettes.	3.14	28
7. Satisfactory resolution of logistics or administrative problems during preparation.	3.14	28
8. Satisfactory resolution of problems during OBW I/ITSEC events.	3.07	28
9. Learning about challenges/limitations of integrating Cyber technologies into traditional LVC training environments.	3.63	19
10. Demonstrating interoperability between systems.	3.37	30
11. Realization that regardless of challenges, solutions can be found.	3.43	28
12. Finding partners that have compatible technologies with your LVC system.	3.37	28

Overall, the responses indicate that expectations for OBW planning were met. The highest ratings received were associated with learning the challenges and limitations of cyber integration into LVC environments (mean = 3.63) and understanding technical challenges of integration (mean = 3.60). The lowest rating was for having adequate time and ability to test and rehearse before the event (mean = 2.63). Additional need for testing and rehearsal time was identified during the 2015 after-action review, and as a result more time was allocated for OBW 2016 test and rehearsal. Nonetheless, results show that more time was needed to prepare than was estimated. Further, the open-ended response questions within the Reactions Survey (described below) indicate that some industry partners did not participate in the integration week, which could have led to testing and rehearsal challenges. Starting integration even earlier, allowing more time for testing and rehearsal, and ensuring all industry partners attend the test and rehearsal could increase ratings on these questions for OBW 2017.

Reactions Survey

The Reactions Survey contained 8 questions and was partially, or wholly, completed by 29 respondents, of which approximately 90% were from industry and 10% were from government. Within industry, organizational roles were categorized by management (senior management or management) or technical (senior technical staff or technical staff). The majority of respondents overall were in a senior technical staff role (24%). And overwhelmingly, those in one of the two technical roles responded to more questions than those in both management roles (44.8% compared to 34.5%). The majority of the respondents (75.9%) had participated in OBW the previous year, which means most respondents had baseline OBW knowledge and experience that may contribute to more perceptive reaction ratings. Questions that were focused on planning, preparation time, and overall experience from OBW 2016 were included from the full survey and are summarized in Table 2. The remaining questions, not pertaining to planning of aspects of OBW 2016, were excluded. These included questions such as, “Would you recommend participating in OBW to other industry partners?”

The majority of the respondents (63%) felt they devoted more planning and preparation time than originally estimated. This is unsurprising, as more time is often needed when multiple partners engage in an event across multiple days. Providing this feedback to new and returning participants may stress the importance of starting these phases early.

Table 2. Select Results from the Reactions Survey

Did you find that the time devoted to your planning and preparation was more, less, or about the same as what you estimated going into the event?	Mean	*N*
More time needed than estimated	63%	17
Less time needed than estimated	4%	1
Same time needed as estimated	33%	9
Overall, how would you rate your 2016 OBW experience?	Mean	***N***
Above average	33%	9
Average	60%	16
Below average	7%	2

The majority of respondents (60%) rated OBW 2016 as average, and 33% rated it above average. The open-ended responses in the survey provide more insight into these results. Positive feedback on planning activities included: team-to-team collaboration and web meetings, communications, effectively run working groups, top-notch Navy network professionals, and helpful planning conferences. Areas identified for improvement included: tie down the database and enumerations at least a month in advance, use better visualization tools during the integration session to support verifying a model or that an action is taking place, have a single week for integration that everyone is required to attend and provide Wi-Fi to support troubleshooting, define simulation timeframes during each block earlier to permit better synchronization, and try to identify methods for simplifying the time and resource intensive planning and coordination effort.

STRATEGIC IPT VISION

In addition to the OBW Operational Integrated Product Team (IPT) responsible for planning the I/ITSEC show events, an OBW Strategic IPT provides the vision and guidance for the events and ensures alignment with overall I/ITSEC goals. One of the tasks of the Strategic IPT was to establish measurable goals for each of the focus areas. In the case of performance measurement, the group attempted to define a measurement framework that could facilitate annual, subject-matter agnostic comparisons of measurement activities at OBW. The Strategic IPT also hoped this framework would encourage OBW participants to treat the event as a measurement sandbox, encouraging them to design and validate new LVC performance measurement tools. To this end, the Strategic IPT defined three general categories of measures that LVC systems, in general, should include: Programmatic, Technological, and Learning (see Table 3). Further, the Strategic IPT recommended that LVC developers consider both Measures of Effectiveness (MOEs, i.e., process measures) and Measures of Performances (MOPs, i.e., outcome measures) in each of these categories.

After the completion of I/ITSEC 2016, OBW participants completed a survey that asked them to identify which of these performance assessments their systems did (or could) collect. That is, for each item listed in the Table 3, respondents indicated whether they (1) did collect data on this at OBW, (2) can/have previously collected this, (3) could, if extended, (4) could not have, or (5) N/A. Responses to the survey were intended to gauge the state-of-the-art in LVC performance assessment. Thirteen responses were collected.

Currently, these limited results reveal little. The trends suggest that outcome verification (46%) and dependability (30%) were the two areas more often assessed during OBW 2016—albeit by fewer than half of participating organizations that responded to the survey. An even smaller percentage collected data on outcome validation (23%), cost to administratively set-up a training event (23%), hardware and software acquisition costs (23%), and learning outcomes (level 2) (23%). Meanwhile, the other measurement areas were only addressed by one or two OBW participating organizations. However, many respondents indicated that they could have, or have previously, collected such data during LVC events. Interestingly, the measure that more respondents (30%) indicated they could *not* collect involved LVC-specific -ilities, such as the quality of “fair play” (i.e., the accuracy of time, environment, model synchronization). Consequently, the level of participation in measurement objectives and the integration of unique LVC technical measures could be enhanced during OBW 2017. The survey will be administered again after OBW 2017 and the annual comparison data should be made available.

Table 3. OBW Strategic IPT Measures

	CLASSIFICATION	EXAMPLES
Programmatic	**Outcome Verification** *Does the system do what we said it would do?*	• Effective coverage of the training capability compared to desired KSA outcomes (e.g., Fundamental Competency Sets (FCSs), Mission Essential Competencies (MECs)) • For other example, see the Verification, Validation and Accreditation (VV&A) document developed by the Defense Modeling & Simulation Coordination Office
	Outcome Validation *Does the system do what the customer originally wanted it to do?*	• For other example, see VV&A document developed by the Defense Modeling & Simulation Coordination Office
	Time Factors *How much time and associated person-hours are required to make the system work?*	• Efficiency to build content (time to build 1 hour of training) • Efficiency to execute (time to "turn on" one training event) • Learning usability (time to fully trained on "buttonology")
	Cost Factors *What is the total financial cost (including labor costs) required to acquire and operate the system?*	• Projected total cost of ownership • Hardware and software acquisition costs • Cost to develop 1 hour of training content • Cost to administratively set-up (hardware/software) training • Cost ("on the floor") to execute 1 hour of training • Return on investment (cost x readiness gains) • Cost avoidance, i.e., possible cost savings (fuel, bullets, infrastructure, etc.) versus traditional training forms (e.g., live)
Technological	**General Technical -ilities** *Does the technology (at many different levels) work according to general system quality principles?*	• System Quality Attributes list, e.g.: accessibility, dependability, recoverability, durability, standards compliance
	Specific Technical Capabilities *Does the technology (at many different levels) achieve specific LVC technology best practices?*	• Fair play: Accuracy of time/environment/model synch • Plug-and-Play: Did LVC protocols require modification to work? • Fidelity: Level of fidelity and similarity across federated systems • Instructional Strategies: Use of intentional pre/post/during-instructional strategies (not just exposure to the simulation) • After-Action Review Integration: Efficiency/effectiveness of integrated AAR
	Human-System Interface *Is the system "user friendly" to all end-users?*	• Usability (for trainees, instructors, administrators, curriculum designers, etc.) • Satisfaction • Engagement • Subjective experience
Learning	**Learning Outcomes 1** <u>Reactions</u>*: What were the stakeholders' (e.g., trainees, instructors) reactions to the event?*	• Reactions from trainees • Reactions from other stakeholders (admins, instructors, etc.)
	Learning Outcomes 2 <u>Learning</u>*: How much did the trainees learn from the event?*	• Pre/post knowledge/skill/attitude tests • Criterion- or norm-referenced tests
	Learning Outcomes 3 <u>Behavior</u>*: How did the trainees' actions "on the job" change as a result of the training?*	• Longitudinal follow-up • Change in immediate job performance pre/post-training • Potentially, leader-, self- or 360° subjective pre/post-job surveys
	Learning Outcomes 4 <u>Results</u>*: How did organizational outcomes (e.g., mission effectiveness, readiness) change as a result of the training?*	• Mission effectiveness change over time as a result of training • Readiness estimate change over time as a result of training • Secondary indicators, e.g., reduction in equipment breakage

Note: Learning concept from Kirkpatrick, Donald L., & Kirkpatrick, James D. (2006).

The Strategic IPT survey also included questions about whether OBW participants collected measures of the costs and benefits of the event, for their internal use. Table 4 shows these data. Cells with uniquely greater percentages have been highlighted; although, the low response rate means these results should be considered with caution.

Table 4. Organizational Utility Measures

Did your organization measure these outcomes of OBW 2016 - specifically?	Yes	No	N/A
Outcome Verification: Were quantitative metrics or targets established for your participation?	53.9%	46.1%	0.0%
Outcome Verification: Did the system fulfill the OBW operational assignment?	**84.6%**	7.7%	7.7%
Time Factors: Was an overall participation development and execution schedule established?	**92.3%**	7.7%	0.0%
Time Factors: Did you measure how much time it took for OBW network entry from boot-up?	38.5%	53.8%	7.7%
Time Factors: Did you measure how much time users / administrators took to learn your OBW system?	7.7%	**76.9%**	15.4%
Cost Factors: Did you set a budget for OBW participation?	**76.9%**	23.1%	0.0%
Cost Factors: Did you measure the total cost of OBW equipment involved (existing + new investment)?	**83.3%**	16.7%	0.0%
Cost Factors: Was a Return on Investment (ROI) strategy established?	30.8%	61.5%	7.7%

OBW 2017 AND BEYOND

OBW 2017 continues to explore challenges related to standards, AAR, cybersecurity, live integration, MLS/CDS, performance measurement, multiple terrain providers/formats, and centralized distribution of data. Additionally, as lead service for I/ITSEC 2017, the Army has developed additional objectives for OBW 2017:

- Gaining better insights and improving the human dimension/performance
- Integration of dense urban areas into terrain data, and improving database correlation and interoperability
- Emphasizing the importance of a uniform and centralized distribution of authoritative data
- Promoting LVC use of new and existing network technologies, including service/cloud-based capabilities
- Using LVC in domains beyond traditional Training (e.g., test, analysis, other)
- Using commercial gaming capabilities to address training needs

The lessons learned during OBW 2015 and 2016 are being incorporated into OBW 2017.

Cost Analysis

Due to the issues that hampered the cost analysis effort during OBW 2016, the objective for this measurement area during OBW 2017 is to obtain more consistent results in the cost analyses produced by Calytrix Technologies and PricewaterhouseCoopers through incorporating the planning strategies, discussed previously. While this may seem quite simple, it is a critical factor for industry and DoD acceptance of the results produced by the analyses. Once this goal is achieved, the analysis approach can be applied to additional LVC training events beyond OBW to generate the return on investment data needed by acquisition authorities.

Network Performance

The number of network participants will likely grow in the future, and addressing the valid (and invalid) data on each of VLANs is important for network performance. As additional information is placed on the network, and increased bandwidth is utilized to accommodate the additional systems and participants, there is a greater potential for issues that can negatively affect the outcome of an exercise. Staging exercises were used each year to ensure participants adhere to networking standards, yet networking anomalies have occurred during both OBW 2015 and 2016—sometimes just minutes before the start of an exercise. The OBW 2016 results show excellent progress in reducing unwanted or unexpected network traffic. Despite this, anomalies will likely occur during OBW 2017, and an increased use of firewalls may be needed to minimize unexpected and/or invalid packets and increase exercise reliability. The network performance measures and strategies for mitigating network performance issues documented across all three years of OBW should be shared with organizers of major training events, such as Northern Edge, that are incorporating increased LVC capabilities.

Trainee Performance

Many of the measures collected during OBW 2016 focused on the performance of the operators of airborne assets. These measures are still applicable to OBW 2017 and will be collected. Learner proficiency measures of tactical ground performance will also be introduced in 2017, including squad-level goals. Planning for the performance measurement activities is underway at this time, so the specific details of these measures are not available for this paper. We have also established additional performance measurement goals to support the Army's objectives for the human dimension/performance. First, we are defining measures that are applicable across the vignettes that make up each scenario block, which will characterize trainee performance improvements throughout the event. We are also considering block level or event level measures that truly showcase the advantages of LVC training.

Exercise Planning Metrics

The same Expectations Met and Reactions surveys will be administered after the OBW 2017 event. The results will be compared to the survey data from OBW 2016. Two additional surveys, still in development at the time of this writing, will be administered. The first focuses on partners that participated during all three years of OBW (2015, 2016, and 2017) to learn what differences were noticed, what improvements were made, and how lessons learned were applied. The second will seek to understand why several partners who participated in 2016 did not return in 2017. How well OBW emphasizes and addresses the criticality of uniform and centralized distribution of authoritative data (Army objective #3) will have been influenced by aspects of, and those participants contributing to, the planning process before and after I/ITSEC 2017. Survey questions regarding this critical piece are being developed, but at this writing have not been finalized. The ultimate goal of collecting this category of performance measures is to identify methods for making the exercise planning process more efficient, and ultimately do the same for LVC exercises in general.

SUMMARY

Performance measurement was one of many critical focus areas explored during the OBW events. This paper describes the lessons learned and challenges associated with measuring performance in LVC distributed simulation exercises, specifically the OBW 2015 and 2016 exercises. It is our goal that by broadly sharing these lessons learned with industry and the DoD, we can enable others to push the boundaries of the state-of-the-art without confronting the same issues and ultimately improve the effectiveness of LVC training.

ACKNOWLEDGEMENTS

We would like to thank Mr. Steven Slosser and Mr. Farid Mamaghani who contributed by meticulously analyzing and cross-checking survey data as well as providing valuable insight with regard to OBW planning and lessons learned. We also thank the members of the OBW Strategic IPT who defined the performance measurement MOEs and MOPs. We extend our sincerest gratitude to all of the OBW 2016 Performance Measurement participants (in alphabetical order): AFRL; Aptima, Inc.; Calytrix Technologies; NAWCTSD; PricewaterhouseCoopers, LLP; Rockwell Collins, Inc.; & Saab Defense and Security USA. **Disclaimer:** Presentation of this material does not constitute or imply its endorsement, recommendation, or favoring by the U.S. Navy or DoD. The opinions of the author expressed herein do not necessarily state or reflect those of the U.S. Navy or DoD.

REFERENCES

Kirkpatrick, D.L., & Kirkpatrick, J.D. (2006). *Evaluating Training Programs: The Four Levels (3rd Edition)*. Oakland, CA.

National Training and Simulation Association. (2016). *Operation Blended Warrior 2016 Playbook*. Arlington, VA.

Salas, E., Milham, L.M., & Bowers, C.A. (2003). Training Evaluation in the Military: Misconceptions, Opportunities, and Challenges. *Military Psychology, 15*(1), 3-16.

Salas, E., Rosen, M., Held, J., & Weissmuller, J. (2009). Performance measurement in simulation-based training: A review and best practices. *Simulation & Gaming, 40*, 328-376.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*, 321-326.