Employing Metrics to Balance Live, Virtual, and Constructive Simulation to Meet Training Objectives

Tim Cooley, PhD
DynamX Consulting
Oxford, MA
TCooley@DynamX.com

Ivar Oswalt, PhD, CMSP
The MIL Corporation
Lexington Park, MD
IOswalt@MilCorp.com

ABSTRACT

As the application of live, virtual, and constructive (LVC) simulation proliferates there is an increasing need to assess the effectiveness of each type of training delivery approach within a desired scenario. For instance, in an exercise, what training is best accomplished with constructive simulation versus virtual simulation versus live training, or a blend of two, or all three types together? To determine the best blend, a rigorous, repeatable examination of results, cost, risk, and time metrics within a structured framework is essential. This paper modernizes and extends foundational concepts developed previously while also narrowing their field of view to training, thus enabling the creation of a practical approach to LVC capability assessment. It then examines this approach relative to a notional but reasonable LVC training example, that includes current and emerging synthetic capabilities, such as the use of both real opposing forces in live training and those created using artificial intelligence and injected into the live training space in real time. This examination will consider training effectiveness (results) relative to other critical dimensions as well as the major trade-offs and relationships within a training evolution and/or set of training events that takes place over time. The goal is to provide a methodology that allows an efficient and effective blended and composable LVC solution to be defined that can be rapidly constructed to meet training needs. The discussion concludes with a glimpse into a future, where human-machine teams will need to be trained to work together within LVC simulation environments, to meet very challenging operational objectives.

ABOUT THE AUTHORS

Dr. Tim Cooley is the President and Founder of DynamX Consulting, a veteran owned consulting firm. Dr. Cooley spent 17 years on the U. S. Air Force Academy faculty, both in uniform and as a civilian, holding numerous positions including the DMSO Modeling and Simulation Chair, Deputy Department Head, and Senior Researcher. As the principal investigator, he completed many innovative mathematical and cost analyses for the USMC, USAF, and OSD and has performed counter-terrorism research for Joint Staff. He has significant experience using Machine Learning and Neural Networks for pattern recognition/classification in the medical community.

Dr. Ivar Oswalt is a senior modeling and simulation (M&S) analyst, who supports the Navy's M&S Office, the Office of Naval Research, and others in the DoD M&S Community as a member of the MIL Corporation. With significant experience in multi-attribute utility theory, he has developed M&S metrics and measurement approaches; analyzed overseas intelligence operations to assess the potential role of M&S-type predictive systems to forecast events; and currently focuses on M&S plans, policy, standards, outreach, and education. Dr. Oswalt contributed to this paper in his personal capacity and so the views expressed are his own and may not represent those of The MIL Corporation.

Employing Metrics to Balance Live, Virtual, and Constructive Simulation to Meet Training Objectives

Tim Cooley, PhD
DynamX Consulting
Oxford, MA
TCooley@DynamX.com

Ivar Oswalt, PhD, CMSP
The MIL Corporation
Lexington Park, MD
IOswalt@MilCorp.com

INTRODUCTION

It is well understood that training at all levels is essential to an efficient and effective workforce and even more crucial to a military or defense force. However, training, no matter the mode, is time-consuming and expensive. Therefore, innovative techniques and methods that decrease training time with the same result or increase the effectiveness of training at the same time are highly desired. With the proliferation of live, virtual, and constructive (LVC) simulation worlds and training suites, a plethora of training options are now available. This is a blessing and a curse. Amongst the multitude of options, how can one determine the best modes (live, virtual, and constructive) and methods for training such that the force is trained at the highest level in the minimum time? Essentially, this is the same as asking what amount, if any, of live, virtual, and constructive training for this task or set of tasks/scenario will train the trainees to the required level in minimum time? To solve this dilemma, a repeatable, effective method/framework that can be broadly applied is required. Modifying a proven framework previously developed (Aegis, 2008 and Oswalt, et al, 2011) we propose a method that is reliable and repeatable that provides an optimal or close to optimal mix of live, virtual, and constructive simulation-based training.

OPERATIONAL COMMAND DIRECTION

Your mission is to develop a scenario-based training plan. At your disposal is a fully fielded LVC environment, the complete list of Training and Readiness (T&R) manuals, and the tasks/scenario that needs to be trained. Everyone's desire is to have the training be efficient, realistic, and effective. That is easy. It only requires two battalions of infantry and artillery with associated logistics tail, ammunition, medical staff, etc.; three squadrons of different aircraft; a command and control aircraft in the sky; as well as communications aircraft and a cyber squadron. All of these are available in today's military. However, you also have budget constraints and the mandate to minimize the risk to the live assets. This requires moving some (or maybe all) of the training from a live exercise to a virtual or constructive world. However, how do you know the best mix of Live, Virtual, and Constructive? In the following sections we propose a framework to help guide the allocation of LVC capabilities towards meeting the goals of a training plan.

PROCESS / METHODOLOGY

In overall terms the key steps to conducting an assessment to enable a training provider to balance live, virtual, and constructive simulation capabilities to meet their training objectives are to 1. Describe the training system (deployed and envisioned LVC environment; T&R stated processes and exit criteria; tasks that need to be trained within the context of the scenario, and the trainees, trainers, and support personnel); 2. Develop the metrics to measure the items of interest - in this case, the results, cost, risk, and time characteristics of LVC simulation options; 3. Employ a robust algorithm to assess the options relative to the metrics; 4. Assess the results of the training.

Complete, Interlinked System Description - When determining the "best" mix of Live, Virtual, and Constructive simulation for a training scenario, the trainer should understand the detailed working of the LVC architecture to include the networking capability for each site not only back to the LVC home station, but between field units and range locations. This enables the planner to understand the lack of any capability such as a certain range doesn't have the bandwidth for each unit to have all their members immersed in AR/VR and to potentially train certain tasks differently.

Comprehensive, Distinct Metrics - With a rigorous and structured description of the LVC capability, it is now essential to associate complete and yet distinct metrics (or variables that allow the measurement of key component features). These metrics must allow the inclusion of quantitative and qualitative inputs, reflect the system's benefits/value and drawbacks/cost, allow for the inclusion of internal/integral factors as well as external ones. The metrics should address the areas of results, time, cost, and risk and clearly measure the different aspects of training Live, Virtual, and Constructive.

Robust, Accurate Adjudication Algorithm - Once a structure is established that describes the LVC simulation's capabilities and the metrics are assigned and evaluated (measured), the next key ingredient is a well formulated algorithm for utility calculation. This algorithm must allow for equal or varying weights / importance values to be assigned to each metric (or metric class). It must be able to calculate an expected value from manual (bottom-up) inputs or in an automated (top-down) fashion using a set of user defined rules or criteria (e.g., the ratio of Live, to Virtual, to Constructive). The adjudication algorithm must also be flexible enough to allow excursions and iterations to be run from a set of initial conditions. Also, as the algorithm calculates the maximum utility, best value, or similar measure-of-merit, it must do so in a way that maintains (or can account for) the confidence intervals assigned to the input data.

Critical Prerequisites - It is important to point out that to effectively employ LVC simulation capabilities, there are some key prerequisites. One of the first is the network connectivity needed to enable a federated or distributed LVC simulation event. It needs to include a variety of sub and specialized networks to enable training, operational, voice, telemetry, and similar data to be shared across a geographically dispersed set of participants with a variety of systems and capabilities and operating a many classification and release authority levels. There is also the associated requirement for a wide variety of interface standards to operational, training, hardware-in-the-loop, platforms and systems that are live, virtual, constructive, and in the future automated. Then there is the need for a simulation scenario driver that provides a consistent synthetic representation to all participants and that provides platforms, systems, and behaviors to constructive entities.

Next, are the ever-present and important set of procedures, participants, and physical controls needed to ensure the safety of the training event to the participants and to neutral or commercial entities. After that, are the training and readiness infrastructure needed to employ or deploy this LVC capability, track the training requirements that have been established and potentially achieved by the individual trainees but also at a variety of aggregated / group levels. Finally, and probably most importantly, are the trainers and software administrators needed to orchestrate the exercises, make the always-needed real-time adjustments required, coach the participants as necessary, and provide the after action review inputs that help to ensure the lessons learned are truly remembered. The more of these prerequisites that can be in place at the start of the training event, the more likely it is to be successful.

METRICS

In this assessment of LVC simulation training options there are four key data types: results, cost, risk, and time. The first, results, reflects the impact of the simulation's use. The term results was chosen because it is neutral, connotating neither a positive nor a negative impact (as opposed to terms like value, utility, or contribution, which normally indicate something favorable). The results of LVC use for training are reflected in the short- and long-term changes in the trainees' knowledge, skills, and abilities: both individually and collectively. Results could also include the degree to which an appropriate training venue exists and can be utilized (accessibility), is ready to support training (availability), can be adapted, changed, or reconfigured to meet multiple objectives, warfare and support areas, force affiliations and composition (flexibility), impacts the environment (terrain, plants, animals, etc.), and many others.

The second data type is cost, and can include the resources for the initial design, development, and deployment of the LVC simulation capability, as well as the resources (fiscal, personnel, infrastructure, etc.) required to plan, run, and review the outcome of an LVC event. Risk most often reflects any negative impacts of the using the simulation. In training these can be the possible harm to or death of participants, instructors, or the general public when, for instance, an aircraft on a training mission crashes in a residential area. However, risk can also reflect the potential down-side of standard programmatic variables like technical maturity. Time normally includes any changes in the elapse time of the training event and learning retention, but it can also include the time it takes to reach, configure, conduct, and review the training event.

With these four categories, example components, and important properties in mind, it is possible to further specify the metrics. They need to be decomposed into standards of measurement (i.e., variables) like seconds, hours, and days for time, or other labels based on the features of interest (e.g., fast-time or real time). These values are relative to a scale (a specified graduated reference used to measure) and may be nominal, ordinal, interval, or ration in type. The value type is particularly important if the intent is to combine or assess these values arithmetically. The values may range, with a defined start and ending point, normally zero when the capability under examination includes none of the feature being assessed, to the end point, which signifies having the maximum amount of the feature that is possible. With these aspects defined it is possible to measure (operationalize) the metrics and use the resulting values, combinations of values, or composed aggregate values to assess the LVC training solution option.

LVC OPTION ASSESSMENT

Our goal is to provide a methodology that allows an efficient and effective blended and composable LVC solution to be defined that can be rapidly constructed to meet training needs. So, the first step is to list the tasks that need to be trained and determine which ones have to be done live and which ones can't be accomplished via live training. Certain tasks may be required to be trained live simply because there is no simulator or simulation that is accredited to train that task. In that case, live training is required.

Similarly, there are tasks that are simulated or partially simulated due to the risky nature of training them live. For example, underwater egress from a helicopter that has crashed into the ocean would always be performed in a controlled simulator environment to ensure the safety of the trainee, not to mention the cost of crashing a helicopter. This brings the second reason to train using a simulator is that some tasks may just be too expensive to train in a live environment. For example, the prices of long-range cruise missiles are >$1M each, therefore, the number of live cruise missiles fired in training is extremely low. If the scenario called for offshore naval fire support, that task might be accomplished via a virtual simulator for the operator and perhaps a constructive simulation for the ground elements.

Once you have determined the tasks that must be trained live the remaining tasks will be trained using virtual or constructive methods. The dilemma now is deciding which method is best for each task. One may find multiple options for training a task in both the virtual and constructive realms and choosing the best method may be somewhat overwhelming. In order to accomplish this one must determine the criteria for "best". Many times, these criteria fit into four categories. Certainly, time is one factor. Performing training to the same performance level in a shorter amount of time is usually desirable. Performance level (also known as reward or result) is another metric. Training a task to a higher level of performance is typically desirable and a method where that performance level transfers to real world conditions is critical. Cost may be another factor in the decision. Constructive simulations will typically be less expensive than virtual training. Finally, risk must be considered - risk to completing training, risk to the safety of the trainees, risk of low training transfer to real world, etc.

Once these two sets of training alternatives are addressed, the focus turns to those training requirements that could benefit from a blended LVC solution. Assessing the characteristics of these solutions generically is challenging at best, but we believe doing so provides some useful insights in spite of the generalizations that must be made. In order to decompose this analysis into a set of manageable set of individual cases, LVC options have been divided into six alternatives: those that emphasize Live, Virtual, or Constructive individually, those that emphasize the three pairs of LVC, and those that emphasize all three.

To help to understand the differences between these options, two results, cost, risk, and time metrics have been defined. They are:

- Employment Confidence - The degree to which the trainee feels able to perform the warfighting action required, with all of its essential components, characteristics, and capabilities.
- Learning Scope - The variety in the scenario relative to unit types, scenario objectives, rules of engagement, warfare and support area representation, etc.
- Personnel Safety - The likelihood that the trainee will complete the training without injuring themselves or others.
- Training Environment Replication - The degree to which the simulation accurately imitates the actual training environment and does not provide negative training.

- Efficiency - The minimum amount of time that a training event takes to provide adequate instruction to participating personnel (from event start to event end - not include planning or after-action analyses).
- Capability Availability - The amount of time, relative to 24/7 365, that the LVC environment is available for training.
- Operations - The cost, per unit of time, to operate a system. The system can be a live asset, simulator, or constructive entity; or their combination.
- Infrastructure - The cost of the networks, system authorizations to operate, required standards, training personnel, and similar.

Using these eight metric types and components, it is possible to examine the seven LVC simulation alternatives. After a general observation is provided, the eight options are ranked, with green indicating the top/best assessment (see Table 1).

Table 1 - LVC Metrics-Based Parametric Scores

Type	Component	LVC Option						
		L	V	C	VC	LC	LV	LVC
Results	Employment Confidence	Highest when a real weapon launches from a real system based on operator input						
		Lowest when operational realities are abstracted						
	Learning Scope	Highest when scenario complexity is maximized						
		Lowest for virtual part task trainers (PTTs)						
Risk	Personnel Safety	Highest in constructive and virtual environments						
		Lowest in live simulation environments						
	Training Replication	Highest when using operational assets with accurate stimulations						
		Lowest when using COTS interfaces and approximate algorithms						
Time	Efficiency	Depends on training complexity. When highest, blended LVC is best						
		When lowest, minimal viable product / PTTs are best						
	Capability Availability	Highest for non-operational assets - C (and then V)						
		Lowest for operational assets when training interferes with operations						
Cost Minimization	Operations	Highest when operational units are simulated, or training is concurrent						
		Lowest when all units are live / operational assets						
	Infrastructure	Highest when existing LAN/WAN capabilities can be employed						
		Lowest when new, dedicated capabilities need to be developed						

Given these assessments, simulations that blend virtual and constructive (VC) are very attractive. While they score low on employment confidence, this middle column has a reasonable degree of most of the assessment metrics. Purely live simulations (L), as is well known, are expensive, normally focused relative to scope, take significant time to provide a robust training capability, and can be risky to participants.

Another thing that Table 1 shows, is that the "best" method is rarely immediately discernable because the measures many times compete with each other. For example, the technique with the highest performance and the best training transfer may also take the longest and, therefore, cost the most. The method with the best virtual performance may not have the best training transfer but may be significantly quicker and less expensive. The combinations are typically many and a repeatable, orderly process to measure and evaluate each option is required. To do this metrics as described above are required.

The framework envisioned to evaluate the "best" LVC mix is one very similar to the method we developed to determine the Return on Investment (ROI) of DoD Modeling and Simulation (Oswalt, et al, 2011). In that effort we described a Multi-attribute Decision Making Method (MADM) which uses measured metrics and weights for each metric to calculate an ROI score for an investment. The measured metrics are multiplied by the weights and then the products are summed to determine a final score.

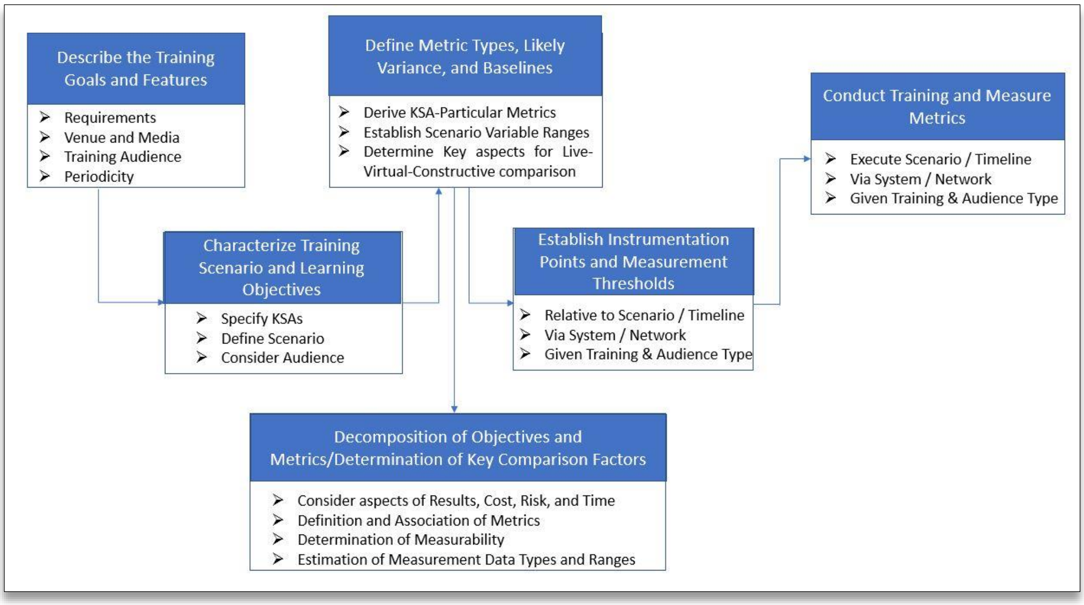

Figure 1 - LVC Simulation Assessment Process

This allows the comparison of several investment opportunities to decide which one provides the greatest bang for the buck. The idea of developing the "best" training plan has many similarities to calculating ROI of an investment. First, there are multiple factors to consider, and they may not all carry the same importance. Second, both situations are desiring to choose the "best" from among several valid options. Third, as in the case of ROI in the DoD the result you are measuring is not monetary but is a quantitative measure of a qualitative state such as unit readiness. In the case of determining the best mix of L, V, and C to train a unit on a scenario the MADM method repeatably and systematically combines factors and calculates a score which can be used to determine the "best" L, V, and C mix.

Once the tasks for the scenario are identified, that list screened for as noted above, some tasks will be required to be trained in a virtual simulator and some will be required to be performed live. For the rest, the first step is to establish which factors might be more important than others. Is it most important that the time is short, or is it more important that the cost is low? How does training transfer rank in importance; or is everything equal? Once the applicable metrics have been chosen, the weights determined, and the metrics weighted, they can be fed into the MADM process and a score calculated for each training option for that task.[1] At this point, the training method for that task with the best score can be selected and the process repeated for the next task until the best training technique for each task have been identified. Now is the time to insert the tasks into the scenario in their correct order and a final sanity check would be required. This final check screens for logistical and sequencing issues such as one virtual task trained amid several live tasks (move the virtual task to live), or a task that would be trained best with a live-virtual mix, yet the link between the simulator and the range where the live task would have to be trained is non-existent. Once the final check is complete the training plan can be fielded, and units can train with this plan using an LVC environment in an efficient and effective manner.

[1] It may certainly be the case that the MADM process is not required to determine the best training method for a given task. It may be that one method is an order of magnitude greater in cost so it wouldn't be considered, or that the time to train the task with a certain method won't fit in the schedule. Obviously, those situations do not require a rigorous process.

LVC TRAINING EXAMPLE

LVC simulations often offer many advantages over other possible training alternatives.[2] Compared to employing operational assets on a tactical training range, LVC events are normally less costly, scenarios can be repeated as needed, and novel and yet potentially dangerous tactics practiced. In addition, LVC exercises can include unusual manned and unmanned concepts of operations using sensors and warfighting plans under development in a way that limits emissions and unwelcome reconnaissance. In this notional but typical LVC training example, some of the nuances in operationalizing these measurement constructs are discussed.

LVC events, by design, include interactions between multiple participants and systems striving to achieve a common mission objective. For that reason, LVC simulation-based training exercises normally include as one of their goals teaching participants effective command, control, and communications (C3) processes and techniques. Such instruction, using LVC simulations was found to "enable enhanced training" in cases that included artillery C3 (Hannay, 2014). As a result, to assess how results, cost, risk, and time metrics could be used to assess the effectiveness of an LVC training event, the use case examined here is instructing effective C3 techniques to the three operators located within the Engagement Control Station (ECS) of a Patriot air defense missile battery (see Figure 2).

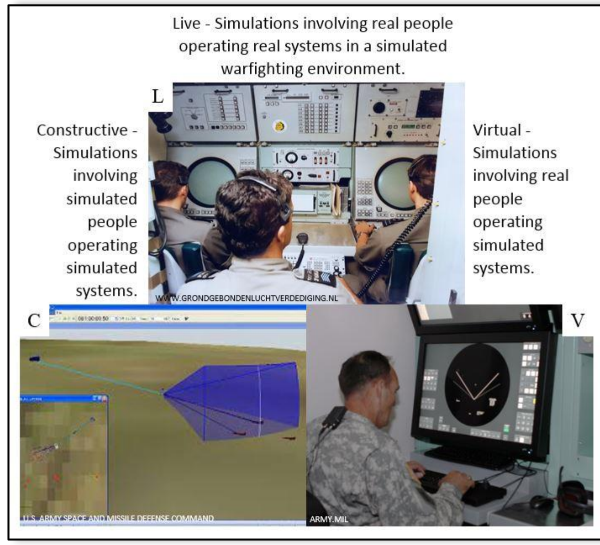

Figure 2 - Patriot LVC Training Operations

Patriot is a guided missile air-defense system with long-range, medium-to-high altitude, all-weather capabilities designed to counter tactical ballistic missiles, cruise missiles, and advanced aircraft. The ECS is the command center of the missile battery and contains stations for three operators seated at two radar consoles and a communications station (see Figure 2). Operators can see the status of all targets that the system is tracking and can select or deselect targets or set the system to run in fully automatic mode. The communication station allows the battery to communicate with other batteries and with the command center for the region. As a result, the operational mission and the training objectives of this use case are to:

- Control the air battle: monitor threats, prioritize targets, and engage targets.
- Communicate with friendly forces: command and coordinate, via messages, with higher headquarters, subordinate, and lateral units.
- Conduct liaison with supported units and other units in the Patriot battery's operating area.

Assessing the effectiveness of these C3 activities can draw from the long history of metrics development and assessment, much of which applies to distributed LVC training simulations. For instance, previous analysis has developed measures of C2 performance and measures of effectiveness relevant to gunnery and air defense scenarios (Sweet, 1985). This analysis included the ability of the operator to understand the data being displayed and then use it to achieve a successful engagement of the target. It also included decision response time (both elapse and relative (timeliness)) and associated system engagement impact variables and assessment measures.

There has also been previous research into how to assess the effectiveness of C3 within LVC-based training (Roberts, 2017). In it, four metrics were proposed for evaluating C2 LVC training effectiveness: observe-orient-decide-act (OODA) time, communications effectiveness, area of responsibility (AOR) familiarity, and communications familiarity. OODA time is the elapse between the estimated earliest entity detection time and when it is classified as suspect, hostile or friendly. Communication effectiveness evaluates an individual's proficiency with communication

[2] This section has been adopted from Operationalizing Artificial Intelligence in Simulation Based Training, Inter-service/Industry Training, Simulation and Education Conference, December 2021 with the permission of the authors (Cooley, 2021).

protocols and brevity codes within the event. AOR and communications familiarity are subjective assessments made by the trainee of how much they learned or increased their knowledge of the region of interest and the required communications procedures, respectively, because of the training event.

Considering the particulars of this use case, four metrics have been chosen as examples of what could be measured to assess the value of a C3 LVC training event. They are listed in Table 2 and include key indicators of C3 proficiency relevant to the Patriot system as well as results, risk, time, and cost variables that reflect the overall value of an LVC option.

These metric types, provide the foundation to estimating baselines, collecting telemetry, determining measurement thresholds, and ultimately to operationalizing them to enable LVC event assessment.

Table 2 - Patriot C^3 Example LVC Metrics

Metric	Type	Range
Engagement Outcome	Result	0-R Threats penetrated. 0-R Threats neutralized.
Personnel Safety	Risk	# Safety Violations / Number of Scenarios Run.
Elapse Time	Time	0-T from training start to achieving effectiveness level X
Training Event Resources	Cost	Amount in $s of hardware set-up, missile expenditure, personnel travel, and training time.

When measuring these metrics for an LVC federation, of course their absolute values will vary significantly by simulation type, but it is useful to rank order them (see Table 3, Note 3 > 1). The rationale behind these rankings can be summarized. In terms of engagement outcome, the level of preparation normally associated with a live shot makes the likelihood of success high, and similarly, although a bit less so for LVC. A purely virtual environment, with less preparation and pressure rank it at the bottom (although the flexibility of the synthetic environment is likely to allow for a wider range of scenarios to be trained against). It is important to point out, that when a comprehensive set of results metrics are to be assessed, the ability to do so will depend on the simulations' interface design specification, which may or may not include the ability to capture and provide the data needed to assess training performance, effectiveness, or learning retention (Beaubien, 2017). Personnel safety is the highest when live assets are not involved, in the middle in the case of LVC because the shot itself is constructive but the systems are in an operational environment, and lowest when real weapons are launched.

Regarding elapse time, or the time it takes to reach a particular proficiency level, the rapidity of the high number of reps and sets afforded in a virtual environment make it score the best. While LVC has the advantage of offering training opportunities within the area of operations, and thus improve the proficiency of deployed units within the closest proximity to fight as possible, in-theater demands will limit the number of training opportunities. Lastly, the time it would take to safely set up, run, and evaluate the outcome of a live shot on a range iteratively, to achieve a particular training level, results in it ranking last.

Table 3 - Ranking LVC Options

Metric	L	VC	LVC
Engagement Outcome	3	1	2
Personnel Safety	1	3	2
Elapse Time	1	3	2
Event Operations Cost	1	3	2
Most Rank Order Outcome	**1**	**3**	**2**

When examining the last assessment metric type, cost, and event operations cost in this case, the highest in this example will be live simulation. This is particularly true because the cost of establishing the LVC capability initially, which is $100s of millions of dollars, is not accounted for in this cost variable. Using this metric alone, the cost score is driven by the expenditure of the missile, which in the case of Patriot, is approximately $3M dollars each. LVC comes in second, given the costs of including an underway or deployed asset, but excluding the missile's cost, since it would be constructively represented. Finally, the highest score / least costly option is a blended solution using virtual and constructive simulation systems.

LVC consistently scores in the middle of the other two options, indicating that it provides a robust ("on average") training capability. As more specifics are added in terms of specifying the LVC capability being provided and the training objectives being sought, the relative value of this option will be better understood and is likely to start to outrank the others. For instance, in fighter pilot training, "Results indicate that pilots are positive toward the LVC scenario design, especially the dynamics that a large-scale scenario brings to training of decision making." (Aronsson et al., 2022).

While this discussion provides some insights into how to apply this approach to a notional LVC training example, there are many additional considerations that need to be contemplated. One is to account for training transfer between different LVC simulation modalities. For example, from constructive to virtual, and probably even more significantly between virtual and live. Next, it will be necessary to be creative in gathering the data needed to assess the metrics chosen. Conducting training via LVC simulation-based events is still relatively new and the data collection requirements (likely in the form of Data Information Descriptions) are still being formulated.

Next, this method requires the development of measurable metrics, some of which may contain subjectivity or subjective assessments from experienced operators and trainers. Thus, applying techniques to quantify subjective assessments are likely to be relevant as will the need to enable operators and trainers to provide LVC exercise feedback using these techniques. All-in-all, there are several real challenges to overcome, but the bottom line is that as LVC simulations are increasingly used within persistent, or persistently available networks, the need to optimize the application of each, and every combination of their capabilities, increases significantly.

LIVE, VIRTUAL, CONSTRUCTIVE AUTONOMOUS / ARTIFICAL INTELLIGENCE (AI)

While the idea of integrating AI into training is not necessarily new (Oswalt & Cooley, 2019), recent advancements in this area mean the idea is now becoming closer to reality. Recent demonstrations show the capability of artificially intelligent pilots flying live aircraft (unmanned aerial systems) on an operationally relevant track (ASDNews, 2023). Real-time updates were given to the aircraft and the AI pilot dynamically adjusted to incorporate the revised flight plan and threat data (see Figure 3). Additionally, these missions were flown in conjunction with a "digital twin" aircraft using a LVC system to conduct a multi-objective collaborative combat training mission.

Figure 3 - Autonomous Aircraft

The advent of AI in live operational systems produces another dimension to LVC training and training in general and that is the man machine interface of AI controlled assets and those operated by humans only. How will this interface and interoperability evolve? For example, will autonomous flight agents operate as a cohesive four-ship or two-ship or will they integrate where there might be half autonomous agents and half live pilots? Could an autonomous pilot be trained for air-to-air combat, or will they be limited to reconnaissance or supply missions? Is it ethical or even legal for an AI piloted aircraft to deliver weapons? Would the autonomous agent perform all the mission except weapons delivery which would be accomplished virtually by a human in the system. Are there methods to use autonomous agents in training that would be beneficial even if they are not used in the live fight?

All these questions will need to be carefully considered and the answers well understood and tested. However, at some point in the future it is easy to envision that there could be another category in addition to L, V, and C; that is L-A, live-autonomous which captures training with both live and AI operated autonomous assets. This is the topic for future work and discussion.

CONCLUSIONS AND RECOMMENDATIONS

The use of simulation by military organizations around the world is greater now than ever before. It supports strike operations planning and rehearsal, allows otherwise impossible training, increases the rigor of campaign assessments, and provides an experimental environment for platform design, development, and testing; to name just a few. Increasingly, however, LVC's technology infrastructure is enabling persistently available capabilities across a wide range of mission applications, especially in support of training. As a result, balancing when, and to what degree, LVC simulations should be blended to best support particular objectives is becoming increasingly critical. Said differently, the LVC training community is moving from "there is no other way" to "how do we optimize this system to yield the best results?" Efficiently applying, blending, and optimizing requires metrics.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the research and editorial support provided by Ms. Margaret Callahan, Senior Modeling and Simulation Researcher at The MIL Corporation.

REFERENCES

AEgis Technologies Group. (2008). *Metrics for modeling and simulation (M&S) investments.* Naval Air Systems Command Prime Contract No. N61339-05-C-0088.

ASDNews. (2023). *GA-ASI Flies Multiple Missions Using Artificially Intelligent Pilots*. Retrieved from www.asdnews.com/news/defense/2023/01/12/gaasi-flies-multiple-missions-using-artificially-intelligent-pilots.

Beaubien, J. M., Stacy, E. W., Wiggins, S. L., & Lucia, L.C. (2016). How Can We Measure Learning? Let's Count the Ways! *Proceedings of the 2016 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.

Cooley, T., & Oswalt, I. (2021). Operationalizing Artificial Intelligence in Simulation Based Training. *Proceedings of the 2021 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*

Hannay, J. E., Mevassvik, O. M., Skjeltorp, A., & Brathen, K. (2014). Live, Virtual, Constructive (LVC) simulation for land operations training: Concept Development & Experimentation (CD&E*). Proceedings of the NATO Modelling and Simulation Group Symposium Integrating Modelling & Simulation in the Defence Acquisition Lifecycle and Military Training Curriculum (MSG-126).*

Oswalt, I., & Cooley, T. (2019). Simulation-Based Training's Incorporation of Machine Learning, *Proceedings of the 2019 MODSIM World Conference*.

Oswalt, I., Cooley, T., et al. (2011). Calculating Return on Investment for U.S. Department of Defense Modeling and Simulation. *Acquisition Research Journal*.

Aronsson, S., Artman, H., Mitchell, M., Ramberg, R. & Woltjer, R. (2022). A live mindset in Live Virtual Constructive simulations: a spin-up for future LVC air combat training. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology.* https://doi.org/10.1177/15485129221106204

Sweet, R., Metersky, M., & Sovereign, M. (1985). *Command and Control Evaluation Workshop*. Military Operations Research Society.