

Validating Simulators for Live, Virtual, Constructive Exercises: A Methodology

**Robb Dunne, Ph.D., Scott Harris,
Lauren Reinerman, Ph.D.**
Institute for Simulation and Training
University of Central Florida
Orlando, FL
rdunn@ist.ucf.edu, sharris@ist.ucf.edu,
lreiner@ist.ucf.edu

Nathan Jones
Program Manager Training
Systems (PM TRASYS)
Orlando, FL
nathan.jones1@usmc.mil

Alexander Arrieta
Marine Corps Synthetic
Training Environment,
Training and Education
Command
Quantico, VA
alexander.arrieta@usmc.mil

ABSTRACT

The Commandant's Planning Guidance for 2015 states, "My intent is for Marines to encounter their initial ethical and tactical dilemmas in a simulated battlefield vice actual combat" (USMC, 2015). Live training can never be completely replaced, but live training can be optimized or augmented, or both, by preparing in live, virtual, constructive (LVC) training events. However, enabling readiness through training depends on the capabilities of the simulators to enable training of operational tasks as required by training objectives. Typically, however, validation of simulator capabilities is conducted while in stand-alone configuration. Validating simulator capabilities while linked to other simulators in a distributed environment is only now being undertaken. This paper reports on an assessment and analysis methodology based on the systematic team assessment of readiness training (START) process to determine if the method utilized can inform potential validation efforts for future LVC distributed mission training environment (DMTE) exercises. The capabilities of three simulation systems, used in a LVC DMTE were first assessed in stand-alone configuration and then, after conduct of the DMTE exercise, assessed to determine the simulators' capabilities while in distributed configuration. Then, the stand-alone and distributed configuration START results were compared and analyzed. While general findings of this effort are outside the focus of this paper, the salient take-away is the development of a methodology for identifying simulator capabilities and value-added training delivered in distributed, LVC configurations. A job-aid for validation of distributed systems in the form of a protocol, based on analysis, observations and subject matter expert input was developed and herein provided. This protocol will be employed during a follow-on effort. Accordingly, adjustments and revisions will be made and reported. Implications, downstream application, and potential impacts of this methodology upon validation and verification efforts, LVC simulator requirements, and recommended adjustments to the START process are also discussed.

ABOUT THE AUTHORS

Robb Dunne, Ph.D. received his degree in Instructional Technology from University of Central Florida and is now a Research Associate at the Institute for Simulation and Training. He has conducted numerous training system evaluations including front-end analyses, verification, and validations, training effectiveness evaluations, and systematic team assessment of readiness training processes for Marine Corps Systems Command, United States Marine Corps Training and Education Command and Naval Air Warfare Center Training Systems Division.

Scott Harris is a Faculty Research Associate at the Institute for Simulation and Training in the Prodigy Lab of the University of Central Florida in Orlando, FL. In February of 2017, Scott retired after serving on Active Duty for 28 years as a Pilot in the United States Marine Corps. In 2010, Scott was assigned by Headquarters Marine Corps to serve as the Program Manager for all Marine Corps Aviation Training Systems at PMA-205. Prior to his assignment as the PM for USMC Aviation Training Systems, Scott served on the staff of the Chairman of the Joint Chiefs in J8 as the Vertical Lift Strategic Funding and Requirements Analyst for General Glen Walters (currently the Assistant Commandant of the Marine Corps).

Lauren Reinerman, Ph.D. is the Director of Prodigy, which is one lab at the University of Central Florida's Institute for Simulation and Training, focusing on assessment for explaining, predicting, and improving human performance and systems.

Nathan Jones is the Manpower, Personnel, and Training Technical Support Lead at MCSC PM TRASYS. Mr. Jones has 18+ years of experience in human performance, human-systems integration, assessments, and acquisition program experience. He is responsible for overseeing PM TRASYS' training front-end analyses, verification and validations, training effectiveness evaluations, and training domain expertise for PM TRASYS. He has been directly involved in the development of acquisition documents for the past 10 years at PM TRASYS.

Alexander Arrieta is the Verification, Validation, and Accreditation (VV&A) Lead for USMC TECOM, Training, and Education Capabilities Division. Alex is responsible for organizing, coordinating, and executing a comprehensive VV&A program for Marine Corps training Modeling & Simulations systems to provide credibility of USMC training solutions and, as the Accrediting Agent, define accrediting criteria. Alex is a retired Marine, a National Service Rifle Champion with over 25 years of national and international marksmanship competition and training; developing applying and instructing procedure-based training incorporating visualization and mental management.

Validating Simulators for Live, Virtual, Constructive Exercises: A Methodology

**Robb Dunne, Ph.D., Scott Harris,
Lauren Reinerman, Ph.D.**
Institute for Simulation and Training
University of Central Florida
Orlando, FL
rdunn@ist.ucf.edu, sharris@ist.ucf.edu,
lreinerm@ist.ucf.edu

Nathan Jones
Program Manager Training
Systems (PM TRASYS)
Orlando, FL
nathan.jones1@usmc.mil

Alexander Arrieta
Marine Corps Synthetic
Training Environment,
Training and Education
Command
Quantico, VA
alexander.arrieta@usmc.mil

INTRODUCTION

The Commandant's Planning Guidance for 2015 states that he expects the Corps "...to make extensive use of simulators" and that it is his intent, "...for Marines to encounter their initial ethical and tactical dilemmas in a simulated battlefield vice actual combat" (p.11). In answer, the Training and Education Command (TECOM) is identifying and pursuing linkages among the Corps' existing simulators and examining networking requirements to provide Joint and partner nation training thus expanding the employment of simulators. The Marine Corps Synthetic Training Environment (MCSTE) focuses on live, virtual, constructive (LVC) networks to support distributed mission training environments (DMTE) between Marine Air-Ground Task Force (MAGTF) elements that are necessary to enable large-scale exercises such as Talon Exercise (TALONEX), tactical MAGTF integration courses, integrated training exercises and other formal service level training exercises in accordance with the Commandant's guidance.

With the expansion of the MCSTE and increasing investment in technology to provide tactically relevant and networked simulators, verification, validation, and accreditation (VV&A) efforts must extend beyond stand-alone configurations to evaluate simulator capabilities in networked environments. Simulation credibility for use in a particular application or environment is measured by VV&A efforts (Jones & Arrieta, 2017) and this credibility must extend to LVC. To support LVC training that enables operational readiness, replicable, systematic evaluation processes are essential to ensure cost-effective acquisition of new Modeling and Simulation (M&S) systems and to provide research-based evidence as foundation for acquisition requirements of new systems, and upgrades of existing systems. As the MCSTE matures to an envisioned "plug and play" system of systems (SoS) capability, evaluations of MCSTE simulator capabilities to provide value-added training¹ that goes beyond individual or Military Occupational Specialty (MOS)-centric training, are necessary. However, efforts to validate simulator capabilities while in a distributed environment are only now being undertaken. This paper reports on an assessment and analysis methodology using the systematic team assessment of readiness training (START) process to determine if the method utilized can inform validation efforts for LVC distributed mission training environment (DMTE) exercises.

The three systems used in this LVC DMTE exercise are:

- Virtual Battlespace 3 (VBS3): a commercial off-the-shelf software program embedded in the Deployable Virtual Training Environment (DVTE) laptop system used to train Marines from the individual to the battalion staff by using a self-contained network of computers with reconfigurable workstations capable of presenting skills sustainment training scenarios. VBS3 provides a Joint Terminal Attack Controllers (JTAC) station.

¹ Value-added Training is a concept used to guide the identification of potential LVC configurations focusing on the progression of a Marine's competency (knowledge, skills and attitudes). For instance, there is value-added for an aviator and a Joint Terminal Attack Controller (JTAC) to train close air support (CAS) objectives together; both trainees accountable for mission objectives.

- Combat Convoy Simulator (CCS): a 360-degree immersive training environment that places Marines in realistic, convoy operations that allow for repetition and remediation to train basic and advanced combat convoy tactics, techniques, and procedures, command and control, improvised explosive device countermeasure skills, attack, and response with small arms and crew served weapons.
- F-35 Pilot Training Aid (F-35 PTA): a touchscreen monitor that displays out-the-window views and touchscreen instrumentation. The F-35 PTA fully replicates the F-35B hands-on throttle and stick. It is Distributed Interactive Simulation (DIS) network compatible and contains communication radios, constructive simulation, and virtual reality capability, with an expandable library of 3D entities that include munitions effects and aural cues, simulation of visual conditions such as night and infrared and customized entities and assets that can send datalink messages to the F-35 PTA.

While these three systems comprise the “V” of LVC, the *live* (real people operating real systems) and *constructive* (simulated people operating simulated systems) aspects, as defined by Department of Defense (2010), are satisfied by the participation of Joint Terminal Attack Controllers (JTAC) using operational communications to direct pilot participants in the F-35 PTA. Operators of the CCS and VBS3 systems make inputs, but are not involved in determining the outcomes. Outcomes are a function of the interaction between the simulated entities and systems, and the exercise participants.

Once these systems were solidified as the evaluated systems, an assessment plan that includes appropriate analyses, an integral part of the instructional systems design process, was developed. At the heart of the assessment plan is the systematic team assessment of readiness training (START). START contributes to effective front-end analysis (FEA) by capturing physical and functional baseline requirements. As an augmentation or enhancing dimension to V&V efforts, START provides specific data to support accreditation and, when augmenting training effectiveness evaluations (TEE), START enables a replicable analysis that documents the training affordances of training devices such as part-task trainers, desktop computer based training, and high fidelity simulators to live events.

Possessing characteristics of each of the FEA, V&V, and TEE processes, the START process was used as the evaluation instrument for this LVC DMTE. The START process employs a series of linked spreadsheets to determine the level of capability of training simulators to train tasks (Sheehan et al., 2009). START results also inform decision makers of the suitability of upgrades to improve training by expanding the array of mission essential tasks (METs) that can be effectively supported (Johnston et al, 2015). Initially intended for use on stand-alone systems, the dynamic training environment inherent to LVC DMTE required that START extend the verification and validation criteria (i.e., device capability, training criticality) from stand-alone systems to LVC configurations (Dunne, et al., 2017).

To construct the assessment plan, it was necessary to develop an innovative methodology for the evaluation of simulators in a distributed environment. The LVC DMTE combined battle staff, individual, unit and collective training objectives across MOS and units to simulate the contemporary “conditions, circumstances, and influences that affect the employment of military forces and bear on the decisions of the unit commander” (DoD, 2016). This condition required extending analysis practices beyond stand-alone configurations. This paper describes that methodology, the lessons learned, and recommendations moving forward to other LVC DMTE efforts such as the Marine Air-Ground Task Force-Tactical Integrated Training Environment (MAGTF-TITE) during TALONEX 2-18 in Mar 2018. There is additional potential to apply the methodology in support of Marine Corps Tactics and Operations Group Tactical MAGTF Integration Course during May/June 2018.

METHODOLOGY

To establish suitability of the three training systems for this particular DMTE, it was necessary to conduct stand-alone STARTs to determine system capability baselines. Stand-alone STARTs identified tasks associated to training and readiness (T&R) events that were fully or adequately supported, may require work-arounds to execute, or were not supported at all. System attributes, characteristics that provide the necessary fidelity for conduct of tasks and that affect the training effectiveness of the system, were documented to provide awareness during the construction of the distributed training exercises.

The assessment design required to capture the data and allow efficient observation of all systems and participants is illustrated in the figure below.

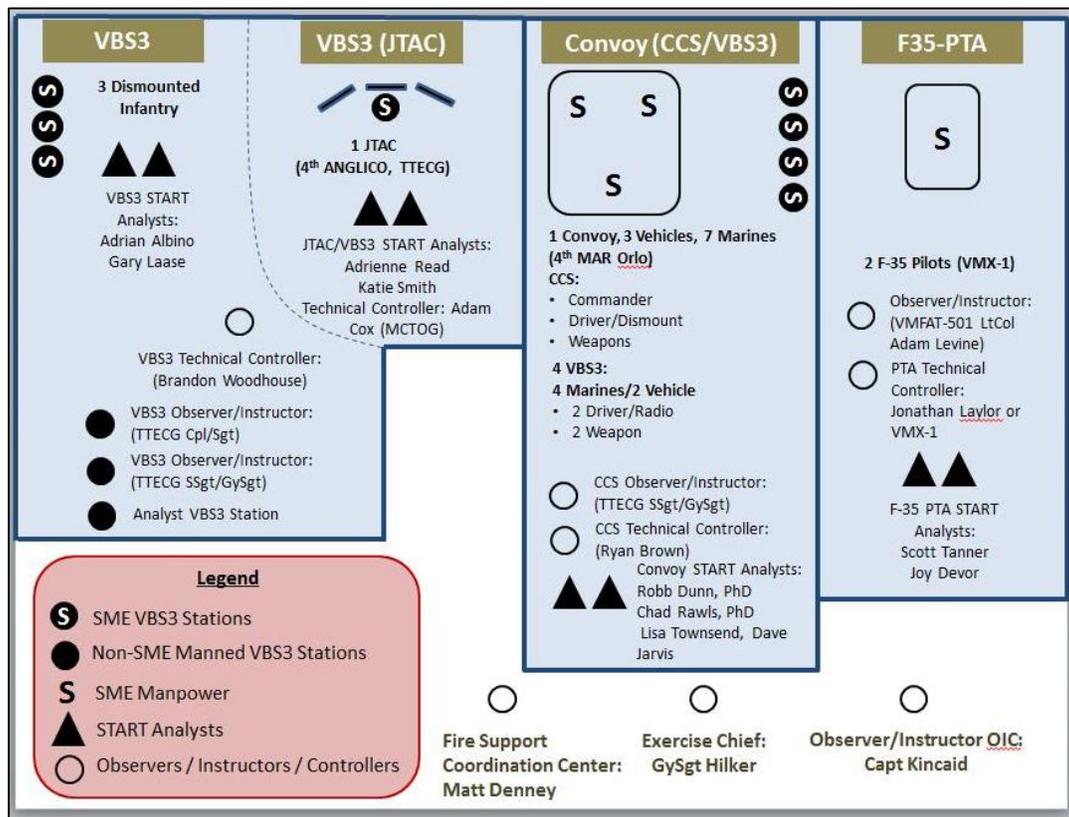


Figure 1. DMTE Assessment Design

START Instrument Inputs and Outputs

START is a systematic, replicable analysis process that assesses training device capabilities to support performance of tasks associated to T&R events and thus training objectives. It is a data-driven evaluation of simulators and their ability to enable and support training of tasks associated to T&R events and identifying areas for improvement with potential upgrades to enhance trainee proficiency and optimize return on investment (Johnston, et al., 2015). Executed in several steps, the START process collects data through SME feedback. This data enables computational determination of task criticality and device capability scores.

START’s algorithms combine task-to-attribute criticality, and attribute-to-capability ratings provided by SMEs, during data collection workshops, to illustrate the level of support the training device provides for tasks associated to T&R events. SMEs evaluate how critical the presence of specific attributes (conditions and stimuli) is to the execution of each validated task. Then SMEs evaluate how capable the training device is at providing those specific attributes, task by task. In this way, analysis outputs illustrate the alignment of criticality levels to levels of capability.

Input: Tasks

Tasks relevant to learning objectives are selected and validated by SMEs and are the basis of the evaluation of simulator capabilities. For credible and actionable impact, these tasks align to doctrine such as T&R manuals.

Input: Attributes

Attributes are the characteristics, conditions, and stimuli required from the simulator in order to enable conduct of tasks. Together, the tasks and attributes are the *x* and *y*-axes that bound the START matrix. Attributes are tailored to address specific system capabilities or requirements. It is important to determine which attributes are necessary for which simulator to avoid artificial skewing of results. That is, if it is not important for the simulator to convey accurate and realistic representations of realistic distances or changing distances of objects in the environment (on the ground or in the air) then the depth perception attribute should not be included in the START. However, this

type of elimination should only occur *holistically* -when it can be said for the complete portfolio of tasks the simulator is intended to address, not when only a particular scenario or set of exercises form the evaluation framework as is found in a limited scope.

The table below shows two examples of simulator attributes that can be used for a START. Additional examples of attributes and definitions may be found in Johnston et al., (2015).

Table 1. Example START Attributes

Attribute	Definition	Example
Awareness	Ability to sense oneself moving in an environment. Ability to sense the movement of the environment or other objects relative to self.	Sense of objects, people, vehicles, etc. moving past as the individual moves through the environment (riding, walking, running, turning). Shifts in visual reference corresponding to body position, such as standing, kneeling, prone, falling, etc.
Damage States	Ability to provide and receive standardized and correct visual depictions of damage states in response to ordnance or other physical conditions.	Simulated damage states are commensurate with munition types and environmental conditions.

Input: Task-to-Attribute Criticality

The level of the attribute’s fidelity required to conduct tasks is determined by the task-to-attribute criticality evaluations. After community validation of the identified tasks, the attributes (those elements that the training device and environment provide to support task performance), of the training device are evaluated in terms of criticality. The task-to-attribute criticality rating, together with the attribute-to-capability rating, provides the outputs. This rating is based on the question: how *critical* is an attribute when performing a given task in live, operational training?

Below is an example of what the START spreadsheet looks like.

Name of Device		Attributes							
Task #	Task	A1	A2	A3	A4	A5	A6	A7	A8
1	Task Statement	1	2	3	1	2	3	3	3
2	Task Statement	3	2	1	3	2	1	1	1
3	Task Statement	3	2	1	3	2	2	2	2
4	Task Statement	1	2	3	1	2	3	3	3
5	Task Statement	3	2	1	3	2	1	3	3
6	Task Statement	3	2	1	3	2	2	1	1
7	Task Statement	1	1	2	2	3	3	1	1
8	Task Statement	1	1	2	2	3	3	2	2

Figure 2. Task-to-Attribute Criticality Example

*Note: Any attribute rated a criticality of “1” is blacked out on the capability spreadsheet and does not require a capability rating. Figure 3 below illustrates this pass-thru.

Input: Attribute-to-Capability

After task-to-attribute criticality data is collected for each of the identified tasks, evaluation of a training device’s *capability* to deliver the necessary attribute’s fidelity is assessed (Dunne, Harris, Arrieta, 2018). The attribute-to-capability rating, together with the task-to-attribute criticality rating, provides the results. The attribute-to-capability rating asks the question: how capable is the simulator in providing the level of fidelity determined to be required when performing a given task in live, operational training?

Below is an example of what the START spreadsheet looks like as attribute-to-capability levels are collected and integrated with task-to-attribute criticality data.

Name of Device		Attributes							
Task #	Task	A1	A2	A3	A4	A5	A6	A7	A8
1	Task Statement		2	2		3	3	2	2
2	Task Statement	3	2		3	3			
3	Task Statement	3	2		3	3	3	2	2
4	Task Statement		1	2		3	3	3	3
5	Task Statement	3	1		3	2		1	1
6	Task Statement	3	1		3	2	2		
7	Task Statement			3	3	2	3		
8	Task Statement			3	3	2	2	1	1

Figure 3. Attribute-to-Capability Input

Output: Task Training Support (TTS)

Once criticality and capability data are collected, the result is two sets of scores: the task training support (TTS) and code training support (CTS) scores. The TTS score derives from SME’s evaluation of the level to which a simulator’s attributes enable training of tasks. Levels, based on the attribute capability description and ratings shown below, are averaged across all tasks.

Table 2. TTS Levels and Definitions

Level	Description
Level 3	Training device is capable of supporting operator performance of the task and sufficient to allow T&R qualification of the operator upon satisfactory performance of the task.
Level 2	Training device provides attributes at a fidelity sufficient for beneficial training but not for T&R qualification.
Level 1	Training device is incapable of supporting training for the task.

Output: Code Training Support (CTS)

CTS scores are calculated by cross-referencing TTS scores to the task criticality evaluations and produce two types of CTS scores: CTS₁ and CTS₂. The CTS₁ score indicates the training device’s capability with respect to tasks critical to successful completion of the associated T&R event and therefore critical for the training device to enable. Tasks deemed critical are weighted heavier than those simply deemed relevant. The CTS₂ calculation indicates the device’s capability with respect to both non-critical and critical training tasks associated to a T&R event. The CTS scores are thus derived from the level of training support a device provides for execution of specific tasks associated to T&R events as defined in the applicable T&R manual, and are expressed on a scale of one to five as shown below.

Table 3. CTS Levels and Definitions

Level	Description	Definition
5.00	Full training capability	T&R code can be trained <i>thoroughly and accurately</i> in the simulator with no compensation required for the individual to execute and accomplish the T&R code.
4.00	High training capability	T&R code can be <i>effectively trained</i> in the simulator with minor compensation required for the individual to execute and accomplish the T&R code.
3.00	Moderate training capability	T&R code can be <i>trained</i> in the simulator, but with considerable distractions requiring significant compensation for the individual to execute and accomplish the T&R code.
2.00	Low training capability	T&R code can be <i>addressed</i> in the simulator, but with severe distractions requiring extraordinary compensation to have a useful affect towards executing and accomplishing the T&R code.
1.00	No training capability	T&R code <i>cannot be trained</i> in the simulator, and no amount of compensation allows the individual to effectively execute and accomplish the T&R code in the simulator.

Output: Attribute-to-Task Comments

Attributes are also used to collect informal input from participants to further illuminate the evaluation results and provide additional data. An example of how the attributes comment sheet is used to gather information about supporting tasks is included below.

Task #	Attribute Comment
P1	Equipment Operation
T001	Purely digital communication set-up; overall communication concept does not include voice radio systems; want to look at SAVT radio mock-up in system A01/A02-1: component equipment not present A04-1: component equipment not present A05-2: Operator has to control frequencies, JTAC can't load frequencies; operated
T002	Does not exist in sim
T003	Does not exist in sim
T004	Does not exist in sim
T005	A03: updated criticality to 1; capability N/A A05- 4: tracking (optical tracking) slightly decreased functioning with sim (slower or unresponsive if move sim too slowly; real-world is highly sensitive to movement) A10/A11- 4: if operator changes sight picture, sim scopes (PLDR) do not change until operator "hits button" to relink sight picture; has ability to link sight pictures but is merely an annoyance—operator needs to know to "hit button" to relink sight pictures
T006	A03: updated criticality to 1; capability N/A Overall—see note regarding bore sighting equipment A05- 4: should not be able to see ISLT light beam with the naked eye (does not appear during the day but the beam is present when operating under night conditions) A15- 3: beam fidelity not accurate for light levels; beam present at night and not present during the day; no gradation in beam representation

Figure 4. Attribute-to-Task Comments

Output: Criticality-to-Capability Misalignment

This output is of central importance to the assessment and vital to resource sponsors, and decision makers. By identifying simulator attributes that are critical to the performance of a task or T&R event, or both, but are not supported sufficiently by the capabilities of the simulator actionable evidence is documented. These gaps may be hardware or software, or both, and once identified they can be addressed by engineering change proposals or upgrades.

DMTE Recommended Protocol

Although the DMTE evaluation methodology is the focus of this paper, the authors note that the stand-alone methodology maps closely to that used for evaluation of the distributed configuration. Using a review of the processes undertaken, their sequencing, outcomes, and lessons learned, a recommended protocol is presented below to aid in follow-on DMTE evaluation efforts. A brief use-case describing how this protocol originated from the DMTE then follows. The protocol is an illustration of the recommended actions; it is not a "cookie-cutter" solution to all similar efforts. The authors' intent is to continue to revise and improve this protocol, and ultimately standardize across the Services as a common reference for similar efforts, both stand-alone and DMTE.

Table 4. Distributed Recommended Protocol

Phase	Definition
Initiating	Identify and engage primary organization(s) and command and control personnel.
	Identify and concretize the system of systems configuration.
	Define scope: Collect requirements of successful evaluation. Is this holistic or limited evaluation?
	Personnel requirements: Identify and secure the necessary operators, instructional systems architects (ISAs), subject matter experts (SMEs), instructors, engineers, exercise command, etc.
	Technical facilitation requirements: Identify and document what is required to ensure desired configuration.

	<ul style="list-style-type: none"> • Enumerations • Mapping data/correlation • Version types (e.g., DIS) • After action review (AAR)/data logging/tools and timing 	<ul style="list-style-type: none"> • Communication architecture • Network connectivity • Command and control systems • Collective objectives 							
	Instructional architecture facilitation requirements: Identify, locate, and archive previous evaluations associated to the systems to be evaluated. Identify the training objectives or tasks, or both, that will provide optimal return on investment for assessment. Draft training task lists.								
Planning	Determine necessary activities.								
	Create and coordinate schedule, activities, and resources with stakeholders.								
Design	For each system, verify and validate identified training objectives (e.g., T&Rs) or tasks, or both, for use in the distributed exercise.								
	Draft the distributed exercise using the identified training objectives or tasks, or both, as the framework.								
	Draft the distributed scenario using the draft distributed exercise as the framework, and align tasks.								
	Define simulator attributes to be evaluated: These may differ from one simulator to another depending on type of simulator, training objectives or requirements.								
Pre-Exercise	Validate training task lists with SME input.								
	Develop and illustrate the assessment plan design. (Example: Figure 1)								
	Prepare and deliver exercise and workshop read-ahead to control expectations and familiarize the START participants with the inputs, the process, and outcomes.								
	Finalize and document the distributed exercise and scenario. Includes exercise design, personnel requirements, technical requirements, communication equipment (e.g., green gear, simulated, etc.).								
	Conduct technical tests evaluating the technical capabilities, training network, enumeration, and terrain correlation. Correct deficiencies. Validate changes, Network stress test, and confirmation brief. Rehearse scenarios in semi-distributed and fully distributed configuration. Verify receiving stimuli (inputs – visual/audio/haptics). Verify output stimuli with associated simulator/simulation.								
	Give brief on START to reinforce read-ahead and prime the participants for data collection.								
Exercise	Conduct exercise with sufficient numbers of ISA, Instructional system designers (ISD) and observers to capture informal comments and answer questions on a <i>non-interference</i> basis.								
Post-Exercise	Criticality workshop: For each system, collect task-to-attribute criticality evaluations from START participants. Note: Depending on schedule, availability or requirements, this data collection may also be done pre-exercise, but after verification and validation of task lists.								
	Conduct capability workshop: For each system, collect attribute-to-capability evaluations from START participants.								
	Qualitative input (survey, questionnaire, interview, etc.) from both technical and training perspectives.								
Analysis	Training tasks mapped to T&R Events, programs of instruction, training objectives.								
	Determine training task support (TTS) and code training support (CTS) Levels.								
	Compare stand-alone ratings to identical tasks conducted in the distributed configuration.								
	Identify qualitative commonalities or other salient finding.								
	Conduct analysis of:								
	<table border="0"> <tr> <td>• High and low ratings</td> <td>• Collective objectives</td> </tr> <tr> <td>• Exercise and scenario design</td> <td>• Criticality-to-capability misalignment</td> </tr> <tr> <td>• Attribute-to-capability gaps</td> <td>• AAR/data logging/AAR tools and timing</td> </tr> <tr> <td>• Technical requirements (trace to impact at task level)</td> <td>• Compare ratings to qualitative feedback to identify areas of misalignment</td> </tr> </table>		• High and low ratings	• Collective objectives	• Exercise and scenario design	• Criticality-to-capability misalignment	• Attribute-to-capability gaps	• AAR/data logging/AAR tools and timing	• Technical requirements (trace to impact at task level)
• High and low ratings	• Collective objectives								
• Exercise and scenario design	• Criticality-to-capability misalignment								
• Attribute-to-capability gaps	• AAR/data logging/AAR tools and timing								
• Technical requirements (trace to impact at task level)	• Compare ratings to qualitative feedback to identify areas of misalignment								
Document all workarounds used, record deficiencies, record lag time, latency, no reaction, correlation issues, fidelity of distributed conditions, record mitigation strategies, record troubleshooting steps, determine engineering upgrade proposals from which costs can be estimated. Document lessons learned.									
Document	Report results, recommendations and deliver to stakeholder(s).								

These phases manifested during conduct and analysis of this DMTE START. For example, during the initiate phase the scope was solidified by recognizing characteristics that define the project as holistic or limited in approach. If holistic, then the evaluation spans the complete list of tasks or events the simulator is intended to address. If limited, then the evaluation focuses on only those tasks present in the scenario or set of exercises serving as the framework of the evaluation. In the case of this DMTE the scope was limited to a single exercise and the tasks identified pertained only to those tasks required of the scenario.

During the planning phase, it is imperative to work with stockholders and other appropriate acquisition professionals to identify synergistic opportunities that may align to ongoing test and evaluation assessments such as operational, and user evaluations. The groundwork for this collaboration is laid during the initiate phase when the value of obtaining organizational commitment cannot be overstated.

At the beginning of the design phase, it is necessary to determine the tasks or T&R events, or both, in order to align them to the exercise and analysis phases. For this DMTE the tasks were aligned to their associated T&Rs based on three sources: community identified T&R events, the LVC training environment (LVC-TE) working integrated product team convened February 2016, and appropriate T&R manuals.

Once identified, analysts conduct pre-exercise data collection workshops to validate that the specified tasks were complete and accurately associated to T&R events or other training objectives and determine how critical each training device attribute is to the training of each task (i.e., task-to-attribute criticality).

Also during the pre-exercise phase, testing determines network stability, connectivity, and maximum capability, and technicians conduct scenario rehearsals. For this DMTE, because these tests had been done, when participants arrived they efficiently completed the pre-exercise phase; given pre-briefings and familiarization exercises on the training devices. The familiarization exercises were first conducted in limited distributed configurations and then in the fully distributed configuration prior to beginning of the exercise phase.

Then, the full DMTE exercise was conducted with all systems connected. Of the tasks and T&R events identified, 137 collective tasks associated to 86 T&R events formed the basis of the DMTE scenario. During this phase, analysts observed the participants, the system interactions and gathered data from both technical and operational SME input.

Upon completion of the exercise to evaluate the capability of each training device to train each task (Dunne, et al., 2017; Johnston et al., 2015) analysts collected additional START data, (i.e., attribute-to-capability) and conducted informal interviews to collect comments and feedback from each of the system participants (i.e., attribute-to-task comments).

During the analysis phase, START analyses determined TTS and CTS scores, comments were aggregated and DMTE exercise START results were compared to stand-alone START results where previous START documentation was available and appropriate. Documentation was undertaken to present results, conclusions, lessons learned and recommendations.

The documentation of the START, either stand-alone or distributed configuration report, presents deltas (e.g., criticality-to-capability gaps) that then can serve as “trigger mechanisms” for initiation of engineering change proposals or additional requirements. For example, if conduct of a task requires a high level of fidelity from a particular attribute (i.e., level 5) but the simulator’s capability to present that attribute is evaluated as low (i.e., level 1) then a gap exists that indicates the need for improvement of that particular attribute.

LESSONS LEARNED

The following lessons are considered the most salient, as they are immediately actionable in conduct of the Marine Air-Ground Task Force-Tactical Integrated Training Environment (MAGTF-TITE) during TALONEX 2-18 in Mar 2018 follow-on. They are not presented hierarchically.

When comparing the DMTE START of the CCS simulator to an available stand-alone START, an overall diminishing of capabilities to train identical critical tasks was indicated. This raised two questions; 1) did the DMTE scenario allow action of those tasks and, 2) was there a technical interoperability issue that affected the simulator’s capability

to enable the tasks in the DMTE that did not affect its capability in the stand-alone configuration? Further analysis found that the design of the distributed scenario did not allow action of all the identified tasks in the DMTE configuration. It is therefore essential that the design of the exercise and scenario ensure alignment to identified training objectives of all simulators within the DMTE.

Attributes contained in the traditional START process do not account for conditions unique to DMTE exercises; they do not capture network requirements or allow for capture of interoperability gaps. This was clearly the case when it was found that an attribute that was addressed to an acceptable degree in the stand-alone configuration of the VBS JTAC station did not transfer to its distributed configuration when it was assumed it would. To be specific, during a stand-alone START the software functions associated to image/model distance, referred to as “object draw distance” performed optimally. However, during the DMTE START the “object draw distance” became a challenge that negatively affected conduct of the DMTE exercise. Software and hardware testing was then conducted to determine appropriate engineering changes, upgrades or enhancements required for mitigation. Attributes that become essential due to the unique conditions created by a distributed exercise should be included in the START list of attributes.

During the task-to-attribute criticality and criticality-to-capability workshops, it is recommended to have at least 3-5 highly qualified SMEs; qualification criteria to be determined in the planning phase. At this point in the evaluation process, the actual simulator or device being evaluated is *not* necessary because the focus is on rating the criticality of attributes to reach training objectives. If there is disagreement on how to rate a criticality for a particular attribute, one SME should be designated as the final decision maker.

Plan the workshops in a way that the tasks can be assessed easily. Although there is no specific formula for assessing tasks, it is recommended that common tasks be bucketed so they can easily be assessed at once. During the criticality-to-capability assessment, the simulators may be used to aid the SMEs in their evaluation. More than likely, the SMEs will have to run through a simulation several times to ensure all capabilities are assessed.

T&R events, tasks, and other training and learning objectives were sourced from the associated communities, and T&R manuals. However, lack of measures or means to determine acquisition, sustainment, or maintenance of trainee proficiency proved problematic. ISAs or ISDs, in collaboration with associated organization SMEs, should determine the tasks in order to ensure measurable learning objectives (e.g., describe, rotate, operate etc.) are part of the task description. In this way, important factors that may affect measures of proficiency are made available for training effectiveness evaluations.

Because the DMTE scenario was not known until the week of the event, and the task list had been validated and verified after development of the scenario had begun, ensuring coverage of all the necessary tasks was not possible. For this reason, the protocol provided herein inverts the order to allow complete coverage of tasks by a scenario.

With this methodology it was determined that value added training was achieved through the interoperation of simulators. This methodology illuminated technical gaps where innovative solutions prevailed. Recommendations that were forwarded as result of this methodology highlighted the importance of cutting-edge technical affordances keeping pace with training objectives.

MOVING FORWARD

The objective of the next phase, *capabilities expansion*, is to improve training effectiveness by fully integrating the virtual and constructive environments during an operational exercise. This will be done based on recommendations; application of these lessons learned and continued revision of the methodology outlined in this paper. Downstream efforts for standardization of this methodology across the services are under consideration. For example, a common menu of simulator attributes that resulted from this effort may soon be accredited by TECOM and put in place for use across MAGTF exercise evaluations. The list that resulted from this effort is currently under review.

This methodology can be applied by the larger Modeling and Simulation community anytime assessment of simulator inter-operability is required, and used by acquisition entities to develop requirements, conduct pre-acceptance testing, and augment verification, validation and accreditation products.

REFERENCES

- Department of Defense, (2010). Modeling and Simulation (M&S) Glossary. Washington, DC: Department of the Defense.
- Department of Defense, (2016). Department of Defense Dictionary of Military and Associated Terms Joint Publication 1-02. Washington, DC: Department of the Defense.
- Dunne, R., Harris, S., Arrieta, A., Tanner, S., Vonsik, B., Lalor, J., & Muir, S. (2017). *Live, Virtual, Constructive Distributed Missions: Results and Lessons Learned*. Proceedings of the Interservice/Industry Training, Simulation and Education Conference 2017. Orlando, FL.
- Johnston, J., & Nolan M., Caldwell, J. (2015). *Capability Assessment of Test and Live Training Systems for Real-Time Casualty Assessment*. Proceedings of the Interservice/Industry Training, Simulation and Education Conference 2015. Orlando, FL.
- Jones, N., & Arrieta, A. (2017). *Effective Face Validation Methodology to Evaluate Simulation Fidelity for Training*. Tutorial in Proceedings of the Interservice/Industry Training, Simulation and Education Conference 2017. Orlando, FL.
- Sheehan, J.D., Merket, D.C., Sampson, T., Roberts, J. Merritt, S. (2009). *Human System Capabilities-Based Training System Acquisition in Naval Aviation*. Proceedings of the 2009 Human Systems Integration Symposium. Alexandria, VA: American Society of Naval Engineers.
- United States Marine Corps. (2015). *36th Commandant's Planning Guidance 2015*. Washington, DC: Department of the Navy.