

## Performance Assessment in a Virtual Simulation for Integrated Austere Medical Operations Training

**Benjamin R Bauchwitz**  
**Peter W Weyhrauch, James M Niehaus,**  
**Michael Makivic, William Manning**  
**Charles River Analytics, Inc.**  
**Cambridge, MA**  
**{bbauchwitz, pweyhrauch, jniehaus,**  
**mmakivic wmanning}@cra.com**

**John Broach**  
**Peter Lancette**  
**University of Massachusetts**  
**Memorial Medical Center**  
**Worcester, MA**  
**{John.Broach,**  
**Peter.Lancette}**  
**@umassmemorial.org**

**Christopher M Garrison**  
**Frank E Ritter**  
**Pennsylvania State University**  
**University Park, PA**  
**{cmg35, fer2}@psu.edu**

### ABSTRACT

The DoD is preparing for future austere environments where it will not be possible to sustain the current level of medical operations across the chain of care. Enemy activity, resource depletion, and lack of safe spaces to deliver care will all provide unique challenges to the performance of medical operations. Under these circumstances, medical operations will require the coordination of personnel outside traditional medical roles, such as those involved in logistics or mission command. The DoD has neither a curriculum nor a suitable training system for teaching the skills necessary to succeed in these operational environments.

Here we report initial work designing the training curriculum and performance assessment metrics for Extensible Field and Evacuation Care Training in a Virtual Environment (EFFECTIVE), a virtual simulation for training, practicing, and assessing the skills for managing medical operations in austere and far forward environments. EFFECTIVE is intended to provide training for personnel in traditional medical roles (combat lifesavers, flight nurses, field surgeons, etc.) as well as those in roles that interface with medical personnel during complex missions. This training is designed to emphasize coordination of activities across these different roles as well as the problem solving and decision-making skills that become necessary when medical operations cannot be performed as intended.

We applied the Methodology for Annotated Skill Trees (MAST), a previously published cognitive task analysis framework, to design the curriculum for this training system. The results of this initial work were (1) a set of plausible training scenarios inspired by challenges faced by today's Warfighters but expanded to represent future austere battlespaces; (2) a set of critical skills needed by medical, logistics, and mission command personnel to successfully perform their duties in these scenarios; and (3) definitions for concrete, objective metrics for assessing proficiency on these skills in the EFFECTIVE simulation environment.

### ABOUT THE AUTHORS

**Benjamin Bauchwitz** is a scientist at Charles River Analytics studying intelligent virtual training environments. His work involves studying the roles of fidelity, adaptability, and modality on training, establishing models of skill for tasks in a variety of domains, and designing medical training technology on platforms ranging from manikins to computers to virtual- and augmented-reality. He received his BS in Cognitive Science from the Massachusetts Institute of Technology and is pursuing a PhD in Computer Science from Duke University.

**James Niehaus** is a Principal Scientist and Director in the Human-Centered Intelligent Systems Division at Charles River Analytics. Dr. Niehaus's areas of expertise include artificial intelligence and training systems, and he is the principal investigator on multiple efforts developing new training technologies. He also specializes in the computational modeling of narrative, collaborating on efforts to develop effective narratives for target populations and to develop software that enables intelligence analysts to use narrative as an organizational principle in analysis. Dr. Niehaus has a BS in computer science from the College of Charleston and a PhD in computer science from North Carolina State University.

**Michael Makivic** is a Software Engineer at Charles River Analytics. His areas of expertise include game development, virtual environment design, and virtual reality and augmented reality user interaction. Mr. Makivic works on game-based training and simulation projects spanning a broad set of domains ranging from medical care to aircraft maintenance. Additionally, Mr. Makivic regularly develops and consults on projects related to 3D modeling and animation, particularly as they relate to game design and virtual environments. Mr. Makivic has a BS in Computer Science from Brandeis University.

**William Manning** is a Software Engineer at Charles River Analytics who specializes in game design. His areas of expertise include topics in computer graphics such as shader programming, rendering techniques, and optimization; topics in procedural generation such as simulation and creation of artistic or organic content, textures, and 3D models; and topics in gaming such as virtual reality, augmented reality, game engine architecture, pathfinding, game AI, simulation, and optimization. He received a BS in Computer Science and Game Design from Northeastern University.

**John Broach** is an Emergency Physician at the University of Massachusetts Medical Center in Worcester, MA and is the Division Director of EMS and Disaster Medicine. His academic work focuses on the forward deployment of technology in disasters and best practices in disaster response as well as EMS practice. He earned his BS, MPH, and MD from Northwestern University and an MBA from George Washington University.

**Peter Lancette** is Director of Heart and Vascular Intervention Labs and the LifeFlight paramedic flight program at UMass Memorial Hospital. He is an emergency department nurse with a decade of experience working at a University Level I Trauma Center, working with high patient volumes. He served for 26 years in the US Army Reserves while fulfilling leadership positions where he completed the Army's Critical Care Nursing Program and was the Chief Nurse of the 912th Forward Surgical Team, responsible for direct patient care, clinical operations, and training during a tour of duty in Iraq. MAJ Lancette received his BS and MS in Nursing as well as an MBA, all from Grand Canyon University.

**Christopher Garrison** is an Associate Teaching Professor in the College of Nursing at the Pennsylvania State University and the Director of the Simulation Laboratory. His clinical background includes experiences in medical-surgical, cardiology, home health, and as an adult/gerontological nurse practitioner. He is a Certified Nurse Educator (CNE) and a Certified Healthcare Simulation Educator (CHSE). He holds a BS degree from the Pennsylvania State University, an MSN from George Mason University, and a PhD from Nova Southeastern University. His research interests include evaluating the effectiveness of simulation and virtual simulation as an educational strategy.

**Frank Ritter** is a professor in the College of IST at the Pennsylvania State University. He researches the development, application, and methodology of cognitive models, particularly applied to interface design, predicting the effect of behavioral moderators, and understanding learning. He has published a textbook on psychology principles for systems designers with the Director of User Experience at Google, and has a similar book coming for the design of operation centers. Dr. Ritter holds a BSEE in Electrical Engineering from the University of Illinois at Urbana-Champaign, an MS in Psychology from Carnegie Mellon, and a PhD in Artificial Intelligence and Psychology from Carnegie Mellon.

**Peter Weyhrauch** is a Principal Scientist and the Vice President of the Human-Centered Intelligent Systems Division at Charles River Analytics. He researches intelligent adaptive training, cognitive models of human performance, and AI-driven decision aids. He is particularly interested in applying ontology-based models of human performance to scenario generation, adaptive training, and human behavior modeling. He earned his BS in Computer Science and Engineering from MIT and his MS and PhD in Computer Science from Carnegie Mellon.

## **Performance Assessment in a Virtual Simulation for Integrated Austere Medical Operations Training**

**Benjamin R Bauchwitz  
Peter W Weyhrauch, James M Niehaus,  
Michael Makivic, William Manning**

**Charles River Analytics, Inc.  
Cambridge, MA**

**{bbauchwitz, pweyhrauch, jniehaus,  
mmakivic wmanning}@cra.com**

**John Broach  
Peter Lancette**

**University of Massachusetts  
Memorial Medical Center  
Worcester, MA**

**{John.Broach,  
Peter.Lancette}  
@umassmemorial.org**

**Christopher M Garrison  
Frank E Ritter**

**Pennsylvania State University  
University Park, PA**

**{cmg35, fer2}@psu.edu**

### **BACKGROUND**

The DoD expects to operate in future contested environments where the current practice of providing comprehensive medical care to all wounded casualties will not be feasible. Future conflicts with near peer adversaries may produce situations in which enemy activity and high casualty volumes overwhelm available personnel, supplies, and treatment facilities (Cronk, 2018). For example, the increase of low-scale terrorism as a warfare tactic may result in overwhelming civilian casualties or damage to key medical care facilities (Johnson, 2015). Similarly, highly capable adversaries may disrupt important communications or other infrastructure vital to the medical operation (Dian, 2015; Garcia, 2015; Morris et al., 2015; Thomas, 2014). This evolving landscape is likely to at times require substantial deviations from traditional doctrines regarding casualty care (Becker & Link, 2011; Keenan, 2007). Overall, anticipated future medical operations will require highly synchronized interaction between medical, combat, and logistical teams, as well as other entities, such as host nation medical services.

Training personnel to perform in this domain is inherently problematic because the set of activities to be mastered is so diverse and because the threats that will be encountered may not be well understood in advance. DoD training in large part depends on personnel performing in-depth training modules geared specifically toward certain tasks they will perform in the course of their duties. For example, the military's Joint En Route Care Course (JECC) and the similar civilian Transport Professional Advanced Trauma Course (TPATC) are courses that are commonly taken by DoD en route care personnel such as flight nurses, paramedics, and physicians. While these courses may provide strong training for en route care-specific skills and tactics, they do not provide the type of cross-role integrated response training necessary for coping in mass casualty environments (Hoang et al., 2012). Courses for other personnel are similarly constrained (e.g., C-STARS, for first responders). Additionally, this training model is highly reactive and insufficiently adaptive to emerging threats; it takes months to years to revise the courses, and even once revised, they address threats that were new 6-12 months before.

The gaps described above cannot be adequately addressed through live, in-person training because it is infeasible to gather individuals from all required disciplines in a setting where they can regularly train. The DoD has endorsed the use of virtual training, particularly for rehearsal of skills for austere environments where the individual performers are numerous and often stationed at geographically distant locations (Lang, 2016). The effectiveness of virtual training has been demonstrated in many domains, but particularly for medical skills (Agarwal & Bandeli, 2018; Satava & Jones, 1997). In addition to enabling organization of training among distributed groups that would otherwise be infeasible, virtual training provides an opportunity to use cutting edge technology to enhance the training experience beyond what could be achieved in traditional live training. For example, advanced AI technology can be used to create simulated agents, such as teammates with whom a trainee can rehearse certain team skills (Andronikos et al., 2016; Rickel & Johnson, 1999; Seeber et al., 2020). Despite these advantages, virtual training has not been widely adopted for the particular challenges of integrated training for austere medical operations. Additionally, providing a means to rapidly update training in response to emerging trends remains an important challenge in this domain; virtual training environments do not naturally lend themselves to ready modification by training stakeholders (who are not themselves software engineers), and the instructional design process itself may not be sufficiently agile to provide training that

remains at the cutting edge. Our overall motivation in this work was to establish a curriculum for managing medical care in austere, far-forward, and mass casualty environments and design a training platform, Extensible Field and Evacuation Care Training in a Virtual Environment (EFFECTIVE), for delivering this training in a way that captures emerging operational challenges and provides a means to rapidly update the training content.

The instructional design methodology we employed in the design of this system is based on the ADDIE process (Analyze, Design, Develop, Implement, Evaluate) (Allen, 2006; Morrison et al., 2019). As part of this process, we identified the key skills for the target domain by eliciting a comprehensive set of representative scenarios from end users. There are a variety of techniques that have been used to model skills in complex domains like medicine to varying degrees of formality (De Novi et al., 2012; Mackel et al., 2007; Ottensmeyer et al., 2012; Reznick et al., 2003; Risser et al., 1999; Tanner, 2006). Methods based on hierarchical task decomposition (Cao et al., 1996, 1999; MacKenzie et al., 2001) are particularly useful in this domain. From an instructional design perspective they make it easy to conceptualize and organize the skills at a variety of levels of abstraction, which is essential when thinking about an integrated skills training environment where multiple roles are receiving training in the same system (it is necessary to have not only a model of the skills for performing a particular task, but also a model for how that task relates to other tasks an individual will perform, as well as a model for how that individual's actions interact with the activities of other individuals). They are also very useful from a technical standpoint. Hierarchical decompositions make it easy to abstract the skill model into a "network", which makes it easier to encode it in knowledge representation systems like ontologies (Middleton et al., 2018; Niehaus et al., 2014), which can be used by AI-based systems for advanced reasoning (e.g., helping an intelligent training system reason about the skill of a trainee based on their actions). They are also helpful for developing protocols for the objective, data-driven assessment of skills, which is critical for a virtual environment where an instructor may not have as much oversight (skill A is mastered if trainee successfully completes tasks X, Y, and Z; task X is successfully completed when subtasks x1, x2, and x3 are performed). In that way, it can enable a virtual training system to provide feedback that explains the root causes of failure as well as providing appropriate enrichment content.

Our approach to the design of curriculum is called the Methodology for Annotated Skill Trees (MAST); it is derived from other forms of hierarchical task decomposition, but augments the task hierarchy with "annotations" that enable it to be conceptually linked to the properties of virtual training that make use of the skill model content (Bauchwitz et al., 2018, 2019; Grosdemouge et al., 2013; Perez et al., 2013). The set of annotations is customizable and the exact annotations used in any particular model vary, though annotations we have typically used to design simulation-based training include (1) Assessment Metrics, which describe objective criteria used to determine whether a task was performed correctly, as well as whether an underlying skill can be considered mastered; (2) Learning Curves, describing progression of proficiency on the task in response to other independent variables, such as time since last refresher training or number of exercises performed; (3) Cognitive Information Elements, which link a task to things a trainee would think about while learning or performing the task, such as required background knowledge, goals for performing the task, or relevant situations with which to maintain awareness; (4) Instructional Material, which link to material external to the simulation that provide instruction on declarative knowledge necessary to initially learn the skill; (5) Decision-Making Rules, which capture the cues a trainee would use to decide between different courses of action; and (6) Meta Information, which capture information about the task taxonomy itself, such as how difficult a skill is to train and what types of consequences exist if the skill is not trained (this is information a training designer would use, rather than the person practicing the skill).

## **METHODS**

The first step in our skill modeling process was to elicit illustrative scenarios from representative end users. We assembled a team of subject matter experts including officers from both the Army and Air Force medical systems, civilian emergency medicine specialists with both physician and nursing expertise, leaders in mass casualty response training for two different emergency services, and three university professors specializing in medical education, nursing education, and intelligent training systems. We elicited descriptions of scenarios they had encountered in the course of their duties that tested the limits of their organization's ability to deliver effective care. We elicited anecdotes describing times they had to improvise on the job. We asked them to characterize the "soft skills" that were crucial to success in the challenging scenarios they encountered that were not related to their core technical expertise. We conducted exercises in which we had the subject matter experts extrapolate their past experiences into plausible future scenarios that could be encountered given various geopolitical climates.

From these exercises, we characterized the core types of scenarios that could be encountered. This process focused on establishing scenario archetypes, which describe the unique thematic elements of a scenario, rather than establishing specific details such as setting or number and type of casualties present. This was done because training for this domain is intended to be flexible—within a given training framework, the underlying goals and core skillset for the trainee remain the same regardless of the specific mission objectives, casualties, and locations. Scenario descriptions describe the general spectrum of tasks performed by different types of personnel in each of the scenarios; within a simulation-based training system, this helps inform the design of user interfaces, perspectives, and levels of fidelity so that each training module is optimized for the particular role and skillset it addresses.

From the scenario and task descriptions, we then used the MAST hierarchical task decomposition framework described above to build an inventory of the skills needed to perform each task. There were several aspects to this analysis. First, we worked with subject matter experts to define the conditions under which tasks could be completed; this provided insight into secondary tasks that needed to be performed, either by that individual or other individuals, and defined a structure for the relationship between the skills needed to carry out tasks. Second, we worked with the subject matter experts to define all of the non-trivial declarative knowledge that a person would need to learn in order to carry out that task. For procedural skills, this involved defining any domain knowledge a person would need to receive initial instruction on the task, as well as defining explicitly the goals, hazards, and consequences that an individual would keep in mind while carrying out the task. For decision-making skills, we attempted to characterize the intuition that someone would use to arrive at an acceptable decision under uncertainty (note that in this domain, the goal is not always to arrive at a “correct”, or even “optimal” decision, as constraints on time, resources, and available information may render such an outcome infeasible). Third, we asked the subject matter experts to define as concretely as possible metrics by which they would consider skills to have been mastered. Because of the heterogeneity in skills required for this domain, these metrics took a variety of forms. For example, when an individual uses a procedural skill to manipulate their environment, a success metric may define whether or not they manipulated the environment into the desired end state, as well as how reliably they were able to repeat that process. For decision-making skill, successful completion is defined by how positively the decision impacted the mission, and so this process involved defining the mission parameters that could be affected by that decision downstream. Finally, we applied evidence from learning theory to identify effective instructional strategies for each skill, which define the type and sequence of training content used to teach that skill.

## RESULTS

The result of this skill modeling and instructional design process was that we identified a set of core scenario archetypes that cover the broad set of activities likely to be performed in support of austere, far-forward medical operations. The three major sets of themes in these scenarios are (1) *Mission Management via Casualty Care*, (2) *Sustaining Casualties after Resource Exhaustion*, and (3) *Managing Medical Logistics*. Here we define these scenarios in terms of their objectives, the skills they address, and how they are portrayed to trainees in the education process.

Scenarios focused on *Mission Management via Casualty Care* are those in which trainee must provide medical care to a large number of casualties during a “must-do” mission that cannot be abandoned. This type of scenario is experienced from either the perspective of the medical group commander or the senior medic. When the trainee takes the perspective of the medical group commander, they receive a sequence of updates from the line officer responsible for the operational mission regarding the goals and status of that mission as the scenario progresses. The trainee must take these commands and make decisions on how to allocate medical resources so that the mission has the greatest chance of success. They must then communicate these decisions to the senior medic to implement. When the trainee takes the perspective of the senior medic, they receive a series of instructions from the medical group commander and use that direction to manage the treatment of 6-8 casualties. During the scenario, the trainee must make critical judgements about which casualties to treat first, decide how to treat them, and understand where those decisions differ from standard doctrine. In addition, the trainee must exhibit communication skills by translating mission requirements into medical actions and communicating them between medical and logistic personnel.

Scenarios focused on *Sustaining Casualties after Resource Exhaustion* are those in which the trainee must provide medical care to a large number of casualties under austere conditions. In this scenario type, supplies are depleted and there is no possibility for them to be replenished within the timeframe of the scenario. The trainee must creatively use natural and non-medical resources at their disposal to continue delivering care and at times may need to interact with non-DoD personnel to obtain support from sources outside the US Military (such as host-nation medical facilities).

These scenarios require situational awareness and creativity to successfully provide treatment with makeshift supplies, and include key decision points where the trainee must break with traditional practice and obtain support from outside the established DoD chain of care.

**Table 1: Key skill categories for Mission Management via Casualty Care scenario archetype**

Declarative Knowledge	Communication Skills
What are the mission goals and requirements?	Can you translate mission-domain requirements into medical-domain commands?
What are the likely sources of bottlenecks to delivering care in the operational environment?	Can you concisely and effectively update relevant personnel on evolving mission needs?
What are the capabilities of your available equipment?	Can you barter with other actors to obtain assistance or supplies?
What are the potential wounds you will encounter?	Can you convince personnel to deviate from standard doctrine when appropriate?
What are the roles and skillsets of the personnel in your vicinity?	
What friendly, hostile, and host-nation actors are in your area?	
What role does each active personnel play in the mission?	
What are all the possible means by which the injuries you encounter can be treated?	
Perceptual/Situational Awareness Skills	Judgement/Decision-making Skills
Awareness of wider impact of care given to individual casualties on the mission and on care of others	Which casualties to treat first or not treat at all
Presence and state of available resources	How to conserve limited/bottleneck supplies
Multi-functionality of resources that may not be traditional medical supplies	When and when not to deviate from standard doctrine or medical rules of engagement
Awareness of goals/needs of other actors who may be able to help you	Which non-medical supplies can be rigged to function as medical supplies
	What should the source of treatment be for different casualties (e.g., who can be sent back for buddy aid, who needs to use medical team resources)

Scenarios focused on *Managing Medical Logistics* are those in which the trainee acts as a logistics coordinator, such as personnel at TRANSCOM. In this type of scenario, the trainee is supporting a mission where there are no available appropriate vehicles to transport the casualties resulting from that mission. Before the mission, the trainee must create an initial plan for how to modify available resources to transport the types of casualties expected from the mission. This planning process requires an understanding of the capabilities of different assets in the available fleet of transport vehicles, an understanding of the types of casualties that may result from a particular mission, and an understanding of the resources required to care for that particular set of casualties. The trainee also needs the situational awareness to identify likely bottlenecks in care for the particular set of patients and to identify potentially unforeseen complications (e.g., casualties whose injuries will progress differently when transported at altitude vs. on the ground). As the mission evolves, the trainee must adapt the plan and coordinate allocation of resources. For instance, they may need to coordinate land- or water-based evacuation if aircraft become unavailable or if the combat team becomes pinned down such that air access is denied.

From these scenario archetypes, we identified the critical skills used by different personnel to carry out tasks in those scenarios. To carry out these tasks, personnel need to possess specific declarative knowledge, perceptual/situational awareness skills, communication skills, and judgement/decision-making skills. Though reporting the full model and analysis of skills used in this domain is outside the scope of this paper, in we present the key skill categories for the *Mission Management via Casualty Care* scenario archetype to illustrate the novel training requirements for this domain.

For each skill described in the inventory, we created a set of metrics defining when a skill was considered to have been successfully executed within a training session and when it was considered to have been mastered across training sessions. Again, we do not document the full skill model or these metrics, but we do describe the process by which they were constructed and the unique characteristics these metrics have as a result of the specific training domain.

Metrics can be described fairly simply in more traditional medical care settings. For example, in the trauma care domain, there are straightforward and well established means for assessing the proper performance of an intubation (checking for specific indications from a CO<sub>2</sub> monitor or chest X-ray), the insertion of a chest needle (observing the signature sound it makes or changes in the patient's chest rise), or the application of a tourniquet (identifying changes in distal pulses and the presence of new blood). Similar metrics exist for related domains such as casualty extraction, and triage. Defining metrics for the domain of far-forward, mass casualty medicine is more difficult because it is necessary to extract away potentially relevant details such as the enemy force, the type of attack and casualties, and the physical properties of the scene; as a result, there are no pre-defined gold standards and performance must nonetheless be estimated in scenarios that may have never been previously described.

We illustrate this process using two examples from our curriculum: assessment metrics defining how well a casualty collection point (CCP) is placed and how efficiently subordinate medical personnel are utilized. To objectively assess how well a Casualty Collection Point (CCP) was placed in an arbitrary physical setting, it is necessary to consider the properties that influence the quality of the CCP, such as the amount of cover the surroundings provide from fire, the capacity of the space to enable casualty placement and basic life support activities, and the accessibility of the space to evacuation vehicles. Given an arbitrary scene, an expert mass casualty response trainer can make a subjective but informed assessment of how well a CCP has been placed, but providing this assessment objectively and in the absence of an instructor requires defining a function that evaluates the different attributes that contribute to a CCP's quality. Similarly, assessing how efficiently a team's medical personnel are utilized requires assessing the outcomes that indicate productive or unproductive utilization, and this assessment is related to the type of environment in which the team operates. In an urban mass casualty event, first responders perform a few key activities including Basic Life Support and search and rescue. Therefore, an assessment metric for medical teams in this environment compares the benefit each team member provides, in terms of treating and rescuing casualties, versus the cost, in terms of the amount of time spent transitioning between tasks and not rendering aid to victims. Thus, a digital assessment metric considers the team member's initial position, time spent navigating the scene, number of casualties treated/extracted, and extent to which each casualty was stabilized. We implemented these metrics for CCP placement and personnel usage in the EFFECTIVE virtual simulation, and the other assessment metrics for this training curriculum are defined in terms of similar functions.

Finally, we defined instructional strategies for each skill in terms of the modality for training that skill, the type and characteristics of content presented in training, and the sequence of different instructional stages. We adhered to the KRK three-stage model for learning procedural skills, in which initially, purely declarative representations of the skill sequence are formed, and through repeated practice the skill is "proceduralized" until it can be done automatically with minimal conscious thought (Kim et al., 2013). For most skills in this category, our model recommends a general sequence in which first the declarative knowledge is trained through some traditional didactic medium, and the skill is proceduralized through repeated practice in progressively higher fidelity virtual training environments. A similar process applies to perceptual skills, such as recognizing the signs of an injury, where the trainee must first learn declarative knowledge underlying the relevant anatomical structures and then hone their perceptual discrimination by training on progressively more realistic, cluttered, and subtle signals. The choice of realism and fidelity in simulation is thus driven primarily by the stage of learning and the specific primary skills being trained. Depending on the nature of the skills being trained, different interface modalities are also preferable. For many of the procedural and decision-making skills of interest, screen-based simulation provides sufficient visual and functional fidelity to provide an effective learning environment. However, certain skills are more or less sensitive to the fidelity of the environment; for example, while a high-fidelity 3D environment has not been shown to be critical for training teamwork skills, it is

critical for training spatial and navigational skills (Hoang et al., 2012; Salcedo et al., 2013). Therefore, we recommend virtual- and augmented- reality as appropriate training modalities for front line personnel, such as ground medics, where a keen understanding of the spatial relationships between threats and care must be obtained.

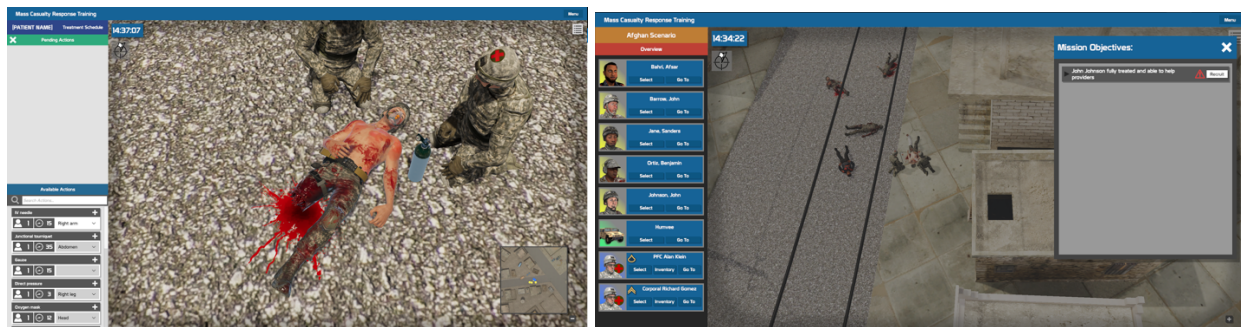
## DISCUSSION

Using the MAST instructional design and skill modeling process described above, we were able to produce a catalog of critical skills and a set of core scenario archetypes that established requirements for the design of EFECTIVE, a training simulator addressing skills used in the broad set of activities likely to be performed in support of austere, far-forward medical operations. This training system is intended to familiarize multiple personnel classes, including frontline personnel, logistics operators, and mission planners, with concepts related to mass casualty response and resource management, as well as provide a rehearsal environment for role-specific skills.

EFECTIVE separates different components of training into individual modules targeted at specific sets of skills and uses a unique user interface customized for each one; as such, the system is “mixed-reality” in that trainees experience different levels of fidelity and immersion as they move through modules for different skills. The MAST process directly influenced the selection of interfaces for different training modules by defining the type of skill needed to perform different tasks (e.g., whether execution of a task critically depends on a perceptual skill, procedural/motor skill, recall of declarative knowledge, decision-making, etc.). A significant body of existing research describes the tradeoffs in using different training modalities for various types of skills; as a result, we were able to remove much of the subjectivity in selecting the training interface for various tasks given our modeling of the operational domain.

Scenarios in the EFECTIVE simulation environment vary widely in scope depending on the goals of that particular training module. These can include, for example: (1) scenarios focused on trauma care, where the trainee performs medical diagnostics and interventions on a single virtual casualty; (2) scenarios focused on triage, where the trainee performs brief assessments and makes decisions about the relative acuity of a small set of casualties; (3) scenarios focused on scene management and resource allocation, where the trainee can observe an area the size of multiple city blocks, establish extraction and casualty collection points, direct the movement of the medical response team, and perform basic assignment of care plans to individual casualties; and (4) scenarios focused on high level operation management, where the trainee observes an area the size of a city, no longer can observe details of single patients, and directs the broad treatment response strategy including establishing perimeters to a hot zone, selecting evacuation and resupply routes, and brokering arrangements with local care facilities. To date, we have built modules addressing inter- and intra-team resource management, trauma management, triage, and search and rescue, and are in the process of adding additional components for training communication and evacuation route planning.

Figure 1 shows two different user interfaces from within the current training system. At the left is the viewpoint shown to trainees when they are training on triage and casualty care management. At the right is the viewpoint shown when the trainee is directing a team of personnel and responding to mission events in resource management training. Both views correspond to a computer-based training module.



**Figure 1: Two different views of the EFECTIVE simulation. Left: the trainee directs a team of simulated medical personnel to perform specific interventions on a casualty. Right: the trainee observes the scene from a high level, with multiple providers and casualties visible at once.**

The curriculum itself and the training platform on which it is implemented derive from the scenarios, the key skills, and the assessment metrics described above. While there are three main scenario categories, training itself is delivered using progressively increasing degrees of scope, starting with just small chunks of the scenarios defined above. For example, training for the first responders includes didactic training to address the core declarative knowledge, delivered in the form of text-, image-, and video-based tutors with electronic quizzes. After demonstrating mastery of the core declarative knowledge, these trainees move onto extremely low-scope scenarios—for example, practicing trauma management sequences on a single virtual patient to demonstrate their understanding of the primary and secondary trauma assessment algorithms. They then progress onto medium-scope scenarios, such as performing just the triage operation—e.g., responding to a “revolving” door of virtual patients, assigning them triage levels, and selecting an appropriate treatment plan. This training may be repeated with a variety of mission contexts presented, because in the version of austere medicine being trained the specific care doctrine depends on greater mission goals. Later, they move onto higher and higher scope scenarios capturing a greater portion of their duties.

In addition to role-specific training, the training system also includes a “cross-training” simulation, which is intended to familiarize trainees with the overall organization of incident response and how the different relevant components relate to each other. Here, rather than performing the training from a specific fixed viewpoint, the trainee experiences the scenario from the perspective of an omni-present entity that can perform a variety of roles at different levels of granularity. This viewpoint is not meant to be realistic to the experience of a particular operational role and is instead meant to provide a complete picture of how that trainee’s role interacts with the roles of other types of personnel. While logistics personnel, medical personnel, and personnel in more combat-oriented roles might have very different individual training modules, they complete the same “cross-training” modules from this single, omni-present viewpoint. These modules are sprinkled throughout training, and their scope increases in tandem with the scope of the individual modules.

Using the MAST process to define assessment metrics promotes the incorporation of automation, modularity, and extensibility into training. This is because a skill-centric approach to training design like MAST encourages the construction of assessment metrics that are targeted to skills but agnostic to the setting and scenario in which those skills are used. This means that a single piece of assessment logic can be used to drive skill assessment under a wide variety of settings and be reused across scenarios. The alternative task-centric approach to training design generally requires a unique assessment definition for each task that is performed in training, even when multiple tasks share one or more core skills. While this can provide detailed feedback on the performance of specific tasks, it lacks the generalizability of a skill-centric approach and precludes continuous monitoring of trainee skills across a range of behaviors. In EFECTIVE, assessment metrics for any skill can be evaluated whenever the data sources that drive those assessment metrics are available, meaning more abstract concepts like “resource allocation” can be continuously monitored even when a trainee is not taking any immediate actions directly related to resource allocation. This characteristic also makes training more adept to modification, because the logic for an assessment metric can be revised only once and still impact all the scenarios in which that skill is used. Currently, assessment metrics for EFECTIVE have been authored using JSON, XML, and Lua, which can be modified directly in text and reloaded into the training application without the need to formally recompile the software. While the current authoring capability is mainly accessible to software engineers and those with an understanding of the Lua logic and Json/XML data structures, ongoing work is addressing the development of a GUI that enables assessment metric authoring in strictly operational terms, abstracting away the computational representations.

In addition to enabling flexible, automated skill assessment, MAST simplifies the inclusion of content generation and AI-based agents into simulation. Content creation is made feasible by the aspecific definition of scenario concepts, which separate the general thematic elements of the scenarios from the particular physical properties that make them distinct. A number of different simulated environments can be used to depict a scenario, as long as that environment exposes the data sources necessary to assess the relevant skills for that scenario. Other scenario attributes, such as terrain, buildings, number and type of forces, weaponry, medical resources, casualty volumes, and specific mission objectives can be swapped in and out or generated on the fly. For example, we have created multiple instances of the mission management scenario class described above to provide training on the same skills in either a western or a middle eastern urban environment within completely different contexts. This is because environment and scenario-specific details are fully abstracted away from the mechanics that enable a trainee to perform and be assessed on their skills. Similarly, we have a logistics management training environment that can pull open source map data and set training in any arbitrary real location, as well as a search-and-rescue training environment that can be used equivalently in either virtual reality or on a computer display.

While MAST instructional design doesn't directly enable inclusion of AI agents into training, it does simplify the process for obtaining that functionality. The hierarchical definition of skill and task performance used by MAST closely aligns with the "behavior tree" structures typically used to author the behavior of virtual intelligent agents. Hap, a reactive planning architecture, is an example of such a framework with a hierarchical behavioral taxonomy analogous to that of MAST (Loyall & Bates, 1993; Reilly et al., 2010), and MAST models can be authored into Hap relatively simply. In EFECTIVE, because our MAST models captured the tasks performed by multiple different roles, we were able to author the behavior of AI agents which perform roles complementary to the trainee's into the training modules. For example, in one scenario in which the trainee participates as the incident commander, numerous AI teammates depicting individual first responders are present which perform trauma assessments in a way that reflects established best practices, as captured in the MAST model for those skills.

## CONCLUSIONS

The Methodology for Annotated Skill Trees (MAST) is a previously defined instructional design framework that we used to model the curriculum for a novel multi-role training system for austere, far-forward medical operations. The MAST process enabled us to define concrete skill requirements and assessment metrics for training while abstracting away notions of actual adversaries and operating environments. The resulting curriculum highlights the important properties of training while providing significant flexibility of actual training content, enabling training that is grounded in strong instructional design while being robust to evolution in the operating environment. Since the resulting training system has not yet been deployed, training effectiveness has not been formally established, and this will be a focus of future work.

## ACKNOWLEDGEMENTS

This work is supported by the US Army Medical Research and Materiel Command under Contract No. W81XWH-19-C-0044. The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other documentation.

## REFERENCES

- Agarwal, N., & Bandeli, K. (2018). Examining Strategic Integration of Social Disinformation Campaign Coordination. In *Defense Strategic Communications* (Vol. 4, pp. 173–206).
- Allen, W. C. (2006). Overview and Evolution of the ADDIE Training System. *Advances in Developing Human Resources*, 8(4), 430–441. <https://doi.org/10.1177/1523422306292942>
- Andronikos, G., Elumaro, A. I., Westbury, T., & Martindale, R. J. J. (2016). Relative age effect: implications for effective practice. *Journal of Sports Sciences*, 34(12), 1124–1131. <https://doi.org/10.1080/02640414.2015.1093647>
- Bauchwitz, B. R., Curley, T., Kwan, C., Niehaus, J. M., Pugh, C. M., & Weyhrauch, P. W. (2019). Modeling Framework Used to Analyze and Describe Junctional Tourniquet Skills. *Military Medicine*, 184(Supplement\_1), 347–360. <https://doi.org/10.1093/milmed/usy348>
- Bauchwitz, B. R., Niehaus, J. M., & Weyhrauch, P. W. (2018). Modeling and Comparing Nurse and Physician Trauma Assessment Skills. *Military Medicine*, 183(suppl\_1), 47–54. <https://doi.org/10.1093/milmed/usx139>
- Becker, T., & Link, M. (2011). Medical rules of engagement negative patients: the dilemma of forward surgical teams in counterinsurgency operations. *Journal of Special Operations Medicine: A Peer Reviewed Journal for SOF Medical Professionals*, 11(2), 12–15.
- Cao, C., Mackenzie, C., Ibbotson, J., Turner, L., Blair, N., & Nagy, A. (1999). Hierarchical decomposition of laparoscopic procedures. *Stud Health Technol Inform*, 62–83.
- Cao, C., MacKenzie, C., & Payandeh, S. (1996). Task and motion analyses in endoscopic surgery. *Proceedings ASME Dynamic Systems and Control Division*, 583–590.
- Cronk, T. (2018). *Near-Peer Adversaries Work to Surpass U.S.* In *Technology, Official Say*. US Department of Defense. <https://www.defense.gov/Explore/News/Article/Article/1512901/near-peer-adversaries-work-to-surpass-us-in-technology-official-says/>

- De Novi, G., Bardsley, R., Shah, R., Ottensmeyer, M., Moore, J., & Ahn, B. (2012). Event-driven Surgical Gesture Segmentation and Task Recognition for Ocular Trauma Simulation. *Intelligent Environments (Workshops)*, 341–352.
- Dian, M. (2015). The Pivot to Asia, Air-Sea Battle and contested commons in the Asia Pacific region. *The Pacific Review*, 28(2), 237–257. <https://doi.org/10.1080/09512748.2014.995124>
- Garcia, D. M. (2015). *Fighting in a Contested Space Environment: Training Marines for Operations with Degraded or Denied Space-Enabled Capabilities* [Master's Thesis]. NAVAL POSTGRADUATE SCHOOL MONTEREY CA. <https://apps.dtic.mil/docs/citations/ADA632458>
- Grosdemouge, C., Weyhrauch, P., Niehaus, J., Schwaitzberg, S., & Cao, C. G. L. (2013). Design of Training Protocol for Perceptual and Technical Skills in a Minimally Invasive Surgery. *Engineering Systems Design and Analysis*, 855–860. <https://doi.org/10.1115/ESDA2012-82869>
- Hoang, T., Kang, J., Laporta, A., Makler, V., & Chalut, C. (2012). Filling in the gaps of predeployment fleet surgical team training using a team-centered approach. - Abstract - Europe PMC. *Journal of Special Operations Medicine*, 13(4), 22–33.
- Johnson, R. A. (2015). Predicting future war. *Australian Defence Force Journal*, 44(196), 70.
- Keenan, S. (2007). Role of Medicine in Supporting Special Forces Counter-Insurgency Operations in Southern Afghanistan. *Journal of Special Operations Medicine*, 7(1), 48–50.
- Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science*, 14(1), 22–37. <https://doi.org/10.1080/1464536X.2011.573008>
- Lang, N. M. (2016). *Distributed Mission Operations: Training Today's Warfighters for Tomorrow's Conflicts* [Technical Report]. AIR COMMAND AND STAFF COLLEGE, AIR UNIVERSITY MAXWELL AFB United States. <https://apps.dtic.mil/docs/citations/AD1041084>
- Loyall, A. B., & Bates, J. (1993). Real-time Control of Animated Broad Agents. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, 664–669.
- Mackel, T., Rosen, J., & Pugh, C. (2007). Application of hidden markov modeling to objective medical skill evaluation. *Stud Health Technol Inform*, 316–318.
- MacKenzie, L., Ibbotson, J., Cao, C., & Lomax, A. (2001). Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. *Minimally Invasive Therapy & Allied Technologies*, 10(3), 121–127.
- Middleton, V., Weyhrauch, P., Ritter, F., & Dancy, C. (2018). *Dynamic Representation for Evaluating the Effect of Moderators and Stress on Performance (DREEMS)*. 2018 Fall Simulation Innovation Workshop.
- Morris, T. S., VanDriel, M. S., Dries, W. D., Perdew, J. C., Schulz, R. H., & Jacobsen, K. E. (2015). Securing Operational Access. *The National Interest*, 136, 69–73. JSTOR.
- Morrison, G. R., Ross, S. J., Morrison, J. R., & Kalman, H. K. (2019). *Designing Effective Instruction*. John Wiley & Sons.
- Niehaus, J., Romero, V., Koelle, D., Palmon, N., Bracken, B., Pfautz, J., Reilly, S. N., & Weyhrauch, P. (2014). A Flexible Framework for the Creation of Narrative-Centered Tools. In M. A. Finlayson, J. C. Meister, & E. G. Bruneau (Eds.), *2014 Workshop on Computational Models of Narrative* (Vol. 41, pp. 130–138). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. <https://doi.org/10.4230/OASICS.CMN.2014.130>
- Ottensmeyer, M., Ahn, B., Moore, J., Munteneanu, E., Iorino, S., De Novi, G., Bardsley, R., Shah, R., Neumann, P., & Langston, D. (2012). Ocular and Craniofacial Trauma Treatment Training System: Overview & Eyelid Laceration Module. *Intelligent Environments (Workshops)*, 319–330.
- Perez, R. S., Skinner, A., Weyhrauch, P., Niehaus, J., Lathan, C., Schwaitzberg, S. D., & Cao, C. G. L. (2013). Prevention of Surgical Skill Decay. *Military Medicine*, 178(suppl\_10), 76–86. <https://doi.org/10.7205/MILMED-D-13-00216>
- Reilly, S. N., Niehaus, J., & Weyhrauch, P. (2010). Modeling Ground Soldier situational awareness for constructive simulation with rules. *Proceedings of the 2010 Winter Simulation Conference*, 1431–1443. <https://doi.org/10.1109/WSC.2010.5679050>
- Reznek, M., Smith-Coggins, R., Howard, S., Kiran, K., Harter, P., Sowb, Y., Gaba, D., & Krummel, T. (2003). Emergency Medicine Crisis Resource Management (EMCRM): Pilot Study of a Simulation-based Crisis Management Course for Emergency Medicine. *Academic Emergency Medicine*, 10(4), 386–389. <https://doi.org/10.1111/j.1553-2712.2003.tb01354.x>
- Rickel, J., & Johnson, W. (1999). Virtual humans for team training in virtual reality. *Proceedings of the Ninth International Conference on Artificial Intelligence in Education*, 578, 585.

- Risser, D. T., Rice, M. M., Salisbury, M. L., Simon, R., Jay, G. D., & Berns, S. D. (1999). The Potential for Improved Teamwork to Reduce Medical Errors in the Emergency Department. *Annals of Emergency Medicine*, 34(3), 373–383. [https://doi.org/10.1016/S0196-0644\(99\)70134-4](https://doi.org/10.1016/S0196-0644(99)70134-4)
- Salcedo, J., Maraj, C., Lackey, S., Ortiz, E., Hudson, I., & Martinez, J. (2013). Effects of Visual Fidelity on Biometric Cue Detection in Virtual Combat Profiling Training. In R. Shumaker (Ed.), *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments* (pp. 388–396). Springer. [https://doi.org/10.1007/978-3-642-39405-8\\_43](https://doi.org/10.1007/978-3-642-39405-8_43)
- Satava, R., & Jones, S. (1997). Virtual Environments for Medical Training and Education. *Presence: Teleoperators & Virtual Environments*, 6(2), 139–146.
- Seeber, I., Bittner, E., Briggs, R., De Vreede, T., De Vreede, G., Elkins, A., Maier, R., Merz, A., Oeste-Reib, S., Randrup, N., & Schwabe, G. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 1(57), 103–174.
- Tanner, C. A. (2006). Thinking Like a Nurse: A Research-Based Model of Clinical Judgment in Nursing. *Journal of Nursing Education*, 45(6), 204–211. <https://doi.org/10.3928/01484834-20060601-04>
- Thomas, T. (2014). Russia's Information Warfare Strategy: Can the Nation Cope in Future Conflicts? *The Journal of Slavic Military Studies*, 27(1), 101–130. <https://doi.org/10.1080/13518046.2014.874845>