Developing Criteria to Compare Military Medical Trauma Simulations Across Modalities

Shannon Bailey, Michael Brannick, Colleen Reiner, & Luis Llerena
Center for Advanced Medical Learning & Simulation, University of South Florida
Tampa, FL
shannonbailey@usf.edu, mbrannick@usf.edu, reiner@usf.edu, luisllerena@usf.edu

SGM (Retired) F. Bowling
Apogee, Engineering, LLC
Tampa, FL
f.y.bowling.ctr@socom.mil

Dennis Lyons
US Army (Retired); Smith and Nephew, Inc
Whittier, NC
dennis.lyons@smith-nephew.com

Samantha Tromly
Institute of Applied Engineering, University of South Florida
Tampa, FL
stromly@usf.edu

ABSTRACT

The military trains medics on trauma management procedures using various methods and modalities of simulation. Simulation modalities for medical training may include manikins, part-task trainers, augmented/virtual reality, computer-based simulations, cadavers, or live animals. Each modality has benefits and limits, such as suitability for training a certain task, cost, scalability and deployability. To choose the right simulation, it is important to examine modality benefits and limits to train specific medical tasks, and to understand the capability gaps for the given training context. To establish whether a simulation meets the needs of military medical training, evaluation criteria were developed for twelve trauma procedures based on existing military doctrine. Military doctrine was converted to a checklist of steps needed to complete each trauma procedure successfully. The checklists provided data through a pilot study conducted as part of a military Technical Experimentation (TE) event. The TE event included 27 medical simulations of different types presented by industry developers, evaluated by 33 experienced military medics. Following the checklists, the medics rated simulations on factors of suitability for military medical training. Checklists and suitability ratings indicated that the capabilities of simulators varied widely and that the benefits and limitations of simulation modalities vary both between and within modalities. The medics also provided evaluation methodology feedback that will be used in future comparisons of simulations. Future simulation evaluation will provide additional guidance for medical training that meets operational needs.

ABOUT THE AUTHORS

Shannon Bailey, PhD, is a Sr. Human Factors Scientist at the University of South Florida's (USF) Center for Advanced Medical Learning and Simulation, and an Assistant Professor in the USF Morsani College of Medicine's Department of Medical Education. She specializes in human factors and extended reality technologies, leading research efforts to investigate educational best practices and cutting-edge simulation technologies for clinical training.

Michael Brannick, PhD, is a psychometrician at the USF Health Center for Advanced Medical Learning and Simulation and professor emeritus in the Psychology Department at the University of South Florida. He is known for work on medical training program evaluation and for research on methods of data analysis.

SGM (Retired) F. Bowling is a Biomedical subject matter expert (SME) who provides contract support to USSOCOM SOF AT&L-ST. During his 29 years in the U.S. Army, he served on many boards, committees, and panels including the Defense Health Board Committee on Tactical Combat Casualty Care, USSOCOM Special Operations Combat Medic Curriculum & Examination Board, the U.S. Army Medical Research and Materiel Command Telemedicine & Advanced Technology Research Center Joint Program Committee-1 (Steering Committee for Medical Simulation and Training), the NATO Technical Team Member for Human Factors and Medicine Panel 242 (HMF-242)-Technology Alternatives for Medical Training: Minimizing Live Tissue Use, USASOC/USAJFKSWCS

Institutional Animal Care and Use Committee, and the NATO Special Operations Forces Tactical Medicine Panel (SOFTMP).

Colleen Reiner is the Director of Education & Research Administration for USF Health Center for Advanced Medical Learning and Simulation (CAMLS). Her career has focused on developing research-driven, educational ecosystems designed to reduce the lag between scientific discovery and the effective dissemination of results, and she has over 20 years of experience leading instructional design and multimedia technology teams in higher education.

Luis Llerena, MD, is the Medical Director of USF Health CAMLS and Professor, Department of Surgery at the USF Morsani College of Medicine. He has over 23 years of trauma experience: Surgical Director of the American College of Surgeons (ACS) Accredited Educational Institute (AEI) at CAMLS, ACS AEI Committees (Accreditation Review and Curriculum), Florida Committee of Trauma (FCOT) Vice Chair, and Florida Trauma State Faculty. He is a course director for Advanced Trauma Life Support (ATLS®), Advanced Trauma Operative Management (ATOM®), Advanced Surgical Skills for Exposure in Trauma (ASSET®), Stop the Bleed®, and the Fundamental Critical Care Support (FCCS) program.

Dennis Lyons has 20+ years active duty with 14 years as a Special Operations Combat Medic and retired in 2015. He has multiple deployments to both Iraq and Afghanistan, along with additional non-combat deployments. He has provided direct patient care at all roles in combat and garrison. Since retiring in 2015, he has worked as an operational medicine consultant doing both business and product development for distributors and manufacturers.

Samantha Tromly is the Director of Programs at USF's Institute of Applied Engineering where she is responsible for successful execution of programs totaling $34M on behalf of United States Special Operations Command (USSOCOM), United States Central Command (USCENTCOM), the 6th Air Refueling Wing at MacDill Air Force Base, NASA, SOFWERX, and others.

Developing Criteria to Compare Military Medical Trauma Simulations Across Modalities

Shannon Bailey, Michael Brannick, Colleen Reiner, & Luis Llerena	SGM (Retired) F. Bowling	Dennis Lyons	Samantha Tromly
Center for Advanced Medical Learning & Simulation, University of South Florida	Apogee, Engineering, LLC	US Army (Retired); Smith and Nephew, Inc	Institute of Applied Engineering, University of South Florida
Tampa, FL	Tampa, FL	Whittier, NC	Tampa, FL
Shannonbailey@usf.edu, mbrannick@usf.edu, reiner@usf.edu, luisllerena@usf.edu	f.y.bowling.ctr@socom.mil	dennis.lyons@smith-nephew.com	stromly@usf.edu

INTRODUCTION

Combat medics must learn to perform life-saving medical procedures under difficult conditions, such as in active combat, austere or challenging environments, and complex multi-casualty situations (Suresh et al., 2023). To prepare for combat casualty care, simulation training is used to approximate the medical emergencies and environmental conditions medics may face in the field (Suresh et al., 2021). Training for combat medics focuses on learning the evidence-based best practices for prehospital trauma while considering the tactical situation. Since the introduction of Tactical Combat Casualty Care (TCCC) in the 1990s, preventable combat deaths have significantly decreased, and the TCCC standards continuously improve as medical evidence advances (Butler, 2017). Yet, a recent scoping review of combat casualty care training highlighted the need for continued efforts to improve efficiency and efficacy of care to increase survival rates (Knisely et al., 2022). Further, recent surveys of combat medics indicated a desire for more training opportunities pre-deployment and between deployments (Suresh et al., 2021; Suresh et al., 2023). Training that focuses on injury is crucial to maintain skills during times of relative peace to reduce the "peacetime effect," or the degradation of skills during times of lower combat intensity that have historically led to proportionally worse outcomes for early casualties in a new conflict (Cannon, Gross, & Rasmussen, 2021).

Simulation training as defined by the Healthcare Simulation Dictionary (Lioce 2020) is a "technique that creates a situation or environment to allow persons to experience a representation of a real event for the purpose of practice, learning, evaluation, testing, or to gain understanding of systems or human actions." Simulation training has been used to meet the varied needs of combat medical training across a range of methods and technologies. In combat casualty care, simulation can be considered any method of practice that is outside of active patient care. The modalities of medical simulation training may include manikins, standardized patients (i.e., actors or role players), part-task trainers, computer-based simulations (i.e., serious games), augmented/virtual/mixed reality, cadavers, and live tissue training.

Simulation modalities are not "one-size-fits-all" for medical training. To employ the appropriate simulation modality or blend of modalities for specific training content and context, the capabilities of each simulation modality must be considered. The capabilities of modalities vary widely in terms of sensory realism, anatomical and physiological fidelity, durability, deployability, reusability, and cost. A recent technical report reviewed existing medical simulation technologies for TCCC as of 2021, cautioning that "although advances have been made in recent years in both simulation and immersive technologies, a single environment, linking realism, haptics and emotional connection remains elusive" (North Atlantic Treaty Organization [NATO], 2021, pp. 3-11). The report concluded that there is not yet conclusive evidence that a particular modality is superior, and options should be considered to blend modalities taking into consideration available resources (pp. 3-9, 3-11). Further, the report identified needs to benchmark standards for simulation modalities and to develop guidelines for integrating simulation modalities into training curricula.

This study is an exploratory evaluation of capabilities and limitations of medical simulation modalities for combat medical training. It was intended to address the problem of lack of standardized methods for evaluating simulation technologies in medical training. The study was comprised of three phases, whose corresponding aims were to:

1. Develop evaluation criteria to determine capabilities and limitations of medical simulation modalities.
2. Conduct a study in which experienced combat medics evaluated simulation modalities presented by developers at a Technical Experimentation (TE) event.
3. Compare simulation capabilities both within and between modalities by statistically analyzing ratings of simulations made by experienced medics. That is, we aimed to determine whether there are distinguishable differences between simulations within modalities across tasks as well as general differences between modalities.

Evaluation Criteria Development

To achieve the first aim, evaluations were developed to quantify the degree to which simulators could provide training for a specific task and/or task step:

1. Section A: Task Analyses of Simulation Capabilities – a step-by-step evaluation of a simulation's capabilities for each of 12 medical tasks. The analyses resulted in a tally of capabilities for each simulation such that scores could be compared across simulations and across medical tasks.
2. Section B: Subjective Evaluation by Task – subjective ratings of simulations for each medical task such that subjective ratings could be compared across simulation modalities.

Section A: Task Analyses of Simulation Capabilities

The task analysis evaluations (Section A) were developed to assess the capabilities of a simulator for each of 12 combat casualty care tasks (Table 1). These tasks were chosen for inclusion in the study based on the need for simulation training to increase combat casualty care efficiency to reduce preventable combat fatalities as identified by the Committee on Tactical Combat Casualty Care.

Table 1. List of Combat Casualty Care Tasks Included in Current Study

Task		Label
1	Chest Tube	Chest Tube
2	Cricothyroidotomy (Open Surgical)	Cric
3	Extraglottic Airway (EGA)	EGA
4	Head-Tilt/Chin-Lift	Head Tilt
5	Jaw-Thrust Maneuver	Jaw Thrust
6	Nasopharyngeal Airway (NPA) Insertion	NPA
7	Needle Chest Decompression	Needle
8	Recovery Position	Recovery
9	Tactical Trauma Assessment	Trauma
10	Tourniquet Care Under Fire (CUF)	CUF
11	Tourniquet Tactical Field Care (TFC)	TFC
12	Wound Packing and Pressure Bandage	Wound

A modified Cognitive Task Analysis (CTA) approach was used to create the capability evaluations for each of the 12 tasks in Section A. CTA is a technique to identify cues, or simulation features, that are critical for successful task performance (Cannon-Bowers et al., 2013). For example, if a medical task requires identification of body landmarks to place a needle in the correct location, body landmarks are a cue necessary to complete the simulated task. Without this cue, the simulation cannot be used to evaluate the trainee's knowledge and correct performance. CTA can thus be used to create a set of standards for features that should be included in a simulation to either train or assess understanding of a task. To define the needs of simulation-based training for a specific medical task, Cannon-Bowers et al. (2013) recommended analyzing the following for each step of a task to develop a simulation:

1. Cues to perform each step
2. What is missing from a simulation to perform each step
3. Expected trainee errors for each step
4. Observable trainee behaviors related to performance or ability on each step
5. Key cognitive decision-making components of each step

These categories were adapted in the current study for the purpose of quantifying the capabilities of existing simulation technologies. During the study, the following information was reported by the combat medics (i.e., "evaluators") for each task presented by the developers during the TE event (Table 2):

Table 2. Adapted CTA Categories for Current Study

	Current Study Task Analyses (Section A)	**CTA Component**
1	All steps to perform a task were included in a simulation ("Included")	Identify what is missing from a simulation
2	Critical steps were included in a simulation ("C")	Identify what is missing from a simulation
3	A simulation allowed errors on a critical step ("Allows Error")	Identify whether a simulation allows for expected trainee errors
4	Sensory features used as a cue to perform a step or to inform decision-making on a step were included in a simulation (Visual "V", Haptic "H", Auditory "A")	Identify whether a simulation includes cues or key cognitive decision-making components
5	Sensory feedback on performance for a critical step was included in a simulation, if applicable ("Feedback")	Identify whether a simulation includes cues or key cognitive decision-making components

Each task analysis evaluation in Section A differed based on the steps required for each task, so a total of 12 task analysis checklists were created. First, the steps for each of the 12 tasks were listed from the existing military doctrine (i.e., TCCC Assessments and Soldier's Manual and Trainer's Guide, MOS 68W Health Care Specialist, STP 8-68W13-SM-TG), and steps marked as critical in the doctrine were noted. The checklist listed each step in performance order, and evaluators were asked to mark whether the simulator included each step. Then, for each critical step, evaluators marked whether the simulation allowed a trainee to make an error on that step, because if a simulation does not allow for expected trainee errors, then the simulation is not capable of showing a trainee's level of proficiency on that step. For example, if a virtual reality simulation does not allow a trainee to place a needle in an incorrect location, then the simulation is not allowing a typical trainee error and is missing a key cognitive decision-making component necessary to train or assess performance on placing a needle.

With each critical step, we also listed sensory features that are typical cues needed to make a decision or perform a step. Sensory features included important visual, haptic, or auditory information about a step. For example, to perform a surgical cricothyroidotomy, a critical step is identifying the cricothyroid membrane between the cricoid and thyroid cartilages. Cues for correct performance of this step include seeing landmarks (Visual) and feeling tissue shift (Haptic). The sensory cues for each critical step were determined by a subject matter expert in combat casualty care as well as features identified as important from prior research in which 376 combat medics rated the importance of features of medical simulators (Horn et al., 2017).

For critical steps in which feedback was necessary to determine whether the action was correct or whether an expected trainee error occurred, we asked whether sensory feedback was included in the simulation. For example, in the surgical cricothyroidotomy task, for the trainee to know whether they correctly inserted the tracheostomy tube, they should see equal rise and fall of the chest. This is a form of feedback indicating correct or incorrect action on that step, a key cognitive decision-making component that would inform the next step. The analysis was quantitative in nature as it annotated whether the feedback was present but did not measure the quality or realism of the feedback.

Section B: Subjective Evaluation by Task

The subjective evaluations by task (Section B) were developed to record evaluators' subjective views of each simulation presented by developers, including their observations on the acceptability of a simulation for each task, the realism of different features of the simulation, the reusability and deployability of the simulation, and other subjective ratings (Table 3). Unlike the Section A checklists, which were unique to each task, Section B questions were identical across tasks. The questions were created based on the best practices literature and mastery learning principles (Ericsson 2004; Furman et al., 2010; McGaghie 2015; McMurray et al., 2021; Perretta et al., 2020). For example, to support the mastery learning principles of performing deliberate practice of a task multiple times with feedback, we asked whether a simulator was reusable or repeatable, and whether as simulator provided performance feedback to trainees. Evaluators rated their agreement with 14 items on a Likert scale from 1 ("Strongly Disagree") to 5 ("Strongly Agree"). The Section B items are shown in Table 3.

Table 3. Likert-scale Items in Section B with Labels

	Section B: Likert-scale Survey Items	**Label**	**Subscale**
1	The simulation is acceptable for training and evaluation.	Acceptable	General
2	The simulation allows accurate demonstration of all the critical steps in the task.	Accurate	General
3	The visuals in the simulation were realistic.	Realistic Visual	Realism
4	The haptic feel (touch) in the simulation was realistic. (If no haptics, mark "Strongly Disagree")	Realistic Haptic	Realism
5	The sounds in the simulation were realistic. (If no sound, mark "Strongly Disagree")	Realistic Audio	Realism
6	The simulation required realistic motor movements (e.g., holding instrument using correct grip).	Realistic Movement	Realism
7	The simulation appears capable of producing an emotive response.	Emotive	Realism
8	The simulation can be widely deployed.	Deployable	Deploy
9	The simulation is reusable or repeatable.	Reusable	Deploy
10	The simulation is durable enough to be used in a field environment.	Durable	Deploy
11	The simulation rapidly assessed a trainee's skill level.	Rapid Assessment	Measurement
12	The simulation accurately measured the trainee's psychomotor performance on the task.	Psychomotor Assessment	Measurement
13	The simulation accurately measured the trainee's cognitive decision-making on the task.	Cognitive Assessment	Measurement
14	The simulation accurately measured the level of learner engagement.	Engagement	Measurement

METHOD

In phase 2, the data collection for this study was combined with a TE event, allowing combat medics to inspect and evaluate a variety of available simulation technologies from industry developers. For the TE Event and study, the developers were given a list of 12 combat casualty care medical tasks and chose which task(s) to simulate for evaluation. Some developers chose to present simulations for multiple tasks, while others chose to present a simulation that was not on the task list with the understanding that the simulation would not be included in the study. Further, medics were instructed to evaluate the current capability of a simulation, not a potential future capability.

Participants

Experienced medics (*n* = 33) were recruited to evaluate simulation modalities during a TE event. Participants were active-duty military, retired, or reservists from the Army, Navy, Air Force, and Marines, with an average of 13.11 years of military service (*SD* = 5.86 years, Min. = 3 years, Max. = 25 years) and an average of 4.45 combat deployments (*SD* = 4.40). All participants had medical experience (100%), with an average of 11.92 years (*SD* = 6.91 years, Min. = 3 years, Max. = 30 years). Most participants were medical instructors (76%, *n* = 25), averaging 4.52 years of instructing (*SD* = 4.80 years, Min. = 0, Max. = 20 years). A majority of participants had experience treating live casualties in a combat environment (55%, *n* = 18). The participants were between 21 and 52 years old (*M* = 34.85 years, *SD* = 7.74 years), and all were male (100%).

Procedure

The study was structured to coincide with a 5-day TE event. To allow participants adequate time with each developer, participants were divided into four smaller groups of 8-9 participants: Groups A, B, C, and D. Participants completed a demographic survey, which was used to balance the experience of the smaller groups such that all groups had similar years of medical experience and that a mix of military branches were represented in each group. The average years of medical experience were similar in each group (Table 4).

During the 5-day TE event, each of the four evaluator groups met as a group with each industry developer for 45 minutes so that each developer had equivalent time to present their medical simulation(s) for evaluation. There were four simultaneous sessions each hour, with one developer demonstrating to a group of evaluators (A, B, C, and D). After each 45-minute session with a developer presenting their task(s), there was a 15-minute interval between sessions to allow developers to reset before the next group of evaluators. Following the 15-minute reset, the groups of evaluators switched rooms to evaluate the next developer. This repeated until every group of evaluators had reviewed each simulation by industry developers. There were 27 simulations presented by industry developers, though not every simulation was evaluated based on the developer's decision to present a briefing or task not listed for evaluation.

Table 4. Average Years of Medical Experience of Participants by Evaluator Group

Evaluator Group	*N*	***M*** years	***SD*** years
A	9	10.67	8.34
B	8	13.25	7.98
C	8	11.31	7.69
D	8	12.63	7.97

RESULTS

In phase 3 of the study, data from the evaluations were analyzed. Descriptive and inferential statistics are reported for the quantitative items in each of the evaluation sections. An *a priori* determination was made not to include simulations receiving fewer than seven responses in statistical analyses. This is because a low response rate could impact statistical accuracy (i.e., response bias, too small sample to use specific statistical tests). Furthermore, due to limited space to report results, we focus on the analyses of capabilities collapsed across tasks. Contact the authors for further analyses of simulator capabilities within a specific task.

Section A: Task Analyses of Simulation Capabilities

ANOVAs were conducted on the percentage of steps and sensory features included in simulations collapsed across modality type and task (Table 5). As can be seen in the graph (Figure 1), there were significant differences among capabilities corresponding to simulation modalities. Some modalities had more variance (See +/- 1 SE bars). For example, VR had greater standard error than LTT. The difference in variance can be explained in part due to some simulations having fewer evaluation responses than others with greater variance among those evaluations (e.g., VR), while other simulations had more responses that were relatively consistent (e.g., LTT). It is notable that LTT had a ceiling effect such that most evaluations were at the high end of the scale with little variance in responses, indicating more capabilities were available in LTT simulations than other modalities. High fidelity manikins and cadavers were the next highest rated capabilities, with low fidelity manikins combined with high fidelity task trainers following in the middle of the scale. Other modalities, including AR and VR, had fewer overall capabilities as rated by evaluators. Modalities that did not have enough responses (e.g., moulage, role player) were not included in the graph.

At least one, and sometimes several, simulations showed at least 90% of critical steps were included for the following tasks: Chest tube, Cricothyroidotomy, CUF, Head tilt, Needle chest decompression, NPA, Recovery position, TFC, Trauma assessment, and Wound packing. EGA and Jaw thrust did not achieve any simulations with at least 90% of the critical steps included. Note that these tasks were the only ones that did not show statistically significant differences between simulations in total steps included. Although LTT simulations showed the overall highest mean percentage of steps included, for most tasks, other modalities also achieved mean percentages of critical steps greater than 90. For the sensory tallies (haptic, visual, and auditory cues), LTT simulations consistently showed greater than 90% step inclusion except for Needle chest decompression (visual and auditory percentages in the 80s). For some tasks (e.g., CUF), other modalities also showed 90% or better step inclusion. However, for some tasks (e.g., Cricothyroidotomy), only LTT showed 90% or better step inclusion. Table 5 shows statistical tests collapsed across tasks. To conserve space, statistical tests of differences in capabilities by modality within task are not shown.

Table 5. Percentage of Steps and Sensory Features Included in Simulations Collapsed Across Modality and Task

	Source	***df***	***Sum Sq***	***Mean Sq***	***F***	***p***
Critical Steps	Modality	5	35.29	7.058	117	<.001
	Residuals	966	58.25	0.06		
Allows Errors	Modality	5	37.53	7.506	131.8	<.001
	Residuals	734	41.79	0.057		
Visual Realism	Modality	5	47.87	9.575	97.17	<.001
	Residuals	932	91.83	0.099		
Auditory Realism	Modality	5	32.72	6.544	41.18	<.001
	Residuals	564	89.63	0.159		
Haptic Realism	Modality	5	47.63	9.526	92.37	<.001
	Residuals	763	78.69	0.103		
Non-critical Steps	Modality	5	36.34	7.267	99.25	<.001
	Residuals	944	69.12	0.073		
Total Steps	Modality	5	34.31	6.862	119.5	<.001
	Residuals	897	51.52	0.057		

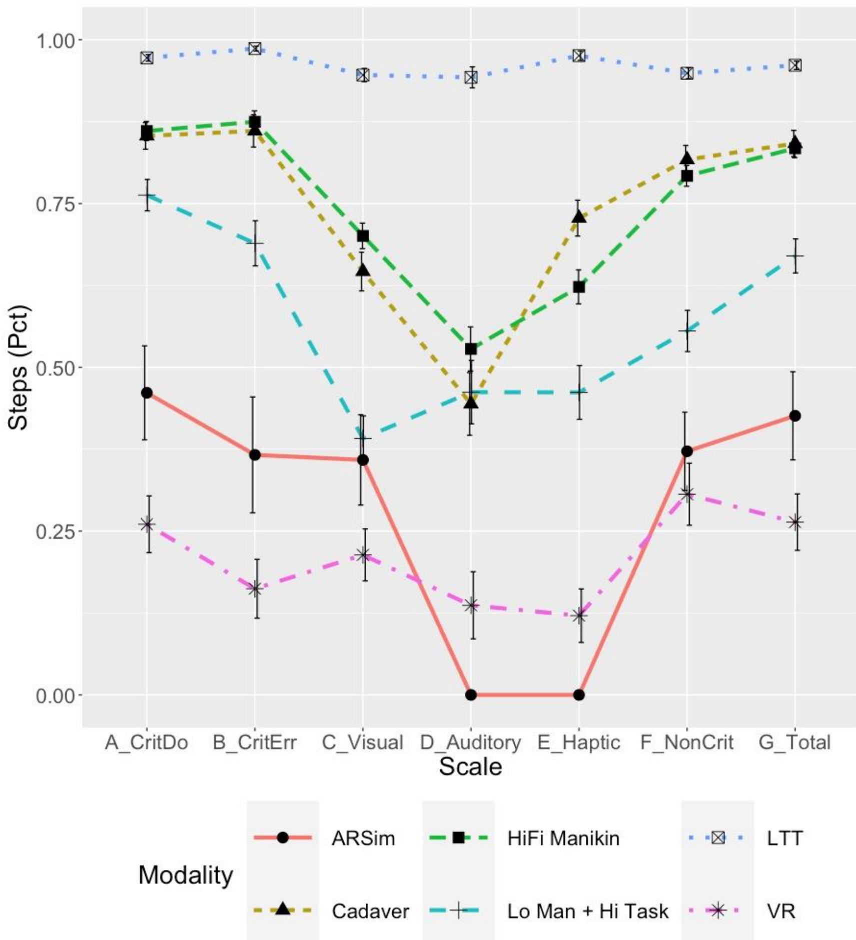

Figure 1. Percentage of critical steps (A), allowing critical errors (B), sensory features (C, D, E), noncritical steps (F), and total steps (G) included in simulations collapsed across modality type and task. Number of responses for each modality differ depending on the developer evaluated and number of simulators in each modality. Error bars represent +/- 1 SE.

Section B: Subjective Evaluation by Task

ANOVAs were conducted (Table 6) on the average responses to combined items in Section B collapsed across modalities on the four subscales (i.e., General, Realism, Deploy, and Measurement; See Table 3 for a list of items in each subscale). As can be seen in the graph (Figure 2), there were significant differences among subjective ratings on items corresponding with general acceptability of the modality for training (General), the realism of the simulations (Realism), the deployability and reusability of modalities (Deploy), and the ability of simulations to measure training outcomes (Measurement). As with the capabilities analysis in Section A, some modalities had more variance on items in Section B (See +/- 1 SE bars). For example, VR and AR had greater standard error than LTT. The AR and VR simulations had fewer evaluation responses than other modalities with greater variance among those evaluations. LTT again had a ceiling effect such that most evaluations were at the high end of the scale with little variance in responses, indicating higher subjective rating scores in LTT simulations than other modalities. High fidelity manikins, cadaveric simulation, and low fidelity manikins combined with high fidelity task trainers were the next highest rated capabilities, averaging slightly above the middle of the scale. AR and VR had lower average subjective ratings overall. Modalities that did not have enough responses (e.g., moulage, role player) were not included in the graph. Descriptively, some simulations were rated more highly than others, and the evaluations generally followed the same pattern as those in Section A. For example, the LTT simulations usually showed the highest means, while the AR and VR simulations showed the lowest.

Table 6. ANOVAs for Section B Attributes Collapsed Across Modality and Task

	Source	*df*	*Sum Sq*	*Mean Sq*	*F*	*p*
General	Modality	5	654.9	131	186.1	<.001
	Residuals	1073	755.3	0.7		
Reality	Modality	5	651.9	130.4	187.4	<.001
	Residuals	1056	734.9	0.7		
Deployment	Modality	5	252.7	50.53	52.85	<.001
	Residuals	1058	1011.7	0.96		
Measurement	Modality	5	572.8	114.55	154.8	<.001
	Residuals	1053	779.4	0.74		

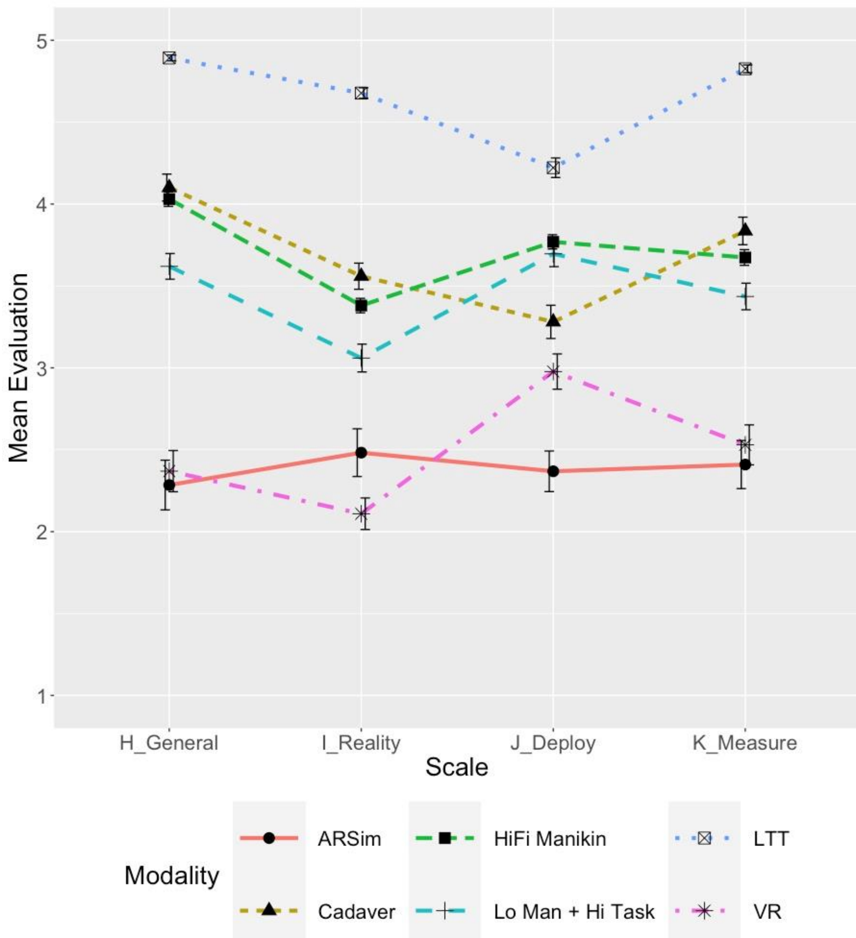

Figure 2. Average subjective ratings collapsed across modality type and task on four subscales in Section B: General (H), Realism (I), Deploy (J), and Measurement (K). 5 indicates high agreement on the subscale items, and 1 indicates low agreement. Number of responses for each modality differ depending on the developer evaluated and number of simulators in each modality. Error bars represent +/- 1 SE.

DISCUSSION

An exploratory study was conducted to serve as a foundation for evaluating and comparing medical simulator capabilities across modalities for combat medic training. During a weeklong TE event, 33 medics evaluated 27 simulations for 12 medical tasks presented by industry developers. In Section A, evaluators completed checklists of steps and sensory features that were included in a simulation that corresponded to the steps to perform a specific medical task. This evaluation allowed for comparison of the capabilities across simulations to determine whether a simulation has the necessary capabilities to train these tasks. Section B asked evaluators to rate the simulations presented by developers on Likert-type scales for capabilities and attributes, such as whether the simulation was realistic for a specific medical task. These ratings allowed for comparison of the subjective quality of the simulations and the perceived appropriateness of modalities for training across medical tasks.

Section A analyses revealed that there were significant differences in the capabilities of simulation modalities for most tasks. For most of the 12 medical tasks, at least one and sometimes several simulations had at least 90% of critical steps; however, some tasks (i.e., EGA and Jaw Thrust) did not have any simulations presented with at least 90% of critical steps included. Although LTT simulations showed the overall highest mean percentage of steps included for most tasks, other modalities also achieved mean percentages of critical steps greater than 90 for certain tasks. For sensory features (i.e., haptic, auditory, and visual cues) included in the simulations, LTT simulations consistently had greater than 90% inclusion, except for Needle Chest Decompression. Other modalities achieved greater than 90% sensory feature inclusion for some tasks (e.g., CUF), but for other tasks, only LTT had more than 90% sensory feature inclusion (e.g., Surgical Cricothyroidotomy). When collapsed across all medical tasks, more capabilities were available in LTT simulations than other modalities, but high-fidelity manikins and cadaver modalities were the next highest rated capabilities. Low fidelity manikins combined with high fidelity task trainers were in the middle of the scale, and AR and VR included fewer overall capabilities.

Section B results indicated there were significant differences among subjective ratings for the general acceptability of simulations for training medical tasks, the realism of the simulations, the deployability and reusability of simulations, and the ability of simulations to measure training outcomes. Following a similar pattern to Section A, some simulations were rated more highly than others. For example, the LTT simulations usually showed the highest subjective ratings. High fidelity manikins, cadaveric simulation, and low fidelity manikins combined with high fidelity task trainers had the next highest subjective ratings generally. AR and VR had lower average subjective ratings. For some attributes, ratings were inversely related. For example, a higher rating for realism tended to mean less reusability, deployability, or availability. The context for training is important because simulator fidelity is often a tradeoff with durability and practicality.

Limitations and Future Research

There were several limitations of this pilot study due to the exploratory nature of the research and the constraints of working within an existing TE event. Because the study coincided with a TE event in which developers could choose the content of their presentations, there was limited experimental control over the simulations presented for evaluation. Some developers chose to present many of the 12 medical tasks included in the evaluations, while other developers chose to simulate a task not under evaluation or to present a briefing instead of a simulation. Additionally, although attempts were made to provide consistent time with each developer, 45 minutes for each developer was perceived as either too much or too little time depending on the simulation(s) presented. For some developers, 45 minutes was not enough time to complete the checklist evaluations, which were notably long, and evaluators expressed wanting more time with some developers to complete evaluations. Other developers chose a simple task to simulate or did not present a task for evaluation, so all 45 minutes were not utilized. A tradeoff with this study was that in order to obtain the detail in the evaluations needed to compare capabilities, the evaluations were long, and the evaluators experienced evaluation fatigue by the end of the TE event. There was a modest sample size overall (n = 33), but response rates were lower due to evaluation fatigue, which varied depending on the developer. Shorter evaluations may have yielded a higher response rate, though this would also result in less detailed evaluations.

To address these limitations, future research may choose to restrict the number of developers, simulations, and medical tasks to evaluate. This may improve consistency with respect to time for evaluation and reduce evaluators' fatigue, though there is the noted tradeoff in terms of detailed results and ability to compare across disparate modalities. Additionally, more time with developers may be provided to encourage more time spent on the evaluations. This may be accomplished through a fractional factorial experimental design in which not every evaluator participates in each simulation. This research design may reduce the workload on evaluators, while providing more time with each developer for evaluation. The results from this study and feedback from the evaluators will be used to refine the evaluations and methodology for future controlled experiments comparing simulation modalities for medical training. We also note that judging developers' approaches to a given context (a given medical task, trainee population, and practice environment) is relatively straightforward compared to judging the value of a developer approach in general, because so many other considerations (e.g., cost, training approaches currently in effect, trainee characteristics, etc.) may influence the overall evaluation. The point is that it is easier to interpret the results of a very restricted comparison than to interpret results of a study where the simulations vary in many ways. If we have simulations that are all the same task, trainee population, and training environment, and differ only by modality, then the effect of modality is more readily apparent as a cause of differences. The trade-off is that the interpretation is also restricted to the specific task, trainee population, and environment.

The data collected during this study represent combat medic experts' judgments of the quality of various simulation training devices, systems, and modalities. This represents an important first step to evaluating the quality of medic TCCC training simulations. Although such data are important for understanding the relative merits of the training approaches, they are not outcome data for trainee skill assessments, training program assessments, or casualty morbidity and mortality. Trainee skill data and training program outcome data are expensive and difficult to obtain, and patient outcomes are subject to both practical and ethical constraints. However, as they become available, such data are important for the evaluation of simulation modalities and training approaches. The current study focused on what capabilities are currently available in medical simulations, not the training outcomes from the simulations.

Conclusion

The aim of the study was to provide empirical data on the quality of various training devices, systems, and modalities for training medics on TCCC. One major conclusion is that the evaluators perceived meaningful differences in the quality of different simulations for TCCC training. However, the underlying condition for the interpretation of the results in all sections is that the context for training and intended learning outcomes are key to determining the appropriate simulation modality and methodology for combat medical training. No modality was considered uniformly superior because views on the suitability of a modality depended on the intended context and goals of training. Modalities have trade-offs. For example, cadaveric simulation was judged less effective in deployment than the low fidelity manikin plus task trainer, but more effective in terms of measurement. Additionally, the current study focused on capabilities currently available in medical simulations, not the training outcomes from the simulations. It is important to view these simulation capabilities in the context of desired training outcomes.

ACKNOWLEDGEMENTS

We would like to acknowledge the United States Special Operations Command who sponsored this work under contract number H9240523F0001 to the University of South Florida Institute of Applied Engineering.

DISCLAIMER

The views expressed in this manuscript are entirely those of the authors and do not necessarily reflect the views, policy, or position of the United States Government, Department of Defense, United States Special Operations Command, or the University of South Florida.

REFERENCES

Butler Jr, F. K. (2017). TCCC updates: Two decades of saving lives on the battlefield: Tactical Combat Casualty Care turns 20. *Journal of Special Operations Medicine, 17*(2), 166-172.

Cannon, J. W., Gross, K. R., & Rasmussen, T. E. (2021). Combating the peacetime effect in military medicine. *JAMA Surgery, 156*(1), 5-6.

Cannon-Bowers, J., Bowers, C., Stout, R., Ricci, K., & Hildabrand, A. (2013). Using cognitive task analysis to develop simulation-based training for medical tasks. *Military Medicine, 178*(suppl_10), 15-21.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*(10), S70-S81.

Furman, G. E., Smee, S., & Wilson, C. (2010). Quality assurance best practices for simulation-based examinations. *Simulation in Healthcare, 5*(4), 226-231.

Horn, G. T., Bowling, F., Lowe, D. E., Parimore, J. G., Stagliano, D. R., & Studer, N. M. (2017). Manikin Human-Patient Simulator Training. *Journal of Special Operations Medicine*, *17*(2), 89-95.

Knisely, B. M., Gaudaen, J. C., Smith, A. V., Perta, J. M., Pamplin, J. C., Quinn, M. T., & Schmidt, P. M. (2022). Evaluating medic performance in combat casualty care simulation and training: A scoping review of prospective research. *Military Medicine.*

Lioce, L., Lopreiato, J., Downing, D., Chang, T. P., Robertson, J. M., Anderson, M., ... & Terminology and Concepts Working Group. (2020). Healthcare simulation dictionary (AHRQ Publication No. 20-0019). *Agency for Healthcare Research and Quality*.

McGaghie, W. C. (2015). Mastery learning: it is time for medical education to join the 21st century. *Academic Medicine, 90*(11), 1438-1441.

McMurray, H., Kraemer, L. S., Jaffe, E., Raiciulescu, S., Switzer, J. M., Dosal, G. C., ... & Bowyer, M. W. (2021). Development of a simulation surgical cricothyroidotomy curriculum for novice providers: A learning curve study. *Military Medicine.*

North Atlantic Treaty Organization (2021). Modelling and simulation technologies for training medical/healthcare professionals (Publication No. RTG-HFM-257).

Perretta, J. S., Duval-Arnould, J., Poling, S., Sullivan, N., Jeffers, J. M., Farrow, L., ... & Hunt, E. A. (2020). Best practices and theoretical foundations for simulation instruction using rapid-cycle deliberate practice. *Simulation in Healthcare, 15*(5), 356-362.

Suresh, M. R., Staudt, A. M., Trevino, J. D., Papalski, W. N., Greydanus, D. J., Valdez-Delgado, K. K., ... & VanFosson, C. A. (2023). A deeper dive into combat medic training. *Military Medicine.*

Suresh, M. R., Valdez-Delgado, K. K., Staudt, A. M., Trevino, J. D., Mann-Salinas, E. A., & VanFosson, C. A. (2021). An assessment of pre-deployment training for Army nurses and medics. *Military Medicine, 186*(1-2), 203-211.