A Systematic Approach for Human Patient Simulation Assessment

Jennifer Winner
US Air Force Research Laboratory
Wright-Patterson AFB, OH
jennifer.winner.1@us.af.mil

Douglas Hodge, David Malek
National Center for Medical Readiness (WSU)
Fairborn, OH
douglas.hodge@wright.edu, david.malek@wright.edu

Caitlan Rizzardo
Aptima, Inc.
Fairborn, OH
crizzardo@aptima.com

Victoria Krumholtz
Leidos, Inc.
Beavercreek, OH
victoria.krumholtz@leidos.com

ABSTRACT

Eastridge et al. (2012) reviewed US Service Member deaths between 2001 and 2011 and found that nearly a quarter of these deaths were potentially survivable, and the majority of those were largely attributable to missed and incorrectly performed interventions for hemorrhage injuries and airway obstruction. The focus of this paper is to describe an approach to improve the empirical assessment and therefore design of Human Patient Simulation (HPS) for training lifesaving airway and hemorrhage interventions. Every action taken by a combat medic has a corresponding outcome for the patient and is in response to the medic's continual assessment of that patient. There are identifiable cues (visual, auditory, or tactile) for successful and unsuccessful airway and hemorrhage interventions, and medics must be capable of detecting them. As more training is performed using HPS devices, the ability to support both procedural practice and patient assessment skills at the right level of fidelity become increasingly important.
Systematic methods are necessary for identifying and addressing deficiencies in HPS and their corresponding impact on training effectiveness. Four considerations relevant to the desired specificity of data are included in the proposed systematic approach to HPS assessment: mapping fidelity assessment to training objectives, representing patient assessment cues, representing variations in injuries and pathologies, and converging measures of trainee performance. While this approach emphasizes a systematic approach to HPS assessment for the training audience of interest (combat medics), the aim is that these considerations will be broadly applicable for medical training.

ABOUT THE AUTHORS

Ms. Jennifer Winner is a Research Psychologist with the Air Force Research Laboratory (AFRL) under the 711th Human Performance Wing's Human Effectiveness Directorate, Warfighter Readiness Research Division (711 HPW/RHA) where she supports the Division's Medical Readiness and Personnel Recovery Team. She has led requirements development for game-based simulation and automated performance assessment systems for use in military training environments. Her most recent work has focused on evaluating the training effectiveness of large-force Personnel Recovery exercises. Ms. Winner received an M.S. in Applied Psychology from Arizona State University and an M.B.A. from Wright State University.

Col (Retired) Douglas Hodge is the Associate Director of Operational & Disaster Medicine at Wright State University and the National Center for Medical Readiness. He has worked in emergency medicine since 1973 and he has been a certified EMT, EMT Instructor, TCCC Instructor, PHTLS Instructor and AMLS Instructor. He served in various medical and leadership positions in the US Air Force including Chief Medical Operations of the 820 SFG, Chief Joint Medical Test, Evaluation & Standards on the Joint Staff, and Deputy Director of the 711th Human Performance Wing. He is a Fellow of the American Academy of Physician Assistants and a Fellow of the Society of Air Force Physician Assistants.

Mr. David Malek is a Research Psychologist at Wright State Research Institute specializing in human factors, human performance, decision making, user-interface design, unmanned systems, and training. He currently leads WSRI's Live, Virtual, and Constructive (LVC) Simulation for training efforts aimed at first responders conducted at the National Center for Medical Readiness (NCMR) and other remote locations. He also supports the 711th Human Performance Wing's Warfighter Readiness Research Division's (711 HPW/RHA) efforts to assess training readiness within the personnel recovery community. Prior to joining WSRI, Mr. Malek led the Joint Training and Experimentation Center (JTEC) located at Camp Guernsey, WY where he led and supported several programs to prototype and evaluate autonomous systems and research the impact of various levels of autonomy on human and system performance. He received an M.S. in Human Factors and I/O Psychology from Wright State University.

Dr. Caitlan Rizzardo is a research scientist with Aptima, Inc. currently supporting warfighter training research at the Air Force Research Laboratory, Warfighter Readiness Research Division (711 HPW/RHA) at Wright-Patterson Air Force Base in Dayton, OH. She has assisted with and led projects studying human performance assessment in simulation-based training; navigation and spatial knowledge acquisition in virtual environments; after-action review augmentation; and 3D-environment display design. She holds a PhD in Human Factors and I/O Psychology from Wright State University.

Ms. Victoria Krumholtz is a behavioral scientist with Leidos currently supporting the Air Force Research Laboratory, Warfighter Readiness Research Division (711 HPW/RHA) at Wright-Patterson Air Force Base, in Dayton OH. She has supported research on a variety of topics, including training assessment and simulation fidelity. She holds an M.A. in Industrial-Organizational Psychology from Xavier University.

A Systematic Approach for Human Patient Simulation Assessment

Jennifer Winner
US Air Force Research Laboratory
Wright-Patterson AFB, OH
jennifer.winner.1@us.af.mil

Douglas Hodge, David Malek
National Center for Medical Readiness (WSU)
Fairborn, OH
douglas.hodge@wright.edu, david.malek@wright.edu

Caitlan Rizzardo
Aptima, Inc.
Fairborn, OH
crizzardo@aptima.com

Victoria Krumholtz
Leidos, Inc.
Beavercreek, OH
victoria.krumholtz@leidos.com

INTRODUCTION

Department of Defense (DoD) guidance for Medical Readiness Training (MRT) specifies that all military service members, in addition to DoD expeditionary civilians, will receive standardized medical training and maintain proficiency as a first responder. The guidance explicitly states that "MRT programs will maximize the use of commercial training simulation, manikins, moulaged actors, and cadavers while reducing the reliance on the live animal model, when appropriate, to prepare Service members to provide effective medical care, minimize casualties, and minimize preventable deaths…" (DoD, 2018, p. 10). Requirements for medical readiness skill sustainment and MRT tracking include a unit-level requirement to report on Tactical Combat Casualty Care (TCCC) initial and recertification training. This latest instruction for MRT reflects the combat medical community's emphases in recent years: reducing the number of potentially survivable combat deaths by improving pre-military medical treatment, increasing the numbers of joint forces trained in combat medical skills, and standardization of combat medical skills training.

Data regarding combat casualties during Operation Iraqi Freedom and Operation Enduring Freedom were analyzed and published in 2011 and 2012 (Eastridge et al. 2012). This work reviewed US Service Member deaths between 2001 and 2011 that occurred prior to arrival at a military Medical Treatment Facility (MTF). According to this report, nearly a quarter of these deaths were potentially survivable (976 out of 3040), and of those potentially survivable deaths, 888 were attributed to hemorrhage and 77 to airway obstruction. Among the conclusions from this work was that the improvement of pre-MTF treatment could significantly impact the survival rates. They concluded that "…emphasis should be placed on trauma system optimization, evidence-based TCCC improvements, and a comprehensive ongoing analysis of all deaths" (Eastridge et al., 2011, p. 436).

In January of 2013, then CENTCOM (The United States Central Command) Commander, General James Mattis, wrote a letter to the Service Chiefs highlighting the success of Tactical Combat Casualty Care (TCCC) training in achieving an unprecedented low incidence of preventable deaths on the battlefield. In this letter he identified TCCC as a means to save lives by improving pre-hospital trauma care. In September of 2015, The Joint Trauma System (JTS) presented a white paper to the Services Surgeons General stressing the importance of standardization of training for all those employing TCCC. The JTS has a considerable amount of ongoing work to ensure the accuracy and consistency of TCCC training across the joint medical lifesaver audiences. Recent communications from the Committee on Tactical Combat Casualty Care (CoTCCC) highlight the importance of standardization for TCCC training, as is required to effectively train increasing numbers of students.

Per the new DoD instruction for MRT, the explicit goal to increase the use of commercial training simulations and manikins implies an increased importance for the fidelity of medical simulation and training devices. Use of human patient simulation (HPS) has increased steadily in recent years; however, systematic methods to quantify their training effectiveness and characterize their fidelity are lacking. Combat medical settings present unique challenges. "Unlike most medical care in civilian settings, treatment for combat trauma casualties involves different injury characteristics, may take place under threat, without proper lighting (or in complete darkness), while evacuating the casualty on uneven terrain." (Nadler et al., 2014, p. 1255). Combat medics must be capable of recognizing patient deterioration in these conditions. Unlike nurses in civilian medical settings who have opportunities for prolonged interaction and can get to "know the patient" (Massey, Chaboyer, & Anderson, 2016), combat medicine involves assessment of largely unfamiliar patients.

In this work we present four considerations aimed at providing a systematic approach to HPS assessment. Our current focus is on the capability of HPS for training lifesaving airway and hemorrhage interventions, since potential deficiencies in these two interventions account for more than 98% of the potentially survivable deaths cited by Eastridge et al., (2012.) These skills are most relevant for the training audience of interest (combat medics); however, our approach emphasizes a systematic approach to HPS assessment that will be applicable across medical training objectives. These considerations can be leveraged for training effectiveness and fidelity research across the gamut of medical simulation capabilities, and will help to provide a most robust picture of the impact of these devices.

We begin our discussion by focusing on fidelity and its relationship to the transfer of skills from training to real-world situations. Next we present four considerations required for systematic HPS assessment: training objectives, patient assessment, injury type/pathology, and converging measures for training effectiveness (e.g., the relationship between confidence and competence). First and foremost, current day characterizations of HPS fidelity are commonly dichotomous or tri-level categories (e.g., low, medium, or high) that lack ties to real-world training objectives or proficiency standards related to combat medical training. Secondly, performing an intervention in the real-world includes assessing patient status, both before and after the intervention. Third, combat medics must be prepared to respond to relevant injury types and pathologies; therefore, these must be accounted for in simulations. Lastly, we emphasize the importance of collecting converging measures of effectiveness, such as interpreting confidence measures relative to actual performance, when evaluating HPS. These four considerations and their importance for fidelity and training effectiveness assessment for combat medical readiness are discussed in the following sections.

FIDELITY AND TRANSFER

Simulation-based training has been utilized in a variety of fields such as aviation, law enforcement, and medical care. Accordingly, fidelity and its link to performance and transfer of skills have been frequent topics of discussion. There is much from historical simulator design and evaluation that may be borrowed for HPS evaluation but, because of unique components for the medical readiness training, the effectiveness of any human patient simulation needs to be evaluated with an eye for the successful emulation of specific patient-related needs.

A review conducted by Doolen et al. (2016) revealed a lack of standardized research on the effectiveness of using simulations in undergraduate nursing, despite the wide-spread use of this type of training. Importantly, Doolen et al. (2016) also emphasizes the lack of systematic research linking so-called high-fidelity simulations to performance and outcomes. Evaluations that do exist do not empirically confirm that HPS, as currently designed, meet the fidelity requirements for effective training. In a pilot study evaluating the impact of simulation fidelity on pre-nursing student test scores, Kardong-Edgren, Anderson, and Michaels (2007) did not observe any difference in student test scores when comparing data from groups learning in low- and high-fidelity training systems. One interpretation of such results is that there is indeed no discernible difference in the training capability of these 'low' and 'high' fidelity systems. However, without a systematic approach to assessment for simulation-based medical care training, an alternative interpretation could be that the ability to discern between fidelity levels is lacking. HPS systems initially designated as 'high' fidelity may not have the requisite types of fidelity for training medical care tasks.

Myers, Starr, and Mullins (2018) summarize the results of studies in aviation attempting to find direct links between low and high fidelity with negative and positive training transfer. Though the study results vary, the authors emphasize the importance of appropriate cues to manufacture training situations that accurately represent real world situations for operator decision making. As summarized by Myers et al. (2018), the National Transportation Safety Board (NTSB) (1990; 1997; 1999; 2001) determined that inaccurate cues in operators' simulation-based training were contributing factors in several aviation incidents. This finding led the NTSB to recommend improvements to specific simulator capabilities. Bürki-Cohen et al. (2000) point out that the lack of fidelity in simulated air traffic control communications leads to simulating an incomplete cognitive environment for pilots, creating a false sense of simplicity. Lee (2009) listed simulator disadvantages to include poor motion cueing, which induces adaptation and compensatory skills. Post simulation training, operators apply improper solutions in real-world situations, but their errors stem from errors in fidelity. Operators have been misled in their expectations of real-world event characteristics and cues.

The importance of fidelity has also been emphasized within Instructional Design Theory. Gibbons, McConkie, Seo, and Wiley (2009) identified principles associated with supplying model content for simulation approaches to instruction. The article specified that simulation designers must account for task fidelity. Task fidelity is focused on the extent to which "the actions taken by the learner resemble actions taken in real environments" (p. 176). Designers must also account for the environmental fidelity. Environmental fidelity is the extent to which "the sensations in simulation response environments resemble the sensations produced in real environments" (p. 176). Additionally, instructional designers must account for realism, which is defined as the "degree to which sensory experiences from external representations of the model conform to sensory information available in real settings" (p. 177). Simulation devices that allow for medical intervention practice in the absence of the real-world cues violate known constructivist instructional principles. Specifically, such simulations do not allow for the design of an authentic task or the design of a task that reflects the complexity of the real-world environment in which the trainee must be able to function (Savery & Duffy, 1995).

It is important to note that the use of the terms 'cueing' or 'cue' within the transfer literature and its use within the medical simulation literature sometimes differ. In the transfer literature, it is often used to denote the meaningful information necessary for decision making (Bürki-Cohen et al., 2000; Myers et al., 2018; Lee, 2009).Within the medical simulation literature, negative transfer has been attributed to failures on the part of the instructor (Fritz, Gray, & Flanagan 2008). Specifically, some have discussed the use of cueing to help calibrate trainees to the intent of the scenario and correct for discrepancies between the simulation and the real-world task (Paige & Morin, 2013). In other words, 'cue' is used to refer to the 'white card' inputs that instructors must give to account for 'sim-isms'. From our perspective, although the instructor plays a significant role in simulation-based training, fidelity does not ride solely on instructors' shoulders. It is worth a closer look at the skills required for mission readiness and the extent to which an HPS can provide practice for those skills, regardless of instructor or simulation facilitator actions or experience. In the following section we outline an approach for assessment that is applicable to the HPS problem by focusing on four considerations.

HUMAN PATIENT SIMULATION FIDELITY ASSESSMENT

The fidelity of current-day HPS is less than perfect. Empirical data has highlighted physical discrepancies in airway anatomy between varieties of HPSs when compared to live human patients (Schebesta et al., 2012). Such discrepancies have potential to negatively impact training in addition to the validation of treatment protocols and lifesaving intervention devices (Klock 2012). Accurate and successful mapping of fidelity requirements to training outcomes is a significant undertaking, regardless of domain (e.g., Muller et al., 2006).

Meaningful fidelity assessment begins with a list of specific training objectives, and evaluates the training capabilities against those objectives. Historically, this method, which is also referred to as training capability assessment, has been conducted for aviation using a breakdown of training experiences coming from either the Mission Essential Competencies (MEC^{SM}) or a list of Ready Aircrew Program task events (e.g., Alliger et al., 2015). Ideally, competency standards such as the MECs are available to be leveraged for standardized evaluations across simulation systems (Schreiber, 2013). Using this approach, the competency standards serve as the basis for experts' ratings of the training capabilities. In cases where ratings are low, deficiencies are identified. As such, dichotomous or tri-level categories such as 'low, medium, or high' are too generic. Human-patient simulators as 'high fidelity' do not ensure that those simulations or devices provide the cues at the suitable level of granularity for some specific training objectives.

The training of medical personnel tasked with point-of-injury medical care requires very specific skill development. These training requirements imply four important considerations for the use of and assessment of simulation and devices.

1. <u>Training Objective Mapping</u>: The training objectives of these medics, in the form of specific medical interventions, need to be considered carefully. There are specific lifesaving interventions that certain medics of varying levels of certification should be qualified to perform and each of those interventions may require different types and degrees of simulation support for assessment, treatment, and re-assessment.

2. Patient Assessment: Patient assessment responsibilities are essential to the medic role. It is vital to the trauma care process and the cues that enable a medic to properly assess a patient be adequately modeled.
3. Injury Type/ Pathology: Medical readiness training requires that different pathologies must be considered in concert with the lifesaving interventions. Traumatic injuries will affect the ways in which patients are assessed and the ways the interventions need to be performed. Assuming a nominal or normal patient state when training lifesaving interventions is insufficient.
4. Converging Measures of Training Effectiveness: In an ideal world, simulation-based training will appropriately calibrate trainee confidence with their ability to perform the practiced skills in real-world situations. Rigorous evaluations will look at a variety of measures to corroborate the effectiveness of the simulations. All of these considerations are addressed in more detail in the following sections.

Consideration #1: Training Objective Mapping

Systematic methods are required to ensure appropriate categorization and, ultimately, meaningful training effectiveness comparisons across HPS. Just as it is important to design medical training assessment around training objectives (Boulet & Murray, 2010), fidelity assessment should be conducted and mapped to medical readiness training objectives.

Previous military training research has focused on the identification of simulator deficiencies by structuring assessments around specific training objectives (Prost, Schreiber, & Bennett, 2007). Under this work, 'deficiency' has been defined as a limitation in the hardware and/or software that limit the training experiences that can be provided by a particular training system (Prost, Schreiber, Bennett, & Kleinlein, 2007). This method offers specificity otherwise lacking in the dichotomous or tri-level characterizations of fidelity. Motivation for this method includes the ability to make data driven decisions regarding simulator upgrades, and map it to the amount of training capabilities (e.g., number of training experiences, priority training experiences) provided by a given simulation-based training system. This approach relies on experts to provide feedback regarding the capabilities of a particular system. Overall, this approach is a utility assessment as described by Bell and Waag (1998), in that it relies on the opinions of experts to characterize the training capabilities of a particular simulation system.

This approach is directly applicable to HPS fidelity assessment and there are definitive sources for training objectives within the combat medical domain that can be leveraged. The Prehospital Trauma Life Support (PHTLS) textbook provides detailed guidance regarding the lifesaving interventions relevant to the combat medic community (NAEMT, 2014). This textbook details the guidelines for administering numerous airway and hemorrhage interventions. A second potential source is the TCCC training content which lists airway and hemorrhage interventions as well as patient assessment criteria mapped to potential deficiencies. Sources such as these, in collaboration with medical experts, can begin to lay the foundation for increasing the granularity of fidelity evaluations.

In coordination with the staff at the National Center for Medical Readiness in Fairborn, Ohio, semi-structured interviews were conducted to identify examples of medical interventions relevant to combat medics for both airway and hemorrhage. The following bulleted list of airway interventions are examples of those of interest to the combat medical community.

- Oropharyngeal airway (OPA)
- Nasopharyngeal (NPA)
- Supraglottic – King LT
- Digital Intubation (Tactile Intubation)
- Video Laryngoscopy
- Rapid Sequence Intubation (RSI)
- Supraglottic – Combi-Tube
- Supraglottic –Laryngeal mask airway (LMA)
- Endotracheal tube
- Nasotracheal Intubation (N-ETT)
- Surgical airway (CRIC key)

Similarly, fidelity assessment for the following hemorrhage interventions are also of interest to this community.

- Dressing
- Direct pressure
- Hemostatic agents – Combat gauze, Celox gauze, ChitoGauze, XSTAT-30
- Wound packing
- Combat Application Tourniquet (C-A-T)
- SOF-T tourniquet
- Junctional tourniquets - Abdominal aortic tourniquet (AAT), CRoC, JETT, SJT
- Tranexamic Acid (TXA)

As TCCC and other comprehensive medical readiness program requirements evolve, so will these lists. The bottom line here is that fidelity assessment should be conducted and mapped to medical readiness training objectives such as those examples listed above.

Consideration #2: Patient Assessment

Prior research has discussed three types of fidelity: environment, equipment, and psychological fidelity, which "Reflects the degree to which the trainee perceives the simulation to be a believable representation of the reality it is duplicating" (Fritz et. al., 2008, p. 2). HPS should endeavor to model a believable representation with a focus on the most relevant environmental, equipment, and psychological components of a training objective. However, HPS manikins commonly lack a variety of critical cues for patient assessment (e.g., change in the manikin's nailbed or lip color when airway is blocked and when an intervention is effectively managed).

Every action taken by a combat medic has a corresponding outcome for the patient and is in response to the medic's continual assessment of that patient. There are identifiable cues (visual, auditory, or tactile) for successful and unsuccessful airway and hemorrhage interventions, and medics must be capable of detecting them. As more training is performed using HPS devices, the ability to support both procedural practice and patient assessment skills at the right level of fidelity become increasingly important.

Fisher and King (2013) presented a review focused on the effectiveness of simulation to train nurses responding to a deteriorating patient. They concluded that, generally speaking, simulations have been shown to enhance nurses' confidence, clinical judgement, knowledge, and competence; however, they identified gaps in existing evidence of simulation-based training as an effective means to detect cues and respond to deteriorating patients. In an integrative review of nursing studies focused on recognition of patient deterioration, Massey et al. (2016) reported on four primary themes: assessing the patient, knowing the patient, education, and equipment. To date, there has been little or no explicit discussion of the fidelity of patient assessment cues in the medical simulation literature. The current work seeks to address this gap.

In coordination with the staff at the National Center for Medical Readiness in Fairborn, Ohio, semi-structured interviews were conducted to identify examples of critical cues relevant to both airway and hemorrhage interventions. The following bulleted list includes examples of critical cues related to airway interventions are important fidelity assessment considerations as it relates to patient assessment within medical readiness training.

- Chest rise – one lung
- Chest rise – both lungs
- Feel breathing (i.e. hand on chest)
- Stomach rise
- Blow air in and see chest rise
- Eye lid color cue
- Fingernail color cues
- Lip color cues
- Skin color cues (e.g., cyanosis, gray, white or blue)
- Capnometry functionality
- Pulseoximetry functionality
- Kinked tube- audible cue
- Kinked tube – resistance in bag valve

- Auditory breathing cues (e.g., whistling, wheezing)
- Stridor – harsh, vibratory sound

Similarly, the following critical cues related to hemorrhage interventions are important to consider in regard to patient assessment.

- Blood flow – slow
- Blood flow – medium
- Blood flow – massive (pulsating/spurting)
- Tissue - color change (e.g., flush/pink from blood flow; blue/gray from lack of flow)
- Tissue – temperature (e.g., warmer/cooler due to changes in blood flow)
- Changes in blood flow as pressure is applied (e.g., slows down as pressure increases, speeds up as pressure decreases/tourniquet loosened)
- Does bleeding slow down/ stop after a certain amount of time given pressure (e.g., does the wound stop/slow down bleeding after tourniquet is loosened, but after tourniquet has been tight for several minutes

Consideration #3: Injury Type/Pathology

In previous sections, discussion has highlighted the importance of simulation fidelity for instructional effectiveness. Standards of best practice for medical simulation dictate that physical fidelity is required and that the simulation content should mimic the situation that will be faced in the real world (Lairet, et al., 2012; Lioce et al., 2015; Timmermann, 2011). In a combat medical setting, this includes the types of injuries likely to be encountered; it is not enough that simulation permits the practice of only basic types of lifesaving airway and hemorrhage interventions medics will need to use (Nadler et al., 2014). Fidelity related to relevant injury types and how intervention procedures are affected (often made more difficult) is also important. How do fidelity requirements, including the needs regarding patient assessment, change based on the injury type? The ability to present clinical variation has been identified as an important feature of medical simulation in that it leads to effective learning (Issenberg, McGaghie, Petrusa, Gordon, & Scalese, 2005). Historical data is able to be leveraged to highlight the likely injuries that medics must be capable of treating (Eastridge et al. 2012; Eastridge et al. 2011).

It is an empirical question as to whether the variety of HPSs can provide effective training across the continuum of all medical intervention/pathology combinations. Which patient assessment cues are enabled by today's HPS, which are lacking, and does this depend on pathology? Likely, all HPS devices have their strengths pertaining to a subset of the medical interventions, cues for patient assessment, and pathologies. Systematic evaluation will enable the identification of those strength along with data that can be utilized for continued advancement.

For the above-mentioned airway interventions, the following pathology examples are presented for consideration.

- Normal
- Blast trauma to head and neck
- Inhalation injury
- Oropharyngeal hemorrhage
- Glottic edema

For the above-mentioned hemorrhage interventions, the following injuries are presented for consideration.

- Extremity – soft tissue injury
- Extremity – minimal bleeding
- Head, chest, abdomen – soft tissue injury – minimal bleeding
- Partial or full amputation
- Junctional hemorrhage – groin/buttocks/perineum
- Junctional hemorrhage – axillae/base of neck
- Massive blunt trauma
- Non-compressible neck and high groin

Consideration #4: Converging Measures Of Training Effectiveness

As previously stated, according to the most recent medical readiness training guidance, “MRT programs will maximize the use of commercial training simulation, manikins, moulaged actors, and cadavers while reducing the reliance on the live animal model, when appropriate, to prepare Service members to provide effective medical care, minimize casualties, and minimize preventable deaths…” (DoD, 2018, p. 10). Use of the phrase ‘when appropriate’ subtly but effectively underlines the importance of understanding the training effectiveness of each method. The considerations discussed thus far have potential to structure HPS fidelity evaluations. The first three considerations are applicable to both utility evaluations and structuring the reporting of in-simulator learning. The fourth consideration addresses measures of training effectiveness applicable to HPS evaluations using in-simulator learning. Specifically, we argue that converging measures are necessary to ensure warranted conclusions of HPS training effectiveness.

An individual measure of training effectiveness cannot stand alone. Both the five-stage model of evaluation by Bell and Waag (1998) and Kirkpatrick’s (1994) four levels for evaluating training programs highlight the value of converging measures when assessing the training effectiveness. A commonly reported measure of training effectiveness is the change in trainee confidence in their ability to perform a task, referred to as self-efficacy (Bandura, 1977/2000). Confidence has been estimated for simulation-based training in a wide variety of medical domains, most frequently nursing but extending to critical care, physical therapy, pharmacy, and doctors as well (Blum et al., 2010, Corbridge et al., 2008, Liaw, Scherpbier, Rethans, Klainin-Yobas, 2012; Marteau, Wynne, Kaye, & Evans, 1990 and as summarized in reviews from Alanazi, Nicholson, & Thomas, 2017, Nishisaki, Keren, & Nadkarni, 2007, Yuan, Williams, & Fang, 2011). Confidence questionnaires are typically created by enlisting a group of experts or using existing checklists of job tasks for items relevant to the simulation scenario. Researchers often ask students before and after simulation-based training to rate their ability on Likert-type scales. These rating scales commonly include scale point labels created by combining a general efficacy statement such as, ‘I am able to’ or ‘I am confident that’ and a task relevant to the training objective (Holbrook and Cennamo, 2014; Stewart, O’Halloran, Barton, Singleton, Harrigan, & Spencer, 2000). The efficacy statement chosen for the item lists needed to, as described by Stewart et al. (2000), “reflect real and meaningful measures of ability and progression” per task or skill. Students largely report increased confidence in their ability to perform tasks post simulation-based training.

However, research on training effectiveness that measures confidence and competence has yielded mixed results for students’ ability to correctly rate their confidence to do a skill compared to their actual ability to perform that skill (Liaw et al., 2012; Marteau et al., 1990). Higher confidence does not necessarily indicate higher competence. Yuan et al. (2011), reviewing the impact of high-fidelity simulators in nursing practice, and Nishikasi, Keren, & Nadkarni (2007), reviewing the impact of simulation on patient safety outcomes, looked broadly at research investigating confidence, competence, or the two together. Confidence often increased following simulation-based training but the pre-post performance scores did not show a consistent increase. The reviewers concluded that the body of literature as a whole does not yet provide a robust conclusion that increased perceptions of self-efficacy also indicate increased performance.

Similar concerns about the relationship between confidence and performance have been documented in literature regarding the Dunning-Kruger Effect. Kruger and Dunning (1999) gathered participant confidence ratings after practicing a task and compared those ratings to task performance, creating a score of confidence calibration accuracy. They found that level of task knowledge impacted confidence calibration. Novices rated themselves as more confident in their skill level than their actual performance warranted, while for experts this was the opposite. This pattern serves as a caution against using a single measure such as confidence. Converging measures are required to make claims about training effectiveness. Additionally, objective and/or subjective measures of outcome and skill-based performance should address the examples of the interventions, patient assessment cues, and/or the injury types presented in Considerations 1-3.

Researchers should continue to jointly assess fidelity and training effectiveness, taking into consideration the type and level of granularity for fidelity as it affects trainee perceptions and performance (O’Donnell, Decker, Howard, Levett-Jones, & Miller, 2014; Paskins & Peile, 2010). Utility evaluations should be accompanied by in-simulator learning evaluations and converging outcome measures are required for valid conclusions about simulator capabilities. These capabilities, mapped to Considerations 1-3, present a more systematic evaluation compared to current-day approaches.

CONCLUSION

The four considerations discussed in this paper are intended to increase the specificity of data for discerning levels of HPS fidelity and then framing evaluations of training effectiveness. These considerations are applicable across the variety of medical simulation capabilities. The first three considerations provide a structured approach for fidelity evaluations. To what extent can HPS be used to train the relevant combat medical interventions with the appropriate patient assessment cues for the likely injury types? Subjective utility evaluations organized by these considerations allow for data to be iteratively amassed and analyzed as capabilities advance. In addition to subjective ratings of training capabilities, evaluations using in-simulator learning will help paint a more robust picture of HPS capabilities. Collectively, these two types of evaluations address the first two stages of Bell and Waag's five-stage model for evaluating simulator training (1998).

The considerations outlined here are also compatible with Kirkpatrick's model for evaluating training effectiveness (1994). The four levels of Kirkpatrick's model (reactions, learning, performance, and results or transfer) are also compatible with evaluating HPS fidelity and training effectiveness with the specific context of the considerations described above. Quite often, reaction data is among the only data available for evaluating training capabilities (Boulet et al., 2011).

In the medical simulation literature, one training evaluation framework has been widely leveraged. Jeffries (2005) introduced a framework to aid in understanding which simulation-based teaching practices led to favorable outcomes for nurses in academic training and in the real-world. In collaboration with the National League for Nursing (NLN) and the Laerdal Corporation, the NLN/Jeffries Simulation Framework (NLN/JSF) was created to structure the design and evaluations of simulation-based training (Jeffries, 2005; Jeffries & Rogers, 2012). The original framework included five conceptual components: teacher factors, student factors, educational practices, simulation design, and outcomes. Fidelity is listed as one of the variables associated with the simulation design. Additionally, skill performance, learning, and self-confidence are all variables highlighted under the outcomes component. In 2011, the International Association of Clinical Simulation and Learning (INACSL) convened a series of expert teams to review each of the NLN/JSF constructs. These teams focused on documenting the current state of the science on each of the constructs and their subcomponents, as well as identifying gaps and future research directions (Durham, Cato, & Lasater, 2014; Groom, Henderson, & Sittner, 2014; Hallmark, Thomas, & Gantt, 2014; Jones, Reese, & Shelton, 2014; O'Donnell et al., 2014; Ravert & McAfooes, 2014). Previous work has found that standardization of terms and methodologies for evaluating HPS is needed, and that there is a gap in empirical data speaking to the relationship between simulation fidelity and skill retention (Groom et al., 2014). Similarly, the teams also reported that additional research is required to further articulate the impact of simulation-based training on confidence and performance (O'Donnell et al., 2014). Despite the wealth of effort to date, empirical data reported using the Jeffries Framework has yet to document training objectives, patient assessment cues, and the injury types at the level of specificity discussed in considerations 1-3.

This work aims to provide a systematic approach for HPS assessment, specifically applied to combat medical readiness training. The four considerations are compatible with the Bell and Waag five-stage evaluation model for simulator training (1998), Kirkpatrick's model for evaluating training effectiveness (1994), and the NLN/Jeffries framework (2005). The considerations support both utility evaluations and in-simulator training studies related to HPS. The considerations provide guidance for training effectiveness and fidelity research across the gamut of medical simulation capabilities, and will help practitioners become increasingly better informed consumers of these devices.

Evidence-based practice is an approach that has been widely adopted as it relates to clinical practice within the medical community. Although conversations within the community have included a focus on training (Institute of Medicine, 2007), there is more work to be done. This effort aims to address evidence-based adoption of HPS for training. Systematic evaluations which account for the four considerations presented in this paper can be leveraged by medical practitioners to ensure that the systems they adopt are aligned with and will be effective for their focused training objectives. Expanding fidelity methods to address these considerations is necessary to meet the growing need for HPS-enabled medical readiness training.

REFERENCES

Alanazi, A. A., Nicholson, N., & Thomas, S. (2017). The use of simulation training to improve knowledge, skills, and confidence among healthcare students: A systematic review. Internet Journal of Allied Health Science and Practice, 15(3). Retrieved from https://nsuworks.nova.edu/ijahsp/vol15/iss3/2/.

Alliger, G. M., Bennett, W., Jr., Rowe, L., Donsbach, J. S., Beard, R., & Colegrove, C. (2015, April). Competencies Across the Training Continuum: Fundamental to Mission Essential Competencies. In W. Bennett, Jr., (Chair), *Determining training needs: Applications in and lessons from military TNA.* Symposium conducted at the annual Society for Industrial-Organizational Psychology (SIOP) Conference, Philadelphia, PA.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. doi:10.1037/0033-295X.84.2.191PMID:847061

Bandura, A. (2000). Self-efficacy. In A. E. Kazdin (Ed.), Encyclopedia of Psychology (Vol. 7, pp. 212–213). Washington, DC, US: American Psychological Association; Retrieved from http://psycnet.apa.org.ezproxy.lib.vt.edu:8080/books/10522/094

Bell, H. H., & Waag, W. L. (1998). Evaluating the effectiveness of flight simulators for training combat skills: A review. *The International Journal of Aviation Psychology, 8(3),* 223-242.

Blum, C. A., Borglund, S., & Parcells, D. (2010). High-fidelity nursing simulation: Impact on student self-confidence and clinical competence. *International Journal of Nursing Education Scholarship, 17*(1), 1-14. doi:10.2202/1548-923X.2035

Boulet, J. R., Jeffries, P. R., Hatala, R. A., Korndorffer, J. R., Feinstein, D. M., & Roche, J. P. (2011). Research regarding methods of assessing learning outcomes. *Simulation in Healthcare, 6,* 48-51.

Boulet, J. R., & Murray, D. J. (2010). Simulation-based assessment in anesthesiology: Requirements for practical implementation. *Anesthesiology, 112,* 1041-1052.

Bürki-Cohen, J., Boothe, E. M., Soja, N. N., DiSario, R. D., Go, T., & Longridge, T. (2000, May). Simulator fidelity: The effect of platform motion. In proceedings of the Royal Aeronautical Society conference on Flight Simulation—The Next Decade, London.

Corbridge, S. J., McLaughlin, R., Tiffen, J., Wade, L., Templin, R. & Corbridge, T. C. (2008). Using simulation to enhance knowledge and confidence. The Nurse Practitioner, 33(6), 12-13.

Department of Defense (2018, March 16). *Medical Readiness Training (MRT;* DoD Instruction 1322.24). Washington, DC: Office of the Under Secretary for Personnel and Readiness. Retrieved from http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/132224p.pdf?ver=2018-03-16-074408-137.

Doolen, J., Mariani, B., Atz, T., Horsley, T. L., Rourke, J. O', McAfee, K., & Cross, C. L. (2016, June). High fidelity simulation in undergraduate nursing education: A review of simulation reviews. *Clinical Simulation in Nursing*, *12*(7), 290-302.

Durham, C. F., Cato, M. L., & Lasater, K. (2014). NLN/Jeffries Simulation Framework State of the Science Project: Participant Construct. *Clinical Simulation in Nursing, 10*(7), 363-372.

Eastridge, B. J., Hardin, M., Cantrell, J., Oetjen-Gerdes, L., Zubko, T., Mallak, C. & Bolenbaucher, R. (2011). Died of wounds on the battlefield: causation and implications for improving combat casualty care. *Journal of Trauma and Acute Care Surgery, 71*(1), S4-S8.

Eastridge, B. J., Mabry, R. L., Seguin, P., Cantrell, J., Tops, T., Uribe, P., & Butler, F. K. (2012). Death on the battlefield (2001–2011): Implications for the future of combat casualty care. *Journal of Trauma and Acute Care Surgery, 73*(6), S431-S437.

Fisher, D. & King, L. (2013). An integrative literature review on preparing nursing students through simulation to recognize and respond to the deteriorating patient. *Journal of Advanced Nursing, 69*(11), 2375-2388. doi:10.1111/jan.12174. PMID 23734570.

Fritz, P. Z., Gray, T., & Flanagan, B. (2008). Review of mannequin-based high-fidelity simulation in emergency medicine. *Emergency Medicine Australasia, 20*(1), 1–9.

Gibbons, A. S., McConkie, M., Seo, K. K., & Wiley, D. (2009), Theory for the design of instructional simulations and microworlds. In C. M. Reigeluth and A. Carr-Chellman (Eds.), Instructional-design theories and models, Volume III. Mahwah, NJ: Lawrence Erlbaum Associates.

Groom, J. A., Henderson, D., & Sittner, B., J. (2014). NLN/Jeffries Simulation Framework State of the Science Project: Simulation Design Characteristics. *Clinical Simulation in Nursing, 10*(7), 337-344.

Hallmark, B. F., Thomas, C. M., & Gantt, L. (2014). The Educational Practices Construct of the NLN/Jeffries Simulation Framework: State of the Science. *Clinical Simulation in Nursing, 10*(7), 345-352.

Holbrook, H. A. & Cennamo, K. S. (2014). Effects of high-fidelity virtual training simulators on learners' self-efficacy. *International Journal of Gaming and Computer-Mediated Simulations, 6*(2), 38-52. doi:10.4018/ijgcms.2014040104

Institute of Medicine (2007). *The Learning Healthcare System: Workshop Summary*. Washington, DC: The National Academies Press. https://doi.org/10.17226/11903.

Issenberg, S. B., Mcgaghie, W. C., Petrusa, E. R., Gordon, D. L. & Scalese, R. J. (2009). Features and uses of high-fidelity medial simulations that lead to effective learning: A BEME systematic review. *Medical Teacher, 27*(1). doi:10.1080/01421590500046924

Jeffries, P. R. (2005). A framework for designing, implementing, and evaluating simulation used as teaching strategies in nursing. *Nursing Education Perspectives, 26*(2), 96-103.

Jeffries, P. R., & Rogers, K. J. (2012). Theoretical framework for simulation design. In P. R. Jeffries (Ed.), Simulation in nursing education: From conceptualization to evaluation (2nd ed., pp. 25-42). New York: National League for Nursing.

Jones, A. L., Reese, C. E., & Shelton, D. P. (2014). NLN/Jeffries Simulation Framework State of the Science Project: The Teacher Construct. *Clinical Simulation in Nursing, 10*(7), 353-362.

Kardong-Edgren, S., Anderson, M., & Michaels, J. (2007). Does simulation fidelity improve student test scores? *Clinical Simulation in Nursing Education, 3*, 21-24.

Kirkpatrick, D. L. (1994). *Evaluating Training Programs: The Four Levels.* San Francisco: Berrett-Koehler.

Klock, P. A. (2012). Airway simulators and mannequins: A case of high fidelity? Anesthesiology, 116(6), 1179-1180.

Kruger, J. & Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.

Lairet, J. R., Bebarta, V. S., Burns, C. J. et al. (2012). Prehospital interventions performed in a combat zone: A prospective multicenter study of 1,003 combat wounded. *Pre-Hospital Clinical Research, 73*(2), 38-42.

Lee, A.T., (2009). Flight simulation virtual environments in aviation. Burlington, VT: Ashgate.

Liaw, S. Y., Scherpbier, A., Rethans, J-J., & Klainin-Yobas, P. (2012). Assessment for simulation learning outcomes: A comparison of knowledge and self-reported confidence with observed clinical performance. *Nurse Education Today, 32*(6), 35-39.

Lioce L., Meakim C. H., Fey M. K., Chmil J. V., Mariani B., & Alinier G. (2015, June). Standards of best practice: Simulation standard IX: simulation design. *Clinical Simulation in Nursing, 11*(6), 309-315. http://dx.doi.org/10.1016/j.ecns.2015.03.005.

Marteau, T. M., Wynne, G., Kaye, W., & Evans, T. R. (1990). Experience Without Feedback Increases Confidence But Not Skill. *British Medical Journal, 300*(6728), 849-850.

Massey, D., Chaboyer, W., & Anderson, V. (2016). What factors influence ward nurses' recognition of and response to patient deterioration? An integrative review of the literature. *Nursing Open*. doi:10.1002/nop2.53.

Muller, P., Cohn, J. Schmorrow, D., Stripling, R., Stanney, K., Milham, L., Jones, D., Whitton, M. C., & Fowlkes, J. E. (2006). The fidelity matrix: Mapping system fidelity to training outcome. *In the Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC): Orlando, FL.*

Myers, P. L., Starr, A. W., & Mullins, K. (2018). Flight simulator fidelity, training transfer, and the role of instructors in optimizing learning. *International Journal of Aviation, Aeronautics, and Aerospace, 5*(1), 1-26. Retrieved from https://commons.erau.edu/ijaaa/vol5/iss1/6.

Nadler, R., Gendler, S., Chen, J., Lending, G., Abramovitch, A., & Glassberg, E. (2014). The Israli Defense Force experience with intraosseous access. *Military Medicine, 179*, 1254-1257.

National Association of Emergency Medical Technicians (U.S.), & American College of Surgeons. (2014). PHTLS: Prehospital trauma life support (military edition.)

National Transportation Safety Board. (1990). *Multiple Operators, Multiple Locations,* (Safety Recommendation A-90-044). Washington, DC. Retrieved from https://www.ntsb.gov/safety/safetyrecs/_layouts/ntsb.recsearch/Recommendation.aspx?Rec=A-90-044

National Transportation Safety Board. (1997). *Airborne Express, Douglas DC-863, N827AX, Narrows, Virginia, December 22, 1996* (Safety Recommendation A-97-047). Washington, DC. Retrieved from https://www.ntsb.gov/safety/safety-recs/_layouts/ntsb.recsearch/Recommendation.aspx?Rec=A-97-047

National Transportation Safety Board. (1999). *US Airways, Boeing 737-300, N513AU, Aliquippa, Pennsylvania, September 8, 1994* (Safety Recommendation A-99-026). Washington, DC. Retrieved from https://www.ntsb.gov/safety/safety-recs/_layouts/ntsb.recsearch/Recommendation.aspx?Rec=A-99-026

National Transportation Safety Board. (2001). *American Air Lines, A300-605R, N14053. Dallas-Belle Harbor, New York. November 12, 2001 (*Aircraft accident report, NTSB/AAR-89/04). Washington, DC. Retrieved from https://www.ntsb.gov/investigations/Accident.Reports/Pages/AAR0404.aspx.

Nishisaki, A., Keren, R., & Nadkarni, V. (2007). Does simulation improve patient safety? Self-efficacy, competence, operational performance, and patient safety. *Anesthesiology Clinics, 25*, 225-236. doi:10.1016/j.anclin.2007.03.009

O'Donnell, J. M., Decker, S., Howard, V., Levett-Jones, T., & Miller, C. W. (2014). NLN/Jeffries Simulation Framework State of the Science Project: Simulation Learning Outcomes. *Clinical Simulation in Nursing, 10*(7), 373-382.

Paige, J. B., & Morin, K. H. (2013). Simulation fidelity and cueing: A systematic review of the literature. *Clinical Simulation in Nursing, 9,* 481-489.

Paskins, Z., & Peile, E. (2010). Final year medical students' views on simulation-based teaching: A comparison with the best evidence medical education systematic review. *Medical teacher, 32(7),* 569-577.

Prost, J. H., Schreiber, B. T., Bennett, W. (2007). Identification and evaluation of simulator system deficiencies. In the proceedings of the *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC): Orlando, FL.*

Prost, J. H., Schreiber, B. T., Bennett, W., & Kleinlein, K. (2007). Providing effective and efficient training: a model for comparing simulator improvements. In the proceedings of the *Simulation Interoperability Workshop (SIW): Orlando, FL.*

Ravert, P., & McAfooes, J. (2014). NLN/Jeffries Simulation Framework: State of the Science Summary. *Clinical Simulation in Nursing, 10*(7), 335-336.

Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology, 35*(5), 31-38

Schebesta K., Hüpfl, M., Rössler, B., Ringl, H., Müller, M. P., & Kimberger, O. (2012). Airway anatomy of high-fidelity human patient simulators and airway trainers. Anesthesiology, 116(6), 1204-1209.

Schreiber, B. T. (2013). Transforming training: a perspective on the need for and payoffs from common standards. *Military Psychology, 25,* 294-307.

Stewart, J., O'Halloran, C., Barton, J. R., Singleton, S. J., Harrigan, P., & Spencer, J. (2000). Clarifying the concepts of confidence and competence to produce appropriate self-evaluation measurement scales. *Medical Education, 34*, 903-909. doi:10.1046/j.1365-2923.2000.00728

Timmermann, A. (2011). Supraglottic airways in difficult airway management: Successes, failure, use and misuse. Journal of the Association of Anaesthetists of Great Britain and Ireland, 66, 45-56.

Yuan, H. B., Williams, B. A., Fang, J. B. (2011). The contribution of high-fidelity simulation to nursing students' confidence and competence: a systematic review. *International Nursing Review, 59,* 26–33. doi:10.1111/j.1466-7657.2011.00964