The Journal of Education in Perioperative Medicine

Original Research

The Physiologic Effect of Augmented Reality Simulation Versus Traditional Simulation: A Noninferiority, Randomized Controlled Trial

Asheen Rama, MD
Marcos S. Rojas-Pino, MD
Ellen Y. Wang, MD
Samuel T. Rodriguez, MD
Man Yee Suen, MMedSc
Janet S. Titzler, BS
Michelle Zuniga-Hernandez, BS
Christian Jackson, MS
Oswaldo Rosales, BA
Faith Collins, BS
Thomas J. Caruso, MD, PhD

Introduction

Simulation is an indispensable, experiential training tool for health care professionals.[1-3] Technological advancements in the late 20th century heralded mannequin task trainers and full-body simulators that mimicked human physiology.[2] In addition to simulations at freestanding facilities, in situ simulations have grown in popularity given their ability to replicate scenarios in hospital care units with actual workflows.[3]

Medical simulations induce physiologic stress responses, which potentiate long-term memories.[4] Unlike lectures and problem-based learning discussions, simulations harness experiential learning to evoke apprehension during challenging scenarios, leading to greater recall.[5] Stress hormones, particularly glucocorticoids, play a significant role in declarative memories.[6,7] In the absence of stress hormones, such as during pharmacological inhibition of glucocorticoids under study conditions, limited recall of declarative memories is observed in humans.[6] Furthermore, when stress hormones increase during a tense situation, they activate the amygdala and hippocampus.[6,7]

Stress plays a significant role in medical simulation learning. High-fidelity simulations can evoke physiological stress responses in students, with heart rate increases observed during training.[8] Moderate stress levels can enhance performance and learning outcomes by providing optimal arousal levels.[9,10] The autonomic nervous system (ANS) response is primarily mediated by a decrease in the parasympathetic activity rather than an increase in sympathetic activity.[11] Learning under stress is a complex process influenced by the ANS, which enhances alertness and attention.[9] Learning also depends on the timing and context of stress hormone release, namely corticosteroids and adrenaline, and there is an interplay between the neural and endocrine response.[9] The learning arousals that activate the amygdala and hippocampus are initiated by the ANS.[12] Stimulating ANS responses during learning are typically not due to sympathetic activation, which occurs over several minutes, but parasympathetic withdrawal, which occurs in seconds.[13] One of the most reliable measures of parasympathetic activity is respiratory sinus arrhythmia (RSA), which indicates subtle changes in cardiac rate.[14] Medical simulations that result in decreases in RSA, indicating parasympathetic withdrawal, further support the hypothesis that learning arousal contributes to the effectiveness of simulations on memory enhancement.[1,15]

In recent years, health care innovators have used augmented reality (AR) simulations to provide dynamic, interactive scenarios. While virtual reality (VR) immerses users in a completely virtual environment, AR overlays digital information onto the real world, and, unlike VR, AR provides opportunities for direct, participant interaction.[15] AR can incorporate real-world objects such as task trainers and medical equipment to create mixed reality experiences. Gaps in the literature involving AR simulation include the lack of research focusing on educational instructional design that explores the added educational value of these tools.[16] In order to understand the educational benefits of AR simulations, it is imperative to establish the physiologic basis of equity between traditional forms of medical crisis simulation, which partly rely on inducing stressful scenarios that potentiate memories to support learning. Given the advantages of AR simulation and the ANS-mediated memory potentiation underlying simulation's effectiveness, this investigation studied the physiologic responses of participants undergoing AR and traditional, in-person simulations.[1,15] The primary aim was to determine if the participants' parasympathetic responses to an AR simulation were not inferior to the parasympathetic responses of a mannequin-based, in situ simulation. The secondary aims explored the usability, ergonomics, satisfaction, and learning effectiveness of the AR simulation.

Methods

Participants

The investigators conducted this study

continued on next page

continued from previous page

in 2 simulation rooms in a freestanding, academic hospital in Northern California. Health care providers with direct patient interaction and resuscitation training were eligible, including nurses, physician assistants, physicians, and physician trainees. Participants with severe motion sickness, seizure disorder, nausea, or currently using chronotropic heart medications, such as β blockers, were excluded. The participants were similar to those studied in other simulation trials of pediatric medical crisis management.[1,17] In these previous studies, nurses, medical trainees, and faculty physicians underwent in-person simulation training in conjunction with components of the American Heart Association (AHA) Pediatric Advanced Life Support (PALS) training.[17,18] One week prior and 2 days before study enrollment, electronic announcements publicized the opportunity to participate. Financial incentives were not offered. There were no changes to eligibility criteria or methods after trial commencement. The study adhered to CONSORT guidelines and was approved by the Stanford institutional review board (IRB 68663), Clinical Trials number NCT 05674188.[19]

Intervention

Research assistants (RAs) recruited eligible participants. After screening for exclusion criteria, participants provided consent, demographic information, and completed pre-simulation questionnaires. Participants were randomized using an electronic random number generator with 1:1 allocation to either a traditional, mannequin-based simulation or an AR simulation. Each simulation had 1 participant who served as the team leader with the ability to direct 2 actors: an airway role and a combined pharmacist and chest compressor role. Before the simulation, the simulation instructor conducted a scripted orientation that reviewed effective communication skills and the AHA PALS algorithm for cardiac arrest to both groups.[18]

After the instructional overview, the simulation instructor explained the scope of the roles of the actors. All participants were then equipped with chest biometric sensors (BioNomadix RSP&ECG Transmitter with the BioNomadix Logger; BIOPAC Systems, Inc.), which measured parasympathetic nervous system activity via electrocardiogram (ECG) RSA at 1000 Hz.[1] Before beginning the intervention, participants sat quietly to record baseline RSA data for 60 seconds. The biometric sensors remained on throughout the intervention and for an additional 60 seconds after the simulation to record the recovery phase. The BIOPAC system is the industry standard for measuring RSA as they transmit both ECG and respiratory data, reflecting the variation in heart rate throughout the respiratory cycle.[1] ECG alone offers suboptimal source signal for heart rate variability (HRV) analysis, as estimating RSA from an ECG assumes a regular breathing pattern.[20] HRV short-term measurements are obtained by the dynamic relationship between the parasympathetic and sympathetic nervous systems and the regulatory control of heart rate from RSA, the baroreceptor reflex, and modulation of vascular tone.[21]

The traditional, in-person simulation was an apneic, pulseless, pediatric mannequin in ventricular fibrillation that participants discovered on arriving at the patient's bedside. During the simulation, participants directed the actors to initiate cardiopulmonary resuscitation (CPR), perform defibrillation, and administer epinephrine. A simulation instructor modulated the vital signs in response to the participants' actions during the simulation, including converting to spontaneous circulation (ROSC) in the final stages of the simulation. On ROSC, the participants directed the actors to check a pulse.

The same scenario was used for the AR simulation. The RA fitted the AR participants to the Magic Leap 1 (ML1, Magic Leap, Inc.) headset and the instructor oriented the participants. Similar to the in-person group, these participants discovered a pediatric patient in ventricular fibrillation, but instead of a mannequin, they saw the actors along with a holographic patient in addition to a holographic gurney, monitor with vital signs, defibrillator, and bag valve mask. The simulation instructor modulated the holographic vital signs, pulses, and breath sounds in response to participant actions. After ROSC, the participants requested a pulse check. At the end of the traditional and AR simulations, after ROSC, the instructor provided a scripted debrief to each participant. After the debrief, participants completed post-simulation questionnaires. Five months after the study intervention, participants received an email questionnaire to assess longitudinal recall.

Objectives

The primary objective was to determine if the participants' parasympathetic responses to the AR simulation were not inferior to the parasympathetic responses of the traditional simulation. The secondary objectives explored the usability, ergonomics, satisfaction, and learning effectiveness of the AR simulation. The authors hypothesized that the AR group's RSA response would be noninferior to the traditional group. In addition, the authors hypothesized that the AR group would report high levels of usability, ergonomics, and satisfaction and that both groups would score similarly on longitudinal recall questionnaires.

Outcomes

Demographic data were collected before the intervention, including, age, gender, race, level of training, and prior exposure to AR. The primary outcome compared the difference in mean RSA from before to during the simulation and from during to after the simulation between groups. RSA was measured using the chest biometric sensors. Parasympathetic responses were indexed as RSA, the high-frequency heart rate variation controlled by efferent fibers of the vagus nerve during breathing.[14] A trained researcher performed a visual analysis of the RSA data for motion artifact and aberrant identification of ECG R waves using MindWare Technologies, LTD. In alignment with standard protocols, the authors compared RSA group differences in 30-second epochs.[1,22]

Secondary outcomes were measured using the System Usability Scale (SUS) to assess usability, the ISO 9241-400 to assess ergonomics, and the Simulation Design Scale (SDS) to assess user satisfaction[23-25] (Supplemental Online Material, Appendices 1 through 3). The SUS and ISO 9241-400 scales were only administered to the AR group because they were used to

continued on next page

continued from previous page

assess the usability and ergonomics of the headsets. Finally, learning effectiveness was evaluated using an assessment sent to participants 5 months after the simulation (Supplemental Online Material, Appendix 4). The assessment included 10 multiple-choice questions related to key learning points about resuscitation management. Each participant received 5 electronic reminders to complete the assessment and data were stored in Research Electronic Data Capture (REDCap).[26]

Sample Size

The mean RSA for a 35-year-old adult was considered to be 6.81 with a standard deviation (SD) of 1.16.[24] As the SD of RSA in adults is approximately 15%, a difference of less than 10% between groups was considered to be noninferior.[27] A sample size of 106 participants (53 in each group) was calculated based on a power of 90%, type 1 error of 5%, noninferior margin of 10%, and a sampling ratio of 1:1.[27,28]

Randomization Sequence Generation

Participants were randomized using an electronic random number generator with 1:1 allocation using Excel (Microsoft).[29]

Randomization Allocation Concealment and Implementation

At the time of enrollment, after application of exclusion criteria, the RAs used the generator to assign groups. The sequence was concealed until interventions were assigned.

Blinding

Before randomization, participants and researchers were not aware of the participants' group assignment. Participant blinding was not possible given study design. Final analyses were performed by a blinded statistician.

Statistical Methods

To test the differences in mean RSA from before to during the simulation (intervention phase) and from during to after the simulations (recovery phase), a split-plot analysis of variance (ANOVA) was used. Confidence intervals (CIs) were also calculated to compare the differences of change in RSA during the intervention and recovery phases between both groups. For noninferiority to be established for the AR group compared with the traditional group during the intervention phase, the lower limit of the 2-sided 95% CI would need to be higher than the noninferiority margin of −0.681 (10%, or less than 1 SD, of the population mean RSA value of 6.81), equivalent to a 1-sided test with an α value of 0.025. Similarly, noninferiority for the recovery phase would be established if the upper limit of the 2-sided 95% CI was lower than the established margin of 0.681. Descriptive statistics were used to analyze the results of the SUS and the ISO 9241-400. The SDS data were non-normal, so Wilcoxon Rank Sum Tests were used to assess equivalence in mean reported scores for each item. The recall assessment survey was analyzed using a 2-tailed unpaired *t* test assuming unequal variances.

Results

Participant Flow

A total of 111 participants were enrolled. No participants were excluded after exclusion criteria were applied. An RA randomized 56 participants to the AR simulation and 55 participants to the traditional simulation (Figure 1).

Recruitment

Participants were recruited March 20-22, 2023. After the sample size needed to sufficiently power the study's primary outcome was met, recruitment stopped. Follow-up surveys were emailed to participants in August 2023.

Baseline Data

The mean ages for the AR and traditional groups were 38.0 ± 10.5 and 38.3 ± 9.3 years, respectively. The AR group had 71.4% women and 55.4% White participants and in the traditional group, 67.3% were women and 50.9% were White. Nursing was the most common occupation and the second most common was attending physician. Most reported using immersive simulation experiences fewer than 10 times (Table 1).

Numbers Analyzed

One participant in the AR group (1 of 56) and 4 in the traditional group (4 of 55) did not complete the simulation due to schedule conflicts that prohibited them from completing the intervention. The remaining 106 participants completed the study protocol and were analyzed for the primary outcome (Figure 1). There were no protocol deviations and no participants crossed assignments.

Outcomes

In the split-plot ANOVA to assess the difference in mean RSA between the intervention and recovery phases of the simulations, there was no interaction effect between the 2 groups and the phase of the experiment ($P = .270$) (Figure 2). Both groups experienced a decrease in mean RSA from baseline to during the simulation ($P < .001$ in both groups) and an increase in RSA from during the simulation to the recovery period afterward ($P < .001$ and $P = .035$ for AR and traditional groups, respectively).

For the intervention phase, the RSA mean change was −0.400 (95% CI, −0.624 to −0.177; $P < .001$) in the AR group and −0.437 (95% CI, −0.661 to −0.213; $P < .001$) in the traditional group. The difference in change between the 2 groups was −0.037 (95% CI, −0.349, 0.276; $P = .816$). The CI's lower limit, −0.349, was greater than −0.681, thus supporting noninferiority at the .025 significance level (Table 2). The RSA mean change from the recovery phase was 0.481 (95% CI, 0.259–0.704; $P < .001$) in the AR group and 0.243 (95% CI, 0.017–0.469; $P = .035$) in the traditional group. The difference in change between the 2 groups was −0.238 (95% CI, −0.551 to 0.075; $P = .216$). Because the CI's upper limit, 0.075, was lower than 0.681, noninferiority at the .025 significance was supported (Table 2).

The results of secondary outcomes for usability, ergonomics, satisfaction, and learning effectiveness were also analyzed. The mean SUS score of the AR group was 70.5 ± 14.5. The ISO 9241-400 results of the AR group showed that 34 participants (65.38%) reported feeling comfortable using the headset for a long time. Four participants (7.69%) felt that the concentration required to operate the device was very high and 3 (5.77%) participants felt that eye fatigue was very high (Figure 3).

After analyzing the results of the SDS, the only item to return a difference was related to fidelity involving whether the simulations had real-life factors, situations, and variables built into the scenarios. For

continued on next page

continued from previous page

this item, the AR group mean was 4.24 and the traditional group mean was 4.56 (P = .016). For the assessment of learning effectiveness, the assessment questionnaire was completed by 12 participants in the AR group and 15 participants in the traditional group. The average recall score was 74.2% ± 24.7% in the AR group and 66.6% ± 19.9% in the in-person group (P = .4).

Adverse Events

There were no adverse events throughout the study. No participants reported any harm and there were no incidents of motion sickness, seizures, vomiting, or nausea from either the AR or traditional group participants.

Discussion

This study's findings demonstrated that both AR and traditional simulations led to significant changes in RSA associated with parasympathetic responses, with the AR group showing noninferior changes compared with the traditional group. Secondary outcomes indicated favorable usability of the AR technology, evidenced by a mean SUS score of 70.5, and positive ergonomics, as most participants disagreed with the negative aspects related to device comfort in the ISO 9241-400 survey. Satisfaction levels were similar in both AR and traditional groups across most categories, with a slight but significant difference noted in fidelity. In terms of learning effectiveness, we observed no difference between the 2 groups based on the outcomes of the recall questionnaire.

The primary result demonstrated that during the same pediatric cardiac arrest simulation scenario involving an apneic, pulseless child in ventricular fibrillation, the AR simulations evoked a similar degree of parasympathetic withdrawal compared with traditional, in situ simulations. This pivotal result suggests that AR simulations effectively induce the same level of stress that is well-documented in traditional medical simulations involving medical emergencies.[30,31] Stress plays a crucial role in improving knowledge acquisition and the AR simulations effectively generated this necessary response.[6,30,31] This finding substantiates a recent crossover study that also indicated that AR and traditional simulations elicited comparable stress responses.[30] Our study adds to those findings with the incorporation of RSA as a stress biomarker, a method validated by previous studies but not explored in previous AR simulation investigations.[12,14] Given the noninferior stress response, AR simulations may achieve similar learning outcomes observed in those learning under traditional simulations.[30-32]

Although AR simulations are unlikely to replace traditional simulations, the availability of AR as a reasonable alternative is important. The financial implications of integrating AR into a traditional simulation program must be considered. Depending on the degree of immersion, the costs associated with programming and hardware can range from several hundred to thousands of dollars.[33] Traditional mannequins can also range in cost from several hundred to thousands of dollars.[33] However, dedicated spaces such as simulation centers come with additional financial burdens including staff, audiovisual equipment, and simulation technicians, as well as substantial initial and fixed operating costs.[34] Although AR simulations also have ongoing operating costs, these are mainly associated with software and hardware updates, which may be substantially more affordable than traditional mannequin maintenance. In addition, AR simulations, although cost-effective in certain aspects, do require user training, information technology infrastructure, licensing renewals, and hardware updates.

Unlike experiences in head-mounted displays in which the natural world is excluded, AR retains direct eye contact and preserves visual cues, which play essential roles in communication among simulation participants.[1,35] This may explain AR's ability to produce a physiological response similar to that of in-person simulations. These elements of AR simulations that promote nonverbal cues and bolster collaboration are critical for experiential learning.

The secondary outcomes highlight the usability and ergonomics of AR, particularly the ML1 headset, as essential factors for its integration into medical education. The SUS score suggests a good level of usability, which is a positive indicator for future adoption and integration.[36] Furthermore, a high SUS indicates that this device has a high likelihood of integrating well into current simulation curricula without a steep learning curve. This score resonates with findings from other studies that emphasize the usability and ergonomics of similar simulation technologies in health care settings.[37] Ergonomically, the ML1 headset was well-received, mirroring positive user experiences in similar settings with the ML1, where most participants were comfortable using the device for extended periods and few found it mentally taxing to operate.[37] This is reassuring, as the headset is not likely contributing to extraneous cognitive load from prolonged use which would affect the educational experience. These results, capturing both usability and ergonomics, support the ML1 headset's potential for widespread adoption.

Both groups reported high satisfaction in key categories such as objectives and information, support, problem-solving, and feedback, affirming the effectiveness of AR for educational simulations. However, a small but statistically significant difference was observed in the fidelity category, in which AR users reported less realism than traditional, in-person simulations. This was attributed to holograms in lieu of actual resuscitation equipment used in the AR simulations. Despite this, the congruence in satisfaction scores across most categories underscores the participants' positive reception of AR simulations as a viable training tool.

Longitudinal recall, a measure of the impact of these simulations on memory, was expected to show parallel results between the groups, given that both groups experienced similar stress.[12,38] Although only about 1 in 4 participants completed the recall assessment, the knowledge recall was similar between groups. Even though this finding lacked statistical power of the primary aim, it did further corroborate that the AR group's experience was noninferior to the traditional group's experience.

Limitations

This study measured RSA, which is dynamic, and selected 30-second epochs in alignment with standard protocol for RSA analyses.[1,22] However, it is possible that shorter epochs may have altered the findings

continued on next page

continued from previous page

of the study in an unpredictable direction. The study also included a heterogeneous group of health care workers, potentially leading to unmeasured differences and variability that is not fully controlled for, especially in de-escalation techniques. However, we intentionally included a variety of health care workers to increase generalizability. Also, the mean age of RSA used to derive the sample size calculation was not identical to the mean age of these participants but, given negligible changes in resting sympathetic tone during this age period, any effects on the results of this study would be expected to be negligible. An additional limitation included the variability in participants' familiarity with the PALS algorithm that could influence the recall assessment scores as more experienced participants would have greater familiarity with knowledge. Of note, the participants were not asked about any interval PALS education before the recall assessment. Third, the recall assessment was internally developed and not based on a previously validated scale. This was intentionally done to provide a custom assessment tool specific to the scenario and debrief, although without prior validation, it may have lacked construct validity. Finally, the completion of the recall assessment experienced significant attrition, with only 27 participants responding to the survey. Even though this degree of attrition is not uncommon, it did reduce statistical power of that particular result. Future studies are needed to explore longitudinal recall assessments related to long-term memory retention.

Conclusion

Parasympathetic withdrawal during AR simulations were not inferior to the parasympathetic withdrawal during traditional simulations, an essential element for successful learning during medical simulations. The notable usability and ergonomic scores highlight AR simulation's potential as a valuable educational tool in health care. Overall, this study underscores AR simulation's promise in medical education, offering an innovative, immersive, and interactive approach to training health care professionals.

Acknowledgments

The authors would like to thank Mary Hunter, PT, and Laura Wasserman, CCLS, for contributions to recruitment.

Ethical approval was granted by the Stanford University Internal Review Board, #68663.

Previous presentations: Suen MY, Titzler JS, Zuniga-Hernandez M, Jackson C, Rodriguez ST, Wang EY, Pino MR, Rosales O, Caruso TJ, Rama A. The physiologic and emotional effects of augmented reality simulation versus in-person simulation. Poster presented at: American Society of Anesthesiologists Annual Meeting, October 15, 2023; San Francisco, CA.

References

1. Caruso TJ, Armstrong-Carter E, Rama A, et al. The physiologic and emotional effects of 360-degree video simulation on head-mounted display versus in-person simulation: a noninferiority, randomized controlled trial. *Simul Healthc*. 2022;17(1):e105-12.
2. Bienstock J, Heuer A. A review on the evolution of simulation-based training to help build a safer future. *Medicine (Baltimore)*. 2022;101(25):e29503.
3. Weinstock PH, Kappus LJ, Garden A, Burns JP. Simulation at the point of care: reduced-cost, in situ training via a mobile cart. *Pediatr Crit Care Med*. 2009;10(2):176-81.
4. Henckens MJAG, Hermans EJ, Pu Z, Joëls M, Fernández G. Stressed memories: how acute stress affects memory formation in humans. J Neurosci. 2009;29(32):10111-9.
5. LeBlanc VR, Posner GD. Emotions in simulation-based education: friends or foes of learning? *Adv Simul (Lond)*. 2022;7(1):3.
6. Lupien SJ, Maheu F, Tu M, Fiocco A, Schramek TE. The effects of stress and stress hormones on human cognition: implications for the field of brain and cognition. *Brain Cogn*. 2007;65(3):209-37.
7. Bahtiyar S, Gulmez Karaca K, Henckens MJAG, Roozendaal B. Norepinephrine and glucocorticoid effects on the brain mechanisms underlying memory accuracy and generalization. *Mol Cell Neurosci*. 2020;108:103537.
8. Kocialkowski C, Rybinski B, Hainsworth L, Yang D. Simulation learning: effectiveness and stressfulness in medical student teaching. *MedEdPublish*. 2020;9:143.
9. Joëls M, Pu Z, Wiegert O, Oitzl MS, Krugers HJ. Learning under stress: how does it work? *Trends Cogn Sci*. 2006;10(4):152-8.
10. Bong CL, Fraser K, Oriot D. Cognitive load and stress in simulation. In: Grant VJ, Cheng A, eds. *Comprehensive Healthcare Simulation: Pediatrics*. Cham, Switzerland: Springer International Publishing; 2016:3-17.
11. Müller MS, Vyssotski AL, Yamamoto M, Yoda K. Heart rate variability reveals that a decrease in parasympathetic ('Rest-and-digest') activity dominates autonomic stress responses in a free-living seabird. *Comp Biochem Physiol A Mol Integr Physiol*. 2017;212:117-26.
12. Beauchaine TP. Respiratory sinus arrhythmia: a transdiagnostic biomarker of emotion dysregulation and psychopathology. *Curr Opin Psychol*. 2015;3:43-7.
13. Hastings PD, Sullivan C, McShane KE, et al. Parental socialization, vagal regulation, and preschoolers' anxious difficulties: direct mothers and moderated fathers. *Child Dev*. 2008;79(1):45-64.
14. Rash JA, Aguirre-Camacho A. Attention-deficit hyperactivity disorder and cardiac vagal control: a systematic review. *Atten Deficit Hyperact Disord*. 2012;4(4):167-77.
15. Kassutto SM, Baston C, Clancy C. Virtual, augmented, and alternate reality in medical education: socially distanced but fully immersed. *ATS Sch*. 2021;2(4):651-64.
16. Asoodar M, Janesarvatan F, Yu H, de Jong N. Theoretical foundations and implications of augmented reality, virtual reality, and mixed reality for immersive learning in health professions education. *Adv Simul (Lond)*. 2024;9(1):36.
17. Grant VJ, Cheng A, eds. *Comprehensive Healthcare Simulation: Pediatrics*. Cham, Switzerland: Springer International Publishing; 2016.
18. Duff JP, Topjian AA, Berg MD, et al. 2019 American Heart Association focused update on pediatric advanced life support: an update to the American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. *Pediatrics*. 2020;145(1):e20191361.
19. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726-32.
20. Sobron A, Romero I, Lopetegi T. Evaluation of methods for estimation of respiratory frequency from the ECG. *Comput Cardiol*. 2010;37:513-6.
21. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. *Front Public Health*. 2017;5:258.
22. Malik M. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Eur Heart J*. 1996;17(3):354-81.
23. Jordan PW, ed. *Usability Evaluation in Industry*. London, England: Taylor & Francis; 1996.
24. Ingrassia PL, Mormando G, Giudici E, et al. Augmented reality learning environment for basic life support and defibrillation training: usability study. *J Med Internet Res*. 2020;22(5):e14910.
25. National League for Nursing. Simulation Design Scale© (Student Version). 2005. Available from: https://www.nln.org/education/teaching-resources/tools-and-instruments

continued on next page

continued from previous page

26. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-81.
27. Tsypes A, James KM, Woody ML, et al. Resting respiratory sinus arrhythmia in suicide attempters. *Psychophysiology*. 2018;55(2):10.1111/psyp.12978.
28. Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. *J Gen Intern Med*. 2011;26(2):192-6.
29. Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomize? *BMJ*. 1999;318(7192):1209.
30. Price JW, Price JR, Pratt DD, Collins JB, McDonald J. High-fidelity simulation in anesthesiology training: a survey of Canadian anesthesiology residents' simulator experience. *Can J Anaesth*. 2010;57(2):134-42.
31. Demaria S, Bryson EO, Mooney TJ, et al. Adding emotional stressors to training in simulated cardiopulmonary arrest enhances participant performance. *Med Educ*. 2010;44(10):1006-15.
32. Toohey S, Wray A, Hunter J, et al. Comparing the psychological effects of manikin-based and augmented reality-based simulation training: within-subjects crossover study. *JMIR Med Educ*. 2022;8(3):e36447.
33. Alam F, Matava C. A new virtual world? The future of immersive environments in anesthesiology. *Anesth Analg*. 2022;135(2):230-8.
34. Hippe DS, Umoren RA, McGee A, Bucher SL, Bresnahan BW. A targeted systematic review of cost analyses for implementation of simulation-based education in healthcare. *SAGE Open Med*. 2020;8:2050312120913451.
35. Kleinke CL. Gaze and eye contact: a research review. *Psychol Bull*. 1986;100(1):78-100.
36. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Hum-Comput Interact*. 2008;24(6):574-94.
37. Friedman N, Zuniga-Hernandez M, Titzler J, et al. Prehospital pediatric emergency training using augmented reality simulation: a prospective, mixed methods study. *Prehosp Emerg Care*. 2024;28(2):271-81.
38. Yoo HH, Yune SJ, Im SJ, Kam BS, Lee SY. Heart rate variability-measured stress and academic achievement in medical students. *Med Princ Pract*. 2021;30(2):193-200.

The following authors are in the Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University School of Medicine, Stanford, CA: ***Asheen Rama*** *is Clinical Assistant Professor;* ***Ellen Y. Wang***, ***Samuel T. Rodriguez***, *and* ***Thomas J. Caruso*** *are Clinical Professors; and* ***Christian Jackson*** *is a Statistician.* ***Marcos S. Rojas-Pino*** *and* ***Oswaldo Rosales*** *are Doctoral Candidates, Stanford University Graduate School of Education, Stanford, CA. The following authors are in the Stanford Chariot Program, Lucile Packard Children's Hospital Stanford, Palo Alto, CA:* ***Man Yee Suen*** *is a Clinical Research Coordinator,* ***Janet Titzler*** *and* ***Michelle Zuniga-Hernandez*** *are Clinical Research Assistants, and* ***Faith Collins*** *is a Laboratory Manager and Teacher.*

Corresponding author: *Thomas J. Caruso, Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University School of Medicine, 453 Quarry Road, MC 5663, Palo Alto, CA 94305. Telephone (650) 723-5728, Fax: (650) 725-8544*

Email address: *Thomas J. Caruso: tjcaruso@stanford.edu*

Note: *Work should be attributed to Stanford University School of Medicine, Department of Anesthesiology, Perioperative, and Pain Medicine.*

Financial Disclosure: *The authors received a $20,000 internal grant from the Stanford Chariot Program.*

Other Disclosures: *Thomas J. Caruso, Ellen Y. Wang, and Samuel T. Rodriguez are on the board of a nonprofit (Invincikids, Inc.) that seeks to distribute immersive technology to hospitalized children at no cost. They receive no direct or indirect compensation for this role. Thomas J. Caruso, Ellen Y. Wang, and Samuel T. Rodriguez direct the Stanford Chariot Program, which has received philanthropic donations from Meta, Inc. and Magic Leap, Inc. All other authors declare no competing interests.*

Abstract

Background: Traditional medical simulations leverage stressful scenarios to potentiate memory. Augmented reality (AR) simulations provide cost-effective experiences using holograms instead of mannequins. This study investigated the physiologic response to AR simulations.

Methods: This was a noninferiority, controlled trial at an academic, pediatric hospital in Northern California among health care workers randomized to AR or traditional, in situ medical simulations. The primary outcome investigated parasympathetic tone. Biometric sensors assessed parasympathetic tone as respiratory sinus arrhythmia (RSA). A difference in RSA of less than 10% between groups was considered noninferior. Secondary outcomes explored usability, ergonomics, satisfaction, and recall with the System Usability Scale (SUS), ISO 9241-400, Simulation Design Scale (SDS), and an electronic questionnaire 5 months after the intervention, respectively.

Results: A total of 111 participants were enrolled and 106 analyzed. Both groups experienced a decrease in mean RSA from baseline to during the simulation ($P < .001$ for both groups). Subsequently, there was an increase in RSA from the simulation period to the recovery period ($P < .001$ for the AR group and $P = .035$ for the traditional group). Regarding secondary outcomes, the mean SUS score of 70.5 suggested good usability, 65.38% of AR participants reported feeling comfortable using the headset, and satisfaction in both groups was similar except for differences in use of real-life factors. The recall assessment was completed by 12 AR and 15 traditional participants, with similar scores between the 2 groups ($P = .4$).

Conclusions: AR simulations produced a noninferior change in parasympathetic tone compared with traditional simulations. Future investigations may explore the effectiveness of AR simulations for developing nontechnical skills during remote training. (Registration: Clinical Trials Registry NCT 05674188.)

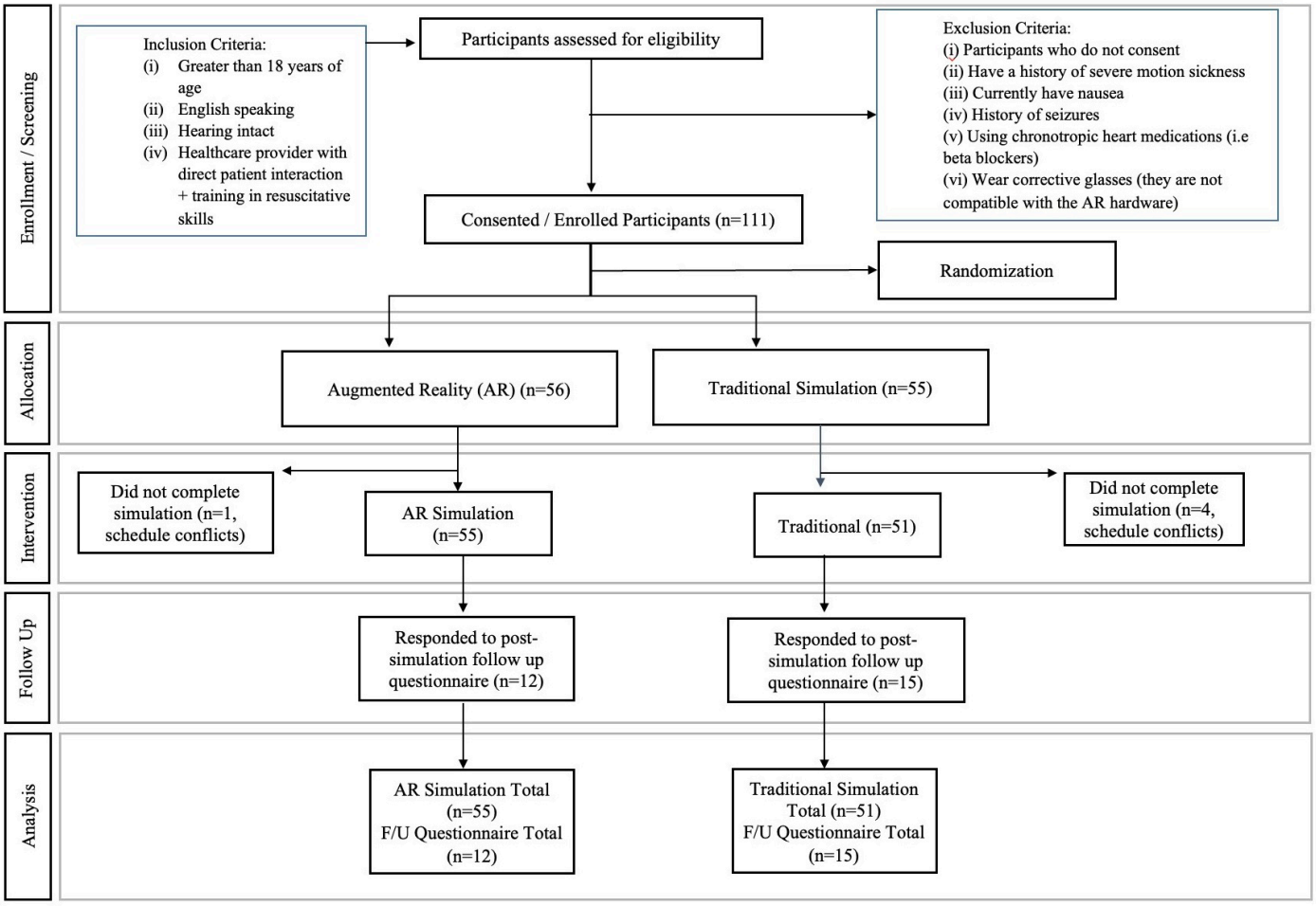
Keywords: Simulation, augmented reality, education, anesthesiology, immersive technologies

continued on next page

continued from previous page

Figures

Figure 1. *Flow diagram of study participants. The number of participants from initial enrollment to randomization in addition to follow-up (F/U) is visualized. A total of 111 participants met inclusion and exclusion criteria and were enrolled. A total of 106 participants were included in the study, as 5 participants did not complete the simulation scenarios due to scheduling conflicts. Twenty-seven participants completed the 5-month follow-up Longitudinal Recall Assessment questionnaire. Abbreviation: AR, augmented reality.*

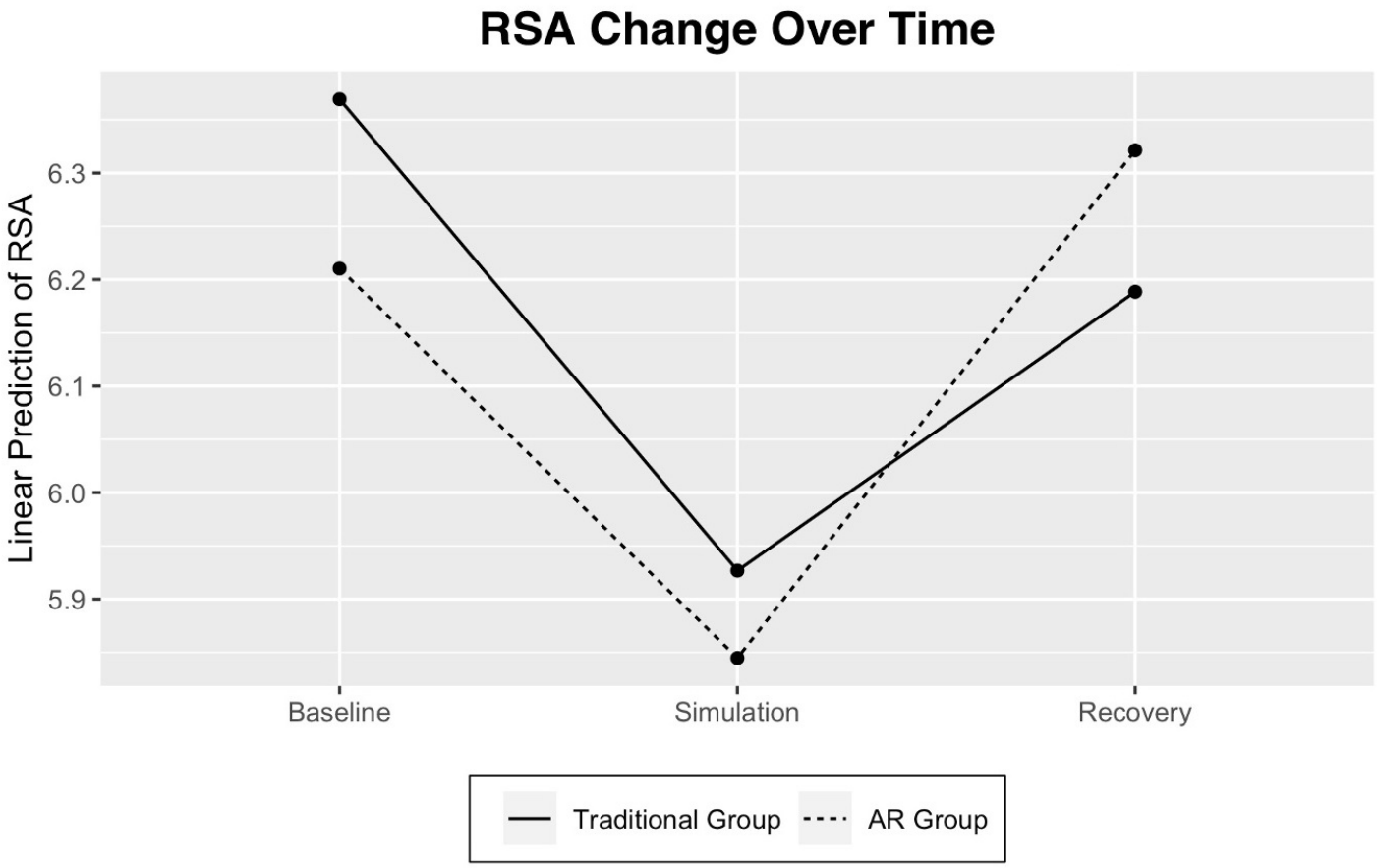


continued on next page

continued from previous page

Figures *continued*

Figure 2. *Trend in respiratory sinus arrhythmia (RSA) changes between augmented reality (AR) and traditional groups. A split-plot analysis of variance was used to assess the difference in mean RSA. The linear prediction of RSA refers to the mathematical model used to estimate RSA based on linear relationships between physiological signals (heart rate variability and respiratory signal). Both groups displayed a significant decrease in mean RSA from baseline to during the simulation followed by a significant increase in mean RSA from the end of the simulation to the recovery phase. These findings established noninferiority between the AR and traditional groups.*

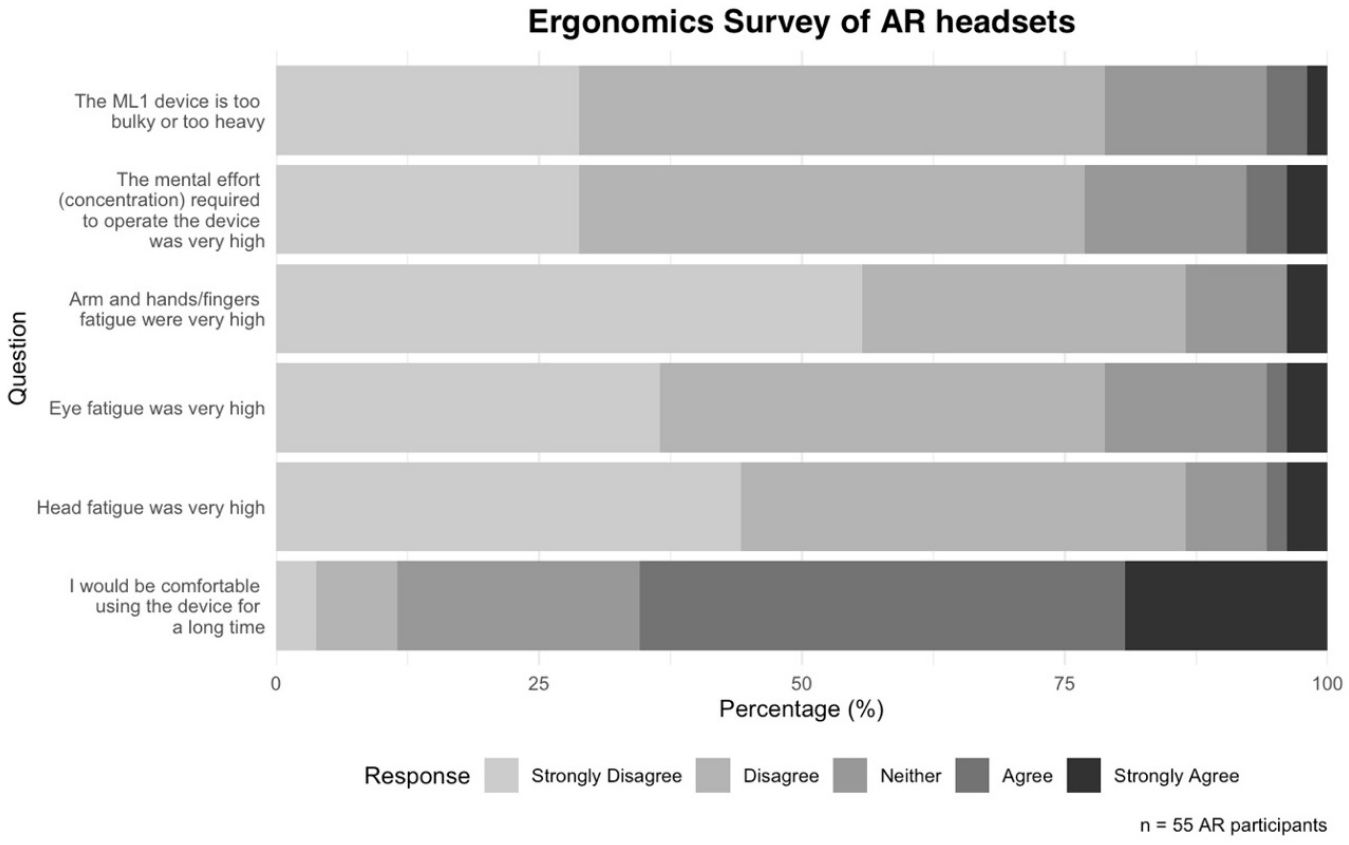


continued on next page

continued from previous page

Figures *continued*

Figure 3. *Ergonomics survey of augmented reality (AR) headsets. The ISO-9241-400 was used to assess ergonomic factors of human-AR interaction. Thirty-four participants (65.38%) reported feeling comfortable using the headset for a long time. Four participants (7.69%) felt that the concentration required to operate the device was very high and 3 (5.77%) participants felt that eye fatigue was very high.*



continued on next page

continued from previous page

Tables

Table 1. *Participant Demographics*

Characteristic	Value	
Cohort	**AR**	**Traditional**
Mean age, SD	38.0 ± 10.5	38.3 ± 9.3
Sex, n (%)		
Male	15 (26.8)	18 (32.7)
Female	40 (71.4)	37 (67.3)
Unknown/Chose not to disclose	1 (1.8)	0
Race,[a] n (%)		
American Indian or Alaskan Native	1 (1.8)	1 (1.8)
Asian	16 (28.6)	16 (29.1)
Black or African American	5 (8.9)	7 (12.7)
Native Hawaiian or Other Pacific	0	0
White	31 (55.4)	28 (50.9)
Unknown / Chose not to disclose	3 (5.4)	3 (5.5)
Other	3 (5.4)	3 (5.5)
Ethnicity, n (%)		
Hispanic or Latino	7 (12.5)	3 (5.5)
Not Hispanic or Latino	47 (83.9)	48 (87.3)
Unknown/Chose not to disclose	2 (3.6)	4 (7.3)
Occupation, n (%)		
Nurse	20 (35.7)	22 (40.0)
Physician	15 (26.8)	13 (23.6)
Physician trainee	10 (17.9)	10 (18.2)
Respiratory therapist	2 (3.6)	3 (5.5)
Other (eg, nurse practitioner, imaging tech, physician assistant)	9 (16.1)	7 (12.7)
Previous exposure to AR, n (%)		
0 times	30 (53.6)	26 (47.3)
1-2 times	22 (39.3)	19 (34.5)
3-5 times	4 (7.1)	7 (12.7)
6-10 times	0	0
>10 times	0	3 (5.5)
Level of resuscitation certification,[a] n (%)		
None	1 (1.8)	0
Basic Life Support (BLS)	38 (67.9)	38 (69.1)
Advanced Cardiovascular Life Support (ACLS)	36 (64.3)	39 (70.9)
Pediatric Advanced Life Support (PALS)	42 (75.0)	41 (74.5)
Neonatal Resuscitation Program (NRP)	13 (23.2)	11 (20.0)

continued on next page

continued from previous page

Tables *continued*

No. times providing resuscitative efforts on a person, n (%)		
0 times	9 (16.1)	9 (16.4)
1-2 times	10 (17.9)	13 (23.6)
3-5 times	13 (23.2)	10 (18.2)
6-10 times	6 (10.7)	5 (9.1)
>10 times	29 (32.1)	18 (32.7)
Received training on effective communication skills during resuscitation, n (%)		
Yes	53 (94.6)	50 (90.9)
No	3 (5.4)	5 (9.1)
Previously worked as a front-line health care worker with direct contact to patients who were critically ill and in need of resuscitation, n (%)		
Yes	51 (91.1)	47 (85.5)
No	5 (8.9)	8 (14.5)
If yes, for how long did you work in this context?		
1-6 mo	2 (3.9)	3 (6.4)
7-11 mo	1 (2.0)	0
1-2 y	7 (13.7)	2 (4.3)
>3 y	41 (80.4)	42 (89.4)

[a] Multiple answers allowed.

Table 2. *Mean RSA Changes During Intervention and Recovery Phases*[a]

	Change in Mean RSA During Intervention		Change in Mean RSA During Recovery	
	Mean (95% CI)	*P*	Mean (95% CI)	*P*
Traditional	−0.437 (−0.661 to −0.213)	<.001	0.243 (0.017 to 0.469)	.035
AR	−0.400 (−0.624 to −0.177)	<.001	0.481 (0.259 to 0.704)	<.001
Difference	−0.037 (−0.349 to 0.276)	.816	−0.238 (−0.551 to 0.075)	.216

[a] Intervention occurred from baseline to during the simulation. Recovery occurred during to after the simulation.

continued on next page

continued from previous page

Supplemental Online Material

Appendix 1. Document of System Usability Scale

System Usability Scale

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

continued on next page

continued from previous page

Supplemental Online Material *continued*

Appendix 2. Document of Ergonomics Survey ISO 9241-400

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
	1	2	3	4	5
The ML1 device is too bulky or too heavy					
The mental effort (concentration) required to operate the device was very high					
Arm and hands/fingers fatigue were very high					
Eye fatigue was very high					
Head fatigue was very high					
I would be comfortable using the device for a long time					

continued on next page

continued from previous page

Supplemental Online Material *continued*

Appendix 3. Document of Simulation Design Scale

Simulation Design Scale (Student Version)

In order to measure if the best simulation design elements were implemented in your simulation , please complete the survey below as you perceive it. There are no right or wrong answers, only your perceived amount of agreement or disagreement. Please use the following code to answer the questions.

Use the following rating system when assessing the simulation design elements:

1 - Strongly Disagree with the statement
2 - Disagree with the statement
3 - Undecided - you neither agree or disagree with the statement
4 - Agree with the statement
5 - Strongly Agree with the statement
NA - Not Applicable; the statement does not pertain to the simulation activity performed.

Rate each item based upon how important that item is **to you.**

1 - Not Important
2 - Somewhat Important
3 - Neutral
4 - Important
5 - Very Important

Item	1	2	3	4	5	NA	1	2	3	4	5
Objectives and Information											
1. There was enough information provided at the begining of the simulation to provide direction and encouragement.	O 1	O 2	O 3	O 4	O 5	O NA	O 1	O 2	O 3	O 4	O 5
2. I clearly understood the purpose and objectives of the simulation.	O 1	O 2	O 3	O 4	O 5	O NA	O 1	O 2	O 3	O 4	O 5
3. The simulation provided enough information in a clear matter for me to problem-solve the situation.	O 1	O 2	O 3	O 4	O 5	O NA	O 1	O 2	O 3	O 4	O 5
4. There was enough information provided to me during the simulation.	O 1	O 2	O 3	O 4	O 5	O NA	O 1	O 2	O 3	O 4	O 5
5. The cues were appropriate and geared to promote my understanding.	O 1	O 2	O 3	O 4	O 5	O NA	O 1	O 2	O 3	O 4	O 5
Support											
6. Support was offered in a timely manner.	O 1	O 2	O 3	O 4	O 5	O NA	O 1	O 2	O 3	O 4	O 5
7. My need for help was recognized.	O 1	O 2	O 3	O 4	O 5	O NA	O 1	O 2	O 3	O 4	O 5
8. I felt supported by the teacher's assistance during the simulation.	O 1	O 2	O 3	O 4	O 5	O NA	O 1	O 2	O 3	O 4	O 5
9. I was supported in the learning process.	O 1	O 2	O 3	O 4	O 5	O NA	O 1	O 2	O 3	O 4	O 5

continued on next page

continued from previous page

Supplemental Online Material *continued*

Simulation Design Scale (Student Version)

Use the following rating system when assessing the simulation design elements:

1 - Strongly Disagree with the statement
2 - Disagree with the statement
3 - Undecided - you neither agree or disagree with the statement
4 - Agree with the statement
5 - Strongly Agree with the statement
NA - Not Applicable; the statement does not pertain to the simulation activity performed.

Rate each item based upon how important that item is **to you.**

1 - Not Important
2 - Somewhat Important
3 - Neutral
4 - Important
5 - Very Important

Item	1	2	3	4	5	NA	1	2	3	4	5
Problem Solving											
10. Independent problem-solving was facilitated.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5
11. I was encouraged to explore all possibilities of the simulation.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5
12. The simulation was designed for my specific level of knowledge and skills.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5
13. The simulation allowed me the opportunity to prioritize nursing assessments and care.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5
14. The simulation provided me an opportunity to goal set for my patient.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5
Feedback/Guided Reflection											
15. Feedback provided was constructive.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5
16. Feedback was provided in a timely manner.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5
17. The simulation allowed me to analyze my own behavior and actions.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5
18. There was an opportunity after the simulation to obtain guidance/feedback from the teacher in order to build knowledge to another level.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5
Fidelity (Realism)											
19. The scenario resembled a real-life situation.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5
20. Real life factors, situations, and variables were built into the simulation scenario.	○ 1	○ 2	○ 3	○ 4	○ 5	○ NA	○ 1	○ 2	○ 3	○ 4	○ 5

continued on next page

continued from previous page

Supplemental Online Material *continued*

Appendix 4. Document of Longitudinal Recall Assessment

1)	1. What of the following does NOT determine cardiac output?	◯ Blood pressure ◯ Heart rate ◯ Heart rhythm ◯ Stroke volume
2)	2. Which is NOT a way of ensuring high quality chest compressions?	◯ Using a backboard ◯ Ensuring adequate chest recoil ◯ Aiming for a rate between 100-120 ◯ Performing a pulse check for less than 20 seconds
3)	3. What is the mechanism of action of epinephrine?	◯ Pulmonary vasoconstriction ◯ Coronary vessel dilation ◯ α 1 agonist / β 1 agonist ◯ α 2 agonist / β 2 agonist
4)	4. What energy is used for defibrillation for the first shock during pediatric cardiac arrest?	◯ 1 J/kg ◯ 2 J/kg ◯ 4 J/kg ◯ 0.5 J/kg
5)	5. What is the code dose of epinephrine?	◯ 0.01mcg/kg ◯ 10 mcg/kg ◯ 1 mcg/kg ◯ 0.001mg/kg
6)	6. What is the in-house phone number used to call a code blue?	◯ 19211 ◯ 211 ◯ 911 ◯ 78000
7)	7. Which of the following is the most common cause of pediatric cardiac arrest in a hospitalized patient?	◯ Coronary artery disease ◯ Pulmonary embolism ◯ Respiratory failure ◯ Stroke
8)	8. A code blue should only be called if a patient does not have a pulse.	◯ True ◯ False
9)	9. H's and T's refer to a mnemonic for possible reversible causes of cardiac arrest. Which of the following is not an H/T used in this mnemonic?	◯ Hypoxia ◯ Hyperchloridemia ◯ Hypothermia ◯ Hypernatremia ◯ Cardiac tamponade ◯ Tension pneumothorax ◯ Toxins
10)	10. Beer's law is a physics principle underlying the science of pulse oximetry. Which of the following represents the equation for Beer's law?	◯ Absorbance is proportional to the path length, through the sample and the concentration of the absorbing species ◯ Absorbance is inversely proportional to the path length, through the sample and the concentration of the absorbing species ◯ Concentration is inversely proportional to the path length, through the sample and the absorbance of the medium ◯ Absorbance is inversely proportional to length and proportional to the concentration of the absorbing medium