



The effect of simulator fidelity on acquiring non-technical skills: a randomized non-inferiority trial

L'effet de la fidélité du simulateur sur l'acquisition d'habiletés non techniques: essai randomisé de non-infériorité

Yuqi Gu, BSc, MD, FRCPC · Tobias Witter, MD, FRCPC ·
Patty Livingston, MD, FRCPC, MEd · Purnima Rao, BSc, MD, FRCPC ·
Terry Varshney, MDCM, FRCPC · Tom Kuca, BSc, MD, FRCPC ·
M. Dylan Bould, MB, ChB, MRCP, FRCA, MEd

Received: 19 December 2016/Revised: 11 August 2017/Accepted: 13 September 2017/Published online: 6 October 2017
© Canadian Anesthesiologists' Society 2017

Abstract

Introduction As simulator fidelity (i.e., realism) increases from low to high, the simulator more closely resembles the real environment, but it also becomes more expensive. It is generally assumed that the use of high-fidelity simulators results in better learning; however, the effect of fidelity on learning non-technical skills (NTS) is unknown. This was a

non-inferiority trial comparing the efficacy of high- vs low-fidelity simulators on learning NTS.

Methods Thirty-six postgraduate medical trainees were recruited for the trial. During the pre-test phase, the trainees were randomly assigned to manage a scenario using either a high-fidelity simulator (HFS) or a low-fidelity simulator (LFS), followed by expert debriefing. All trainees then underwent a video recorded post-test scenario on a HFS, and the NTS were assessed between the two groups. The primary outcome was the overall post-test Ottawa Global Rating Scale (OGRS), while controlling for overall pre-test OGRS scores. Non-inferiority between the LFS and HFS was based on a non-inferiority margin of greater than 1.

Results For our primary outcome, the mean (SD) post-test overall OGRS score was not significantly different between the HFS and LFS groups after controlling for pre-test overall OGRS scores [3.8 (0.9) vs 4.0 (0.9), respectively; mean difference, 0.2; 95% confidence interval, -0.4 to 0.8; $P = 0.48$]. For our secondary outcomes, the post-test total OGRS score was not significantly different between the HFS and LFS groups after controlling for pre-test total OGRS scores ($P = 0.33$). There were significant improvements in mean overall ($P = 0.01$) and total ($P = 0.003$) OGRS scores from pre-test to post-test. There were no significant associations between postgraduate year ($P = 0.82$) and specialty ($P = 0.67$) on overall OGRS performance.

Conclusion This study suggests that low-fidelity simulators are non-inferior to the more costly high-fidelity simulators for teaching NTS to postgraduate medical trainees.

Y. Gu, BSc, MD, FRCPC (✉) · T. Witter, MD, FRCPC ·
P. Livingston, MD, FRCPC, MEd · T. Kuca, BSc, MD, FRCPC
Department of Anesthesia, Pain Management, and Perioperative
Medicine, QEII Health Sciences Centre, Dalhousie University,
Halifax Infirmary Site, 1796 Summer Street, Halifax, NS B3H
3K9, Canada
e-mail: guy@dal.ca

Y. Gu, BSc, MD, FRCPC · P. Rao, BSc, MD, FRCPC
Department of Anesthesiology and Pain Medicine, The Ottawa
Hospital, Ottawa, ON, Canada

Y. Gu, BSc, MD, FRCPC · M. Dylan Bould, MB, ChB, MRCP,
FRCA, MEd
Department of Innovation in Medical Education,
University of Ottawa, Faculty of Medicine, Ottawa,
ON, Canada

T. Witter, MD, FRCPC
Department of Critical Care, Dalhousie University, Halifax, NS,
Canada

T. Varshney, MDCM, FRCPC
Department of Pediatrics, Pediatric Emergency, Children's
Hospital of Eastern Ontario, Ottawa, ON, Canada

M. Dylan Bould, MB, ChB, MRCP, FRCA, MEd
Department of Anesthesiology, Children's Hospital of Eastern
Ontario, Ottawa, ON, Canada

Résumé

Introduction Quand la fidélité (c'est-à-dire le réalisme) augmente, passant de basse à haute, le simulateur ressemble de plus en plus à l'environnement réel, mais devient aussi plus coûteux. On suppose généralement que l'utilisation de simulateurs très réalistes débouchent sur un meilleur apprentissage; on ignore toutefois l'effet du réalisme sur l'apprentissage des habiletés non techniques (HNT). Il s'agit ici d'une étude de non-infériorité comparant l'efficacité de simulateurs haute fidélité et basse fidélité sur l'apprentissage des HNT.

Méthodes Trente-six médecins stagiaires ont été recrutés pour cette étude. Au cours de la phase prétest, les stagiaires ont été répartis aléatoirement pour gérer un scénario utilisant un simulateur « haute fidélité » (SHF) ou un simulateur « basse fidélité » (SBF), suivi d'une séance de débriefing par un expert. Tous les stagiaires ont alors géré un scénario post test sur un SHF enregistré sur vidéo, puis les HNT ont été évaluées dans les deux groupes. Le critère d'évaluation principal était le résultat sur l'échelle d'évaluation globale d'Ottawa (OGRS—Ottawa Global Rating Scale) post test, tout en contrôlant les résultats globaux de l'OGRS prétest. La non-infériorité entre le SBF et le SHF reposait sur une marge de non-infériorité supérieure à 1.

Résultats Concernant notre critère d'évaluation principal, le score global moyen [ÉT] post test de l'OGRS était pas significativement différent entre les groupes SHF et SBF après contrôle pour les scores globaux prétests de l'OGRS (respectivement, 3,8 [0,9] contre 4,0 [0,9]; différence des moyennes, 0,2; intervalle de confiance à 95% : -0,4 à 0,8; $P = 0,48$). Concernant nos critères d'évaluation secondaires, le score total post test de l'OGRS était pas significativement différent entre les groupes SHF et SBF après contrôle pour les scores totaux prétests de l'OGRS ($P = 0,33$). Il y a eu des améliorations significatives dans les scores globaux moyens ($P = 0,01$) et totaux ($P = 0,003$) de l'OGRS entre le prétest et le post test. Aucune association significative n'a été trouvée avec l'année d'étude ($P = 0,82$) ou la spécialité ($P = 0,67$) et la performance globale à l'OGRS.

Conclusion Cette étude suggère que les simulateurs basse fidélité ne sont pas inférieurs aux simulateurs haute fidélité, plus coûteux, pour l'enseignement des HNT aux médecins en formation.

healthcare involves a number of factors, including lower tolerance for medical errors and greater emphasis on patient safety, rapid advances in medical technologies for diagnosis and management, caring for more medically complex patients, evolving models for postgraduate education and healthcare delivery, and recognizing the need for deliberate practice to achieve excellence in patient care.^{1,2} A few studies, including systematic reviews and meta-analyses, have established that technology-enhanced simulation can be superior to traditional teaching methods for learning new knowledge, acquiring skills, and increasing trainee satisfaction.^{3,4}

Driven by the mounting evidence for technology-enhanced simulation, in conjunction with the need to minimize harm to patients, many academic centres have purchased high-fidelity simulators for their respective training programs.⁵ Nevertheless, there is a significant cost associated with purchasing and maintaining high-fidelity simulators, and the monetary to educational ratio of high-fidelity simulators is still unclear when compared with other teaching modalities.^{6,7}

Within the simulation literature, the term “fidelity” describes how the appearance and behaviour of the simulator match those of the real environment, thus a high-fidelity simulator is considered most realistic.⁸ As the simulator fidelity increases from low to high, the technology becomes more advanced and sophisticated, which invariably results in higher cost. Historically, it was assumed that improving the fidelity of the simulator would result in more effective learning.⁹ Moreover, studies have shown that simulator fidelity does not improve learning many of the basic motor or technical skills.⁹ This suggests that large financial investments in purchasing high-fidelity simulators may not always equate to improved learning for certain skills.

Non-technical skills (NTS) are defined as the cognitive (e.g., decision-making, situation awareness) and interpersonal (e.g., communication, leadership) skills important for reducing medical error and improving patient safety.^{10–12} High-fidelity simulators have been shown to be effective in teaching NTS.¹³ In addition, some previous investigations have compared high- vs low-fidelity simulators in teaching NTS, but these studies were limited either by having a small convenience sample or by comparing high- vs low-fidelity models with no actual difference in cost.^{14,15} It is often assumed that learning complex skills, such as NTS, demands an equally complex or high-fidelity simulator.⁹ With limited evidence to support this assumption, the purpose of our study was to compare the effectiveness of a low- vs high-fidelity simulator in teaching NTS to postgraduate medical trainees. If our study results were to show the same teaching effectiveness between the two types of simulators

Simulation in medical education re-creates components of clinical encounters for the purpose of training or assessment.¹ The growing requirement for simulation in

(i.e., non-inferior), then the lower cost simulator might have an overall advantage. Accordingly, we hypothesized that a low-fidelity simulator would be non-inferior to a high-fidelity simulator for teaching NTS.

Methods

Recruitment

After institutional research ethics board (Capital District Health Authority, Halifax, NS, CDHA-RS/2014-262, March 2014) approval, 36 postgraduate year (PGY) 1-5 residents were recruited from training programs at Dalhousie University—i.e., anesthesiology, emergency medicine, internal medicine, or a surgical specialty. One of the study investigators (Y.G.) conducted recruitment through e-mail and in person during August 2014. Written and verbal informed consent was obtained from all participants and demographic data were collected. The intention to evaluate only NTS was not initially disclosed to the participants; instead, we informed them that the purpose of our study was to assess whether simulator fidelity affects learning outcomes. We chose this approach in order to minimize any potential Hawthorne effect where participants would act differently if they knew that only NTS were being assessed. Full disclosure was provided once data collection was complete. Participants received a small honorarium for their participation, which was not revealed until their pre-briefing on data collection day.

Study design (Fig. 1)

Our hypothesis that simulator fidelity will not affect learning NTS was tested as a randomized-controlled non-inferiority trial. The participants were stratified by junior (PGY 1-2) and senior (PGY 3-5) residents and randomly assigned to either the high-fidelity simulator (HFS) or the low-fidelity simulator (LFS) group using a random number generator (Randomness and Integrity Services Ltd., Dublin, Ireland). The resulting assignments were placed in sequentially numbered opaque envelopes according to resident stratification (i.e., junior and senior residents). Following recruitment, one of the study investigators (Y.G.) opened the envelopes to reveal the group allocation. The participant then chose one of several potential times and dates designated to either the HFS or the LFS group. Within their fidelity group, the participants were assigned to teams of three. The teams were arranged by convenience due to the complexity of scheduling residents and the availability of the simulator facility and research assistants. A research volunteer acted as the third confederate for the sessions where only two participants

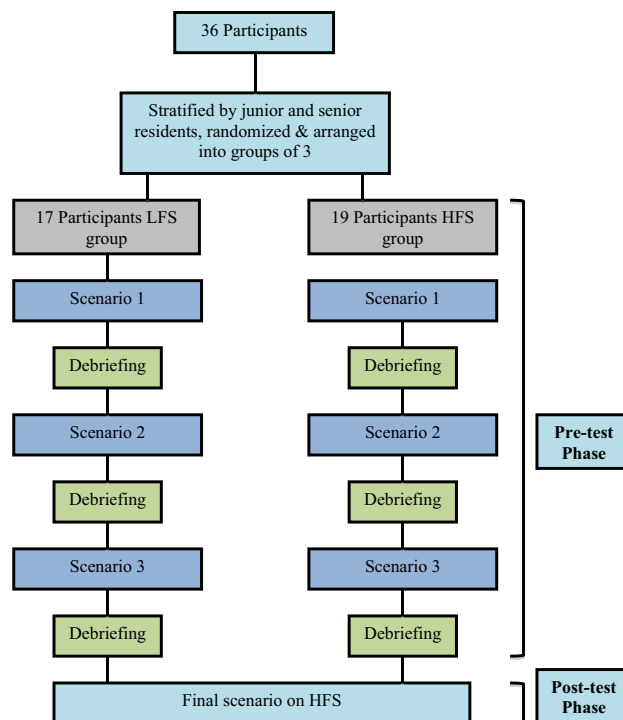


Fig. 1 Representation of the experimental design on both simulators. HFS = High-fidelity simulator; LFS = Low-fidelity simulator

were present. Before each scenario, one of the investigators (Y.G.) conducted a standardized pre-briefing to allow the participants to familiarize themselves with the simulator environment and equipment, address any limitations, broadly discuss the goals and objectives of the scenarios, and address any questions or concerns.

Pre-test phase

There were three different simulation scenarios (described below in the scenario section) followed by a structured debriefing. For each scenario, one participant of the team was evaluated (in the “hot-seat”), while the other two participants acted as confederates—e.g., nurse, respiratory technician, surgeon, etc., depending on the scenario. After completion of the scenario, all team participants were debriefed (described below). Following the debriefing, the participants rotated and one of the other confederates entered the “hot-seat”. The simulation and debriefing process was repeated with a different scenario and then repeated a third time for the final participant. The confederates were provided with pre-defined instructions and scripts on how to respond and behave during the scenarios. In order to minimize deviations from the pre-defined script, each confederate wore a microphone to facilitate back and forth communication with one of the study investigators in the simulation control room. For the

groups where only two participants were present, one of the study investigators acted as the additional confederate according to the same script and instructions. Each scenario and debriefing was video recorded for subsequent evaluation.

Debriefings

All three team participants underwent a 20-min structured debriefing after each of their assigned scenarios. The debriefings were conducted by one of the study investigators (T.W.) who is a simulation instructor in both the Department of Anesthesia and the Department of Critical Care at Dalhousie University and has extensive knowledge and experience in simulation teaching and debriefing. The debriefings consisted of a discussion around both the technical and non-technical aspects of their performance and were based on the principles of crisis resource management (CRM). The debriefings were conducted in a manner that blinded the participants from the focus of NTS assessment. The conversations were video recorded to evaluate the consistency and quality of the debriefings.

Post-test phase

On the same day following completion of the pre-test phase, all three participants (regardless of their HFS or LFS grouping) individually underwent an identical simulator scenario on a high-fidelity simulator. The scenario was video recorded for subsequent analysis. One of the study investigators (Y.G.) and a research volunteer (D.H.) were present to act as confederates for all the post-test scenarios.

There was no formal debriefing after the post-test scenario; however, any critical medical errors would be addressed and time was given for the participants to ask questions related to the scenario.

Simulators

The study was conducted at the simulation centre located at the Victoria General (VG) Hospital in Halifax, Nova Scotia. The high-fidelity mannequin used in our study was the SimMan® 3G (Laerdal Medical Canada Ltd, Toronto, ON, Canada) with the accompanying Laerdal software (Laerdal Medical Canada Ltd, Toronto, ON, Canada), and the low-fidelity mannequin was the Deluxe Difficult Airway Trainer (Laerdal Medical Canada Ltd, Toronto, ON, Canada), a plastic upper torso with no complex functionality used for airway teaching. Pillows and blankets were used for the lower torso. The same Laerdal software was used to control and display the patient's vital signs on a video monitor. Table 1 further describes the similarities and differences between our low- and high-fidelity setup and configuration.

Simulator scenarios

The scenarios were selected from our institution's preexisting intensive care unit (ICU) simulation program and chosen because they represent emergency situations that residents may encounter in the ICU, postanesthesia care unit, or emergency room. The pre-test scenarios included anaphylaxis, pulseless electrical activity (PEA) secondary to septic shock, and acute onset atrial fibrillation secondary to a pulmonary embolus. The pre-test scenarios

Table 1 Simulator Equipment and Room Setup

	Low-fidelity	High-Fidelity
Physical environment	Identical	Identical
Confederates	2 participants	2 participants
Airway equipment	Identical	Identical
Resuscitation equipment/drugs	Identical	Identical
Mannequin	Plastic airway mannequin	SimMan 3G
Airway	Tongue can be inflated	Airway can be manipulated to varying degrees of difficulty for bag mask ventilation and direct laryngoscopy
Physiological functions	None	Auscultate normal and abnormal heart and lung sounds Can palpate arterial pulses Can speak through a built-in microphone Can perform and monitor chest compressions Able to perform cardioversion and defibrillation
Display monitor for vital signs	Identical	Identical
Voice of mannequin	Room speaker	From mannequin
Physical findings	Provided by confederate when prompted	From mannequin

had a fixed order, thus for the two-person groups, these individuals participated only in the anaphylaxis and PEA scenarios. The post-test scenario was management of cardiogenic shock secondary to an acute coronary syndrome. The scenarios occurred in a pre-defined sequence regardless of the participant's action or inaction and were standardized to therapeutic interventions as much as possible. All scenarios began with a clinical stem provided by a nurse confederate, followed by two minutes for patient assessment. In the LFS group, the patient "spoke" via speakers in the room; whereas in the HFS group, the patient "spoke" through speakers in the mannequin itself. After the initial patient assessment, a six-minute Advanced Cardiovascular Life Support event occurred (e.g., PEA, ventricular tachycardia, or unstable narrow complex tachycardia). Once return of spontaneous circulation was achieved, two minutes were provided for patient resolution and disposition. Each scenario lasted ten minutes.

Evaluation and assessment tools

The Ottawa Global Rating Scale (OGRS) (Appendix 1), a tool developed at the University of Ottawa, Canada, to assess NTS, has shown construct validity.^{16,17} The OGRS consists of five domains of CRM skills— i.e., situational awareness, leadership, resource utilization, problem solving, and communication (each scored on a seven-point scale).

For our primary outcome, we used the *overall* OGRS performance score (seven-point scale), which is guided by the scores of the individual domains described above. For our secondary outcome, the *total* OGRS score is the summation of the individual scores in the five domains described above (scored from 5-35). We adjusted for pre-test *overall* and *total* OGRS scores to account for differences in baseline performance of the participants. Three raters from the University of Ottawa (P.R., T.V., D.B.) and unknown to the study participants reviewed the video recordings of performance and applied the OGRS. Each rater was a physician trained by one of the principle investigators (D.B.) in the use of the OGRS and with expertise in the field of medical simulation. Raters familiarized themselves with the OGRS literature and practised rating on video recordings not from this study. Any large variations in scores were discussed and calibrated prior to assessing the video recordings from this study.

The Observational Structured Assessment of Debriefing (OSAD) (Appendix 2) is a tool developed at Imperial College London, United Kingdom to evaluate the consistency and quality of healthcare debriefings.¹⁸ There are eight categories in the OSAD scoring system: approach,

establishing learning environment, learner engagement, reaction, descriptive reflection, analysis, diagnosis, and application. Each category is scored from 1 = done very poorly to 5 = done very well to give a total score from 8-40. To assess for consistency in the quality of debriefing between the HFS and LFS groups, half of the debriefing sessions (18 videos) were chosen by a random number generator (Randomness and Integrity Services Ltd., Dublin, Ireland). Three raters (D.B., T.V., P.R.) scored each debriefing using the OSAD tool.

Statistical analysis

Sample size was calculated based on the mean (standard deviation [SD]) overall OGRS performance scores from a study by Kim *et al.*, where they found that PGY-3 and PGY-1 residents obtained a mean (SD) overall OGRS score of 5.5 (0.9) and 4.0 (0.9), respectively.¹⁵ The difference in OGRS scores between the PGY-3 and PGY-1 residents was 1.5, thus we considered a difference of greater than 1 in overall OGRS score between the LFS and HFS groups as educationally significant. Therefore, the sample size calculation to show non-inferiority between the LFS and HFS groups for teaching NTS was based on a non-inferiority margin of greater than 1. For a power of 0.9 and a type 1 error probability of 0.05, we calculated a total sample size of 36 participants, with 18 participants in each group, using G*Power analysis (Erdfeelder, Faul, & Buchner, 1996).

Data are presented as mean (SD) and/or 95% confidence interval (CI) where appropriate. The Shapiro-Wilk test was used to test for normality, and homogeneity of variance was assessed with Levene's test. Inter-rater reliability between the three OGRS evaluators was assessed using the intraclass correlation coefficient (ICC), and a value greater than 0.6 indicated moderate agreement between the raters.¹⁹ For our primary outcome, a one-way analysis of covariance (ANCOVA) was conducted to examine the difference between the HFS and LFS group on post-test overall OGRS scores, while controlling for pre-test overall OGRS scores *a priori*. The ANCOVA analysis produced estimated marginal means to provide means and SD for the two groups at post-test, adjusted for the *a priori* covariate of the pre-test score as well as the mean difference and its respective 95% CI. For our secondary outcomes, a one-way ANCOVA was performed to examine the difference between the HFS and LFS groups on post-test total OGRS scores, while controlling for pre-test total OGRS scores. A paired Student's *t* test was used to compare the combined pre-test scores of the HFS and LFS groups with the combined post-test scores for both the overall and total OGRS scores. A Chi square test was used to look for any interaction between overall post-test OGRS scores and both

PGY and specialty. The OSAD scores between the two groups were compared using the independent Student’s *t* test. A *P* < 0.05 was considered statistically significant. All data were analyzed using SPSS® version 21 (IBM Corp., Armonk, NY, USA).

Results

We enrolled all 36 residents who were invited to participate in the study. Table 2 summarizes the demographic information of the participants. The LFS and HFS groups had 17 and 19 participants, respectively. The difference between group numbers was due to difficulties in scheduling.

Overall and total OGRS scores

The mean (95% CI) measure of inter-rater reliability was good for the overall OGRS score (ICC, 0.72; 95% CI, 0.51 to 0.85) and total OGRS score (ICC, 0.69; 95% CI, 0.46 to 0.82); thus, we elected to use the means of the three raters for further analysis. Figs. 2 and 3 show the change from pre-test to post-test in the overall and total OGRS scores, respectively.

For our primary outcome, there was no significant difference in the mean (SD) overall post-test OGRS scores between the HFS and LFS groups after controlling for overall pre-test OGRS scores [3.8 (0.9) vs 4.0 (0.9),

respectively; mean difference, 0.2; 95% CI, -0.4 to 0.8; *P* = 0.48].

For our secondary outcomes, there was no significant difference in the mean (SD) total post-test OGRS scores between the HFS and LFS groups after controlling for total pre-test OGRS scores [19.8 (3.6) vs 21.0 (3.6), respectively; mean difference, 1.2; 95% CI, -1.2 to 3.6; *P* = 0.33]. We found no significant association between postgraduate training year and overall post-test OGRS score [χ^2 (4) = 1.55; *P* = 0.817]. We also found no significant association between postgraduate program and overall post-test OGRS score [χ^2 (4) = 2.37; *P* = 0.67].

Comparison between pre-test and post-test OGRS scores

The mean (SD) overall OGRS scores for both groups during pre-test and post-test were 3.3 (0.7) and 3.9 (0.9), respectively (mean difference 0.6; 95% CI, 0.2 to 1.0; *P* = 0.01). The mean (SD) total OGRS scores for both groups during pre-test and post-test were 17.8 (3.5) and 20.4 (3.5), respectively (mean difference, 2.6; 95% CI, 1.0 to 4.3; *P* = 0.003).

Debriefing consistency

The mean (SD) OSAD scores for the debriefings were 21.3 (2.4) in the HFS group and 22.5 (2.4) in the LFS group (mean difference, -1.1; 95% CI, -3.6 to 1.3; *P* = 0.34).

Table 2 Participant demographics

	HFS	LFS
Participants (<i>n</i>)	19	17
PGY (<i>n</i>)		
1	4	3
2	5	6
3	6	5
4	3	1
5	1	2
Male	16	12
Female	3	5
2-person group	2	1
Internal medicine (<i>n</i>)	4	3
Anesthesiology (<i>n</i>)	11	11
Emergency Medicine (<i>n</i>)	1	1
Surgical specialty (<i>n</i>)	3	2
Approximate hours of simulation exposure during residency (mean)	31	28

HFS = high-fidelity simulator; LFS = low-fidelity simulator; PGY = postgraduate year

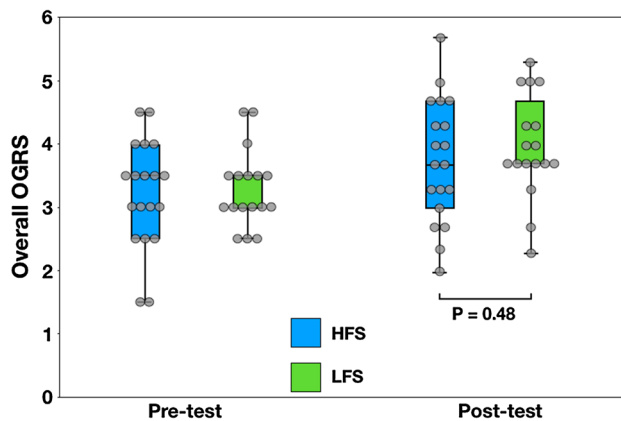


Fig. 2 Box plot with superimposed data points of the mean overall Ottawa Global Rating Scale (OGRS) scores in the low-fidelity simulator (LFS) and high-fidelity simulator (HFS) groups during the pre-test and post-test phase. There was no significant difference in the overall mean post-test OGRS scores between the HFS and LFS groups after controlling for overall pre-test OGRS scores. Box plot line represents the median; the box is the interquartiles and whiskers the range

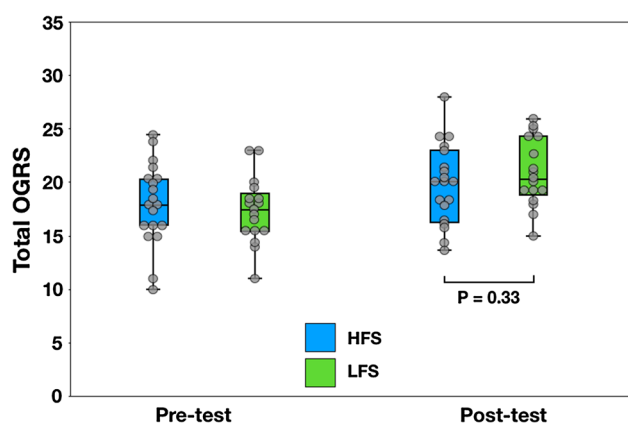


Fig. 3 Box plot with superimposed data points of the mean total Ottawa Global Rating Scale (OGRS) scores in the low-fidelity simulator (LFS) and high-fidelity simulator (HFS) groups during the pre-test and post-test phase. There was no significant difference in the total mean post-test OGRS scores between the HFS and LFS groups after controlling for total pre-test OGRS scores. Box plot line represents the median; the box is the interquartiles and whiskers the range

Discussion

The results of our study suggest that a low-fidelity simulator is not inferior to a high-fidelity simulator for teaching NTS. This supports the notion that the realism, or fidelity, of the mannequin does not strongly influence learning NTS during simulated crisis situations.

Our results are consistent with the findings in a previous study by Cheng *et al.* who compared the use of low- vs high-fidelity simulators for assessing knowledge and team leader behavioural performance during pediatric resuscitation. Their study results showed no difference in NTS between the two groups.¹⁵ The authors used a high-fidelity pre-programmed infant simulator capable of recording vital signs, audio feedback, breath sounds, chest rise, heart sounds, and palpable pulses. Their low-fidelity simulator was identical to the high-fidelity simulator but with all the innate functions turned off. In contrast to our study, they utilized only one pediatric advanced life support scenario, thus making their findings difficult to generalize to other types of complex acute care scenarios, such as the ones used in our study. Their methodology may also limit generalizability, as the difference in fidelity between groups involved a more complex design related to scripted and non-scripted debriefing. Finally, their low-fidelity model was not in fact lower in cost.

A study by Finan *et al.* also compared HFS with LFS for neonatal resuscitation training, and the authors found no significant difference in NTS between the two groups.¹⁴ Interestingly, they measured salivary cortisol levels to

assess stress levels in the participants and found no difference between the two groups. These findings suggest that HFS, despite being more “realistic”, did not elicit a commensurate emotional response from the participants. A major difference between their study and ours was that they had no pre-test on either a HFS or a LFS; instead, they simply required that the participants had completed a course in the Neonatal Resuscitation Program within the previous two months and advanced resuscitation training one month before the study period. The study also used a small convenience sample, which was probably underpowered to find a difference between groups.

We also found that OGRS scores did not improve with PGY of training. A possible explanation is that most of the participants had prior involvement with high-fidelity simulation and debriefing on NTS, and a previous study showed that NTS can improve even after one simulator session.¹³ Another reason may be due to the lack of discriminative ability of the OGRS to detect a difference in more senior residents. This is the likely reason why the study by Kim *et al.* found a significant difference between PGY1 and PGY3 residents, whereas the study by Clarke *et al.* found no significant difference between PGY2 and PGY3 residents.^{16,20} Unfortunately, our study lacked the sample size for a proper investigation of this association, but this issue certainly warrants future studies.

Despite its wide use in the simulation literature, the term “fidelity” has been poorly defined.²¹ As the field of simulation advanced, it became apparent that there were multiple dimensions to describe and define simulator fidelity.⁸ Miller was the first to separate fidelity into both a “physical” and a “psychological” domain, where physical fidelity reflects how closely the training equipment, mannequin, and environment approximate the real situation.⁸ In contrast, psychological fidelity refers to the emotional connection of the learner to the simulation scenario.²² More recently, Diekman *et al.*—then modified by Rudolph *et al.*—proposed that fidelity or realism can be separated into three areas: 1) physical, 2) conceptual, and 3) emotional.^{23,24} The difference between their distinctions and Miller’s is the addition of a conceptual domain that deals with theory, meaning, concepts, and relationships. For example, if there is hemorrhagic bleeding, then there will be hypotension and tachycardia. A common assumption in simulation is that complex skills, such as CRM, require increased physical fidelity. Nevertheless, our results suggest that conceptual and emotional fidelity are perhaps more important for teaching and learning NTS.

The significant cost associated with high-fidelity simulators renders them financially prohibitive to many academic centres, especially those in low- and middle-income countries.⁶ This study shows that effective NTS training can be successfully performed with low physical

fidelity simulators that are often a fraction of the cost to purchase and maintain. A recent study showed that a low-cost and low physical fidelity mannequin was effective in teaching NTS to Rwandan anesthesia providers.²⁵ Our study will hopefully provide useful information to help guide the development of future simulation scenarios intended to teach NTS in centres with limited financial resources.

There were several limitations to our study. First, during the pre-test scenarios, the participants randomized to the HFS group were assessed on the identical HFS setup during the post-test phase. Ideally, the pre-test scenarios should be conducted on another HFS mannequin, in order to minimize pre-familiarization of the HFS simulator. Nevertheless, although the expectations would have been that the HFS group would have higher OGRS scores than the LFS group, our results did not find any added learning benefits to being trained on a HFS during the pre-test phase. Second, there were two groups in the HFS and one group in the LFS with a two-person team during the pre-test phase. Therefore, these groups were exposed to one less scenario and debriefing, which may have introduced some bias. Another limitation to this study was having one study investigator (T.W.), who was not blinded to the objective of this study, conduct all of the debriefings. But based on the OSAD scores, the consistency and quality of the debriefings were the same in both the HFS and LFS groups. During the post-test scenario, the study investigator and research volunteer who acted as team participants were not blinded to the purpose of the study. This may influence how they behaved during the scenarios, which could lead to observer bias. For our study, we assessed immediate post-simulation training OGRS scores and did not conduct a retention test. Therefore, it is unclear how simulator fidelity affects long-term learning of NTS. Finally, most of the participants in this study were recruited from the anesthesia

department, which may reduce the generalizability of our findings.

In conclusion, our study suggests that a low physical fidelity simulator is not inferior to a high physical fidelity simulator for teaching NTS, i.e., based on the overall OGRS score in these two contexts. Our data do not support the assumption that higher fidelity (and higher cost) models result in improved learning of NTS in critical care. Adoption of low-fidelity low-cost models may have the potential to improve both the value of and access to simulation-based medical education.

Acknowledgements The authors sincerely thank the residents from Dalhousie University for their participation in our study, David Hung for his help setting up the audio visual equipment and editing the videos in our study, and the personnel in the VG simulation centre for allowing us to use their facility and equipment.

Conflicts of interest None declared.

Editorial responsibility This submission was handled by Dr. Hilary P. Grocott, Editor-in-Chief, *Canadian Journal of Anesthesia*.

Author contributions *Yuqi Gu* and *Tom Kuca* were involved in the development of the protocol for the study and research ethics submission. *Yuqi Gu* and *M. Dylan Bould* were involved in research grant preparation. *Yuqi Gu* and *Tobias Witter* contributed to data collection. *Yuqi Gu*, *Tobias Witter*, *Purnima Rao*, *Terry Varshney*, and *M. Dylan Bould* were involved in the data analysis and interpretation. *Yuqi Gu* contributed to writing the article. *Tobias Witter*, *Patty Livingston*, *Purnima Rao*, *Terry Varshney*, and *Tom Kuca* contributed their ideas for the study. *Tobias Witter* was the primary debriefer for the study. *Patty Livingston* and *M. Dylan Bould* conceived and designed the study. *Purnima Rao*, *Terry Varshney*, and *M. Dylan Bould* contributed to the video rating. *M. Dylan Bould* was involved in rater training. All authors were involved in critically revising the article.

Funding This work was supported by a research grant from the Capital Health Research Fund, Halifax, Canada, and the Department of Anesthesia, Pain Management, and Perioperative Medicine internal PROJECT fund, Halifax, Canada.

Appendix 1: Ottawa Crisis Resource Management (CRM) Global Rating Scale (“OGRS”)

	1	2	3	4	5	6	7
Leadership Skills							
Stays calm and in control during crisis							
Prompt and firm decision-making							
Maintains global perspective (“Big picture”)							
Situational Awareness							
Avoids fixation error							
Re-assesses and re-evaluates situation constantly							
Anticipates likely events							
Communication Skills							
Communicates clearly and concisely							
Uses directed verbal/non-verbal communication							
Listens to team input							
Overall Performance							
Novice; all CM skills require significant improvement			Advanced novice; many CM skills require moderate improvement		Competent; most CM skills require minor improvement		Clearly superior; few, if any, CM skills that require only minor improvement
Leadership Skills							
Loses calm and control for most of crisis; unable to make firm decisions; cannot maintain global perspective			Loses calm/control frequently during crisis; delays making decisions (or with cueing); rarely maintains global perspective		Stays calm and in control for most of crisis; makes firm decisions with little delay; usually maintains global perspective		Remains calm and in control for entire crisis; makes prompt and firm decisions without delay; maintains global perspective

	1	2	3	4	5	6	7
Problem Solving Skills							
Cannot implement ABC assessment without direct cues; uses sequential management despite cues; fails to consider any alternative in crisis			Incompetent or slow ABC assessment; mostly uses sequential management approach unless cued; gives little consideration to alternatives		Satisfactory ABC assessment; without cues; mostly uses concurrent management approach with only minimal cueing; considers some alternatives in crisis		Thorough yet quick ABC without cues; always uses concurrent management approach; considers most likely alternatives in crisis
Situational Awareness Skills							
Becomes fixated easily despite repeated cues; fails to re-assess and re-evaluate situation despite repeated cues; fails to anticipate likely events			Avoids fixation error only with cueing; rarely re-assesses and re-evaluates situation without cues; rarely anticipates likely events		Usually avoids fixation error with minimal cueing; re-assesses and re-evaluates situation frequently with minimal cues; usually anticipates likely events		Avoids any fixation error without cues; constantly re-assesses and re-evaluates situation without cues; constantly anticipates likely events
Resource Utilization Skills							
Unable to use resources & staff effectively; does not prioritize tasks or ask for help when required despite cues			Able to use resources with minimal effectiveness; only prioritizes tasks or asks for help when required without cues		Able to use resources with moderate effectiveness; able to prioritize tasks and/or ask for help with minimal cues		Clearly able to utilize resources to maximal effectiveness; sets clear task priority and asks for help early with no cues
Communication Skills							
Does not communicate with staff; does not acknowledge staff communication, never uses directed verbal/non-verbal communication			Communicates occasionally with staff, but unclear and vague; occasionally listens to but rarely interacts with staff; rarely uses directed verbal/non-verbal communication		Communicates with staff clearly and concisely most of time; listens to staff feedback; usually uses directed verbal/non-verbal communication		Communicates clearly and concisely at all times; encourages input and listens to staff feedback; consistently uses directed verbal/non-verbal communication

CM = crisis management

Appendix 2: The Observational Structured Assessment of Debriefing tool (OSAD)

	1	2 3	4 5
Objective Structured Assessment of De-briefing			
1. Approach	Confrontational, judgmental approach	Attempts to establish rapport with the learner(s) but is either overcritical or too informal in their approach	Establishes and maintains rapport throughout; uses a non-threatening but honest approach, creating a psychologically safe environment
2. Establishes learning environment	Unclear expectations of the learner(s); no rules for learner(s) engagement	Explains purpose of the debriefing or learning session but does not clarify learner(s) expectations	Explains purpose of debrief and clarifies expectations and objectives from the learner(s) at the start
3. Engagement of Learners	Purely didactic; facilitator doing all of the talking, and not involving passive learner(s)	Learner(s) participates in the discussion but mostly through closed questions; facilitator not actively inviting contributions from more passive learner(s)	Encourages participation of learner(s) through use of open-ended questions; invites learner(s) to actively contribute to discussion
4. Reaction	No acknowledgment of learner(s)'s reactions, or emotional impact of the experience	Asks the learner(s) about their feelings but does not fully explore their reaction to the event	Fully explores learner(s)'s reaction to the event, dealing appropriately with learner(s) who are unhappy
5. Descriptive Reflection	No opportunity for self-reflection; learner(s) not asked to describe what actually happened in the scenario	Some description of events by facilitator but with little self-reflection by learner(s)	Encourages learner(s) to self-reflect on what happened using a step-by-step approach
6. Analysis	Reasons and consequences of actions are not explored with the learner(s)	Some exploration of reasons and consequences of actions by facilitator, but not learner(s), but no opportunity to relate to previous experience	Helps learner(s) to explore reasons and consequences of actions, identifying specific examples and relating to previous experience
7. Diagnosis	No feedback on clinical or teamwork skills; does not identify performance gaps or provide positive reinforcement	Feedback provided only on clinical (technical) skills; focuses on errors and not purely on behaviours that can be changed	Provides objective feedback on clinical (technical) and teamwork skills; identifies positive behaviours in addition to performance gaps, specifically targeting behaviours that can be changed
8. Application	No opportunity for learner(s) to identify strategies for future improvement or to consolidate key learning points	Some discussion of learning points and strategies for improvement but lack of application of this knowledge to future clinical practice	Reinforces key learning points identified by learner(s) and highlights how strategies for improvement could be applied to future clinical practice

Reproduced with permission from: *Arora S, Ahmed M, Paige J, et al. Objective structured assessment of debriefing: bringing science to the art of debriefing in surgery. Ann Surg 2012; 256: 982-8*¹⁸

References

1. *Leblanc VR.* Review article: Simulation in anesthesia: state of the science and looking forward. *Can J Anesth* 2012; 59: 193-202.
2. *Ziv A, Wolpe PR, Small SD, Glick S.* Simulation-based medical education: an ethical imperative. *Acad Med* 2003; 78: 783-8.
3. *Cook DA, Brydges R, Hamstra SJ, et al.* Comparative effectiveness of technology-enhanced simulation versus other instructional methods: a systematic review and meta-analysis. *Simul Healthc* 2012; 7: 308-20.
4. *Cook DA, Hatala R, Brydges R, et al.* Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA* 2011; 306: 978-88.
5. *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 Anesth* 2010; 57: 134-42.
6. *Kurrek MM, Devitt JH.* The cost for construction and operation of a simulation centre. *Can J Anesth* 1997; 44: 1191-5.
7. *Zendejas B, Wang AT, Brydges R, Hamstra SJ, Cook DA.* Cost: the missing outcome in simulation-based medical education research: a systematic review. *Surgery* 2013; 153: 160-76.
8. *Maran NJ, Glavin RJ.* Low- to high-fidelity simulation - a continuum of medical education? *Med Educ* 2003; 37(Suppl 1): 22-8.
9. *Norman G, Dore K, Grierson L.* The minimal relationship between simulation fidelity and transfer of learning. *Med Educ* 2012; 46: 636-47.

10. *Flin RH, Crichton M, O'Connor P.* Safety at the sharp end: a guide to non-technical skills. Farnham: Ashgate Publishing Ltd.; 2013 .
11. *Riem N, Boet S, Bould MD, Tavares W, Naik VN.* Do technical skills correlate with non-technical skills in crisis resource management: a simulation study. *Br J Anaesth* 2012; 109: 723-8.
12. *Fletcher GC, McGeorge P, Flin RH, Glavin RJ, Maran NJ.* The role of non-technical skills in anaesthesia: a review of current literature. *Br J Anaesth* 2002; 88: 418-29.
13. *Yee B, Naik VN, Joo HS, et al.* Nontechnical skills in anesthesia crisis management with repeated exposure to simulation-based education. *Anesthesiology* 2005; 103: 241-8.
14. *Finan E, Bismilla Z, Whyte HE, Leblanc V, McNamara PJ.* High-fidelity simulator technology may not be superior to traditional low-fidelity equipment for neonatal resuscitation training. *J Perinatol* 2012; 32: 287-92.
15. *Cheng A, Hunt EA, Donoghue A, et al.* Examining pediatric resuscitation education using simulation and scripted debriefing: a multicenter randomized trial. *JAMA Pediatr* 2013; 167: 528-36.
16. *Kim J, Neilipovitz D, Cardinal P, Chiu M.* A comparison of global rating scale and checklist scores in the validation of an evaluation tool to assess performance in the resuscitation of critically ill patients during simulated emergencies (abbreviated as "CRM simulator study IB"). *Simul Healthc* 2009; 4: 6-16.
17. *Kim J, Neilipovitz D, Cardinal P, Chiu M, Clinch J.* A pilot study using high-fidelity simulation to formally evaluate performance in the resuscitation of critically ill patients: The University of Ottawa Critical Care Medicine, High-Fidelity Simulation, and Crisis Resource Management I Study. *Crit Care Med* 2006; 34: 2167-74.
18. *Arora S, Ahmed M, Paige J, et al.* Objective structured assessment of debriefing: bringing science to the art of debriefing in surgery. *Ann Surg* 2012; 256: 982-8.
19. *Landis JR, Koch GG.* The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.
20. *Clarke S, Horeczko T, Carlisle M, et al.* Emergency medicine resident crisis resource management ability: a simulation-based longitudinal study. *Med Educ Online* 2014; 19: 25771.
21. *Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA.* Reconsidering fidelity in simulation-based training. *Acad Med* 2014; 89: 387-92.
22. *Farmer E, Van Rooij J, Riemersma J, Moraal J, Jorna P.* Handbook of simulator-based training. Farnham: Ashgate Publishing; 1999 .
23. *Rudolph JW, Simon R, Raemer DB.* Which reality matters? Questions on the path to high engagement in healthcare simulation. *Simul Healthc* 2007; 2: 161-3.
24. *Dieckmann P, Gaba D, Rall M.* Deepening the theoretical foundations of patient simulation as social practice. *Simul Healthc* 2007; 2: 183-93.
25. *Skelton T, Nshimyumuremyi I, Mukwesi C, Whynot S, Zolpys L, Livingston P.* Low-cost simulation to teach anesthetists' non-technical skills in Rwanda. *Anesth Analg* 2016; 123: 474-80.