The minimal relationship between simulation fidelity and transfer of learning

Geoff Norman,[1] Kelly Dore[2] & Lawrence Grierson[3]

CONTEXT High-fidelity simulators have enjoyed increasing popularity despite costs that may approach six figures. This is justified on the basis that simulators have been shown to result in large learning gains that may transfer to actual patient care situations. However, most commonly, learning from a simulator is compared with learning in a 'no-intervention' control group. This fails to clarify the relationship between simulator fidelity and learning, and whether comparable gains might be achieved at substantially lower cost.

OBJECTIVES This analysis was conducted to review studies that compare learning from high-fidelity simulation (HFS) with learning from low-fidelity simulation (LFS) based on measures of clinical performance.

METHODS Using a variety of search strategies, a total of 24 studies contrasting HFS and LFS and including some measure of performance were located. These studies referred to learning in three areas: auscultation skills; surgical techniques, and complex management skills such as cardiac resuscitation.

RESULTS Both HFS and LFS learning resulted in consistent improvements in performance in comparisons with no-intervention control groups. However, nearly all the studies showed no significant advantage of HFS over LFS, with average differences ranging from 1% to 2%.

DISCUSSION The factors influencing learning, and the reasons for this surprising finding, are discussed.

*Medical Education 2012: **46:** 636–647*
doi:10.1111/j.1365-2923.2012.04243.x

Discuss ideas arising from this article at www.mededuc.com 'discuss'

[1]Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada
[2]Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada
[3]Department of Family Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada

Correspondence: Geoff Norman, Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada. Tel: 00 1 905 525 9140 (ext. 22119); Fax: 00 1 905 572 7099; E-mail: norman@mcmaster.ca

INTRODUCTION

There is growing interest in the use of realistic computer-controlled 'high-fidelity' simulation (HFS) in medical education, both for learning and assessment. Simulations have a number of potential advantages over traditional ward- or clinic-based learning. As Teteris *et al.*[1] state:

> 'Simulated learning experiences not only allow learners to practise and err, without patients suffering adverse clinical consequences, they also offer more control over the learning experience… The arguments in favour of simulation are so persuasive that some have proposed that the absence of clinical training is unethical.'

Central to the use of simulations are several assumptions.

1 Instruction on simulators results in meaningful learning.

This issue was recently elegantly addressed in a large systematic review conducted by Cook *et al.*,[2] comprising a total of 609 studies with 35 000 trainees. The results of this review are unequivocal. Learning on simulators showed large gains in performance; typically, effect sizes of around 1.0 are observed for knowledge, time management skills, process skills and product skills.

However, only studies comparing simulator training with no training were examined. The authors[2] deliberately excluded studies in which training on a simulator was compared with training on another simulator or an alternative instructional method.

2 Skills acquired on the simulation can be applied to real patients (i.e. they can be 'transferred').

A number of studies, reviewed by Teteris *et al.*,[1] demonstrate that students who train on simulators show better performance in real situations in a variety of domains, including laparoscopy,[3] catheter insertion,[4] advanced cardiac life support (ACLS)[5] and auscultation.[6,7] Some of these gains have been shown in highly authentic transfer tasks, such as actual response to ACLS events.[5]

However, many of these studies compare the outcomes of formal simulator sessions with those of no formal training. Thus, although transfer is an important outcome of simulator training, there is little evidence that the gains observed are uniquely attributable to the HFS experience and not simply to the time spent in learning.

3 The closer to the 'real world' (and, commensurately, the more expensive the simulation), the better the transfer to real life.

This assumption appears as almost a self-evident truth. As Schuwirth and van der Vleuten state[8]:

> 'Authenticity should have high priority when programmes for… assessment… are designed… The situation in which a candidate's competence will be assessed should resemble the situation in which the competence will actually have to be used.'

The rationale for the assumption, and indeed for many educational practices, such as authentic assessment, situated cognition and cognitive apprenticeship,[9] is the finding that the closer the context of learning to the context of practice, the better the learning.[10] However, this is not a uniform finding within psychology[11] and a number of studies have failed to show an advantage for matching context of learning and test. Even if we accept the argument that context similarity is desirable, a more difficult question refers to precisely what aspects of the context should be the focus of attention. This is explored in the next section.

4 Authenticity – the resemblance of the simulation to an equivalent real-life scenario – is the critical determinant of transfer.

Proximity to 'real life' may not be unidimensional. As Maran and Glavin[12] have suggested, fidelity can be assessed on two levels: 'engineering fidelity' or authenticity (does the simulation look realistic?), and 'psychological fidelity' (does the simulator contain the critical elements to accurately simulate the specific behaviours required to complete the task?).[13] These may be independent. As one example, early laparoscopic simulators were very realistic in their ability to simulate the visual field of the laparoscope and its change with movement. However, to learn to suture laparoscopically, it is critical to 'feel' the needle against the tissue using what is referred to as 'haptic' feedback. Thus, these simulators had good engineering fidelity but poor psychological fidelity. The latter might be better accomplished by using a beef tongue inside a box, although this would clearly lack engineering fidelity.

5 More complex skills demand more complex simulators.

Even if we recognise that basic skills may be taught and learned on relatively simple static simulators without concern for authenticity, it appears a universal assumption that complex skills require complex simulations:

> 'Education in basic procedural skills like suturing… can be delivered using simple task trainers, devices that mimic body parts or regions… Complex clinical events… require training on much more sophisticated medical simulators.'[14]

Although it seems reasonable that more advanced learners performing more complex tasks may require more sophisticated simulators, there is little evidence to support the assertion.

Complex simulators do not come without cost. The METI cardiovascular simulator (Medical Education Technologies, Inc., Sarasota, FL, USA) costs US$80 000; the Harvey® heart sound simulator (Laerdal Medical Corp., Wappinger Falls, NY, USA) is similarly expensive. The cost of HFS has educational consequences. Few schools will have the resources to invest in a Harvey® simulator. Yet even if such a simulator is available and encouraged, student access to the simulator may be restricted simply by the number of learners. By contrast, if each student is given a CD at minimal cost, there is no economic restriction on the amount of practice time a student may accrue. Differentiating between the advantages afforded by each method may indicate a trade-off between substantially greater practice with lower fidelity and much less practice at higher fidelity. The extent to which additional fidelity, at additional cost, leads to more effective learning is a critical factor in this equation. It is certainly not a universal finding; aviation psychology has shown a law of diminishing returns in investment in simulators:

> 'One [misconception is] that greater financial investment in a simulator facilitates training on that simulator […] Relatively low level simulations that capture the salient characteristics of the task are far more versatile for measuring and are very effective for improving performance.'[15]

The same may be true in medical simulation.

In this review, we will investigate the extent to which greater fidelity leads to greater transfer by examining studies that contrast learning based on low-fidelity simulation (LFS) with that based on HFS. To do this, we will essentially adopt the criteria of the respective authors. Although fidelity refers most commonly to the domain of 'engineering fidelity' (as defined earlier), exceptions arise. Further, the HFS may not necessarily be a computer-controlled manikin; in the examination of surgical techniques it is more common to compare the use of live or dead tissue with that of rubber or plastic. What does remain generally true is that HFS is more realistic (authentic) and more expensive (often *much* more expensive) than LFS.

Theoretical framework

This review is not intended as a systematic review article. Such reviews have already been published, and one,[16] in particular, is a primary source for the literature we will review. However, our goal differs from that of earlier reviewers; we do not aim to establish whether or not HFS 'works'. That has been well established. Instead, our review is directed at examining two issues.

1 The relationship between performance on HFS and performance on a structured control intervention, often an LFS but occasionally an instructional video. It is of limited value to prove that an educational intervention is superior to nothing or to 'usual care', which often amounts to the same thing. As our specific interest is the relationship between fidelity and performance, we will primarily examine evidence that contrasts two types of structured intervention.
2 The relationship between learning on a simulator and performance outcomes. Ideally, this implies an outcome measured on real patients, on standardised patients (SPs) or on some credible measure that clearly involves some degree of transfer from the simulation. Assessing performance on an HFS using an outcome assessed with the same HFS says nothing about transfer because specific experience with the simulator is likely to bias any outcome assessed with the simulator. Such results are potentially interesting, however, if they can show that an LFS results in performance as good as those observed on the HFS because the experiment is biased against this outcome.

By implication, then, we are not interested in studies that use self-reported confidence, improvement in skill or any other self-report measure. The literature on self-assessment is uniform in its condemnation of conclusions drawn from such data.[17] Further, we do not intend to examine such issues as the use of feedback or decay of skills; such issues are peripheral to our central question and have been addressed elsewhere.[16]

METHODS

The present study is not intended as a systematic review. As we indicated, a systematic review of HFS was published in 2005[16] and updated in 2010,[14] and we carefully examined these reviews to identify any articles that contrasted HFS with LFS. We have supplemented articles mentioned in these reviews with other studies from other review articles[1] and other studies identified by ourselves.

The present review includes evidence from 18 studies. Of these, seven primarily surgical interventions did not appear in the two systematic reviews, mainly because they involved static simulations.

The review will examine evidence in three broad domains: (i) auscultation skills and use of heart sound simulators; (ii) basic motor skills, and (iii) complex crisis management skills.

RESULTS

Training in auscultation skills

Perhaps the most comprehensive evidence of the role of HFS in skill acquisition refers to Harvey®, a computer-controlled manikin that has been used for about 20 years. As described in one review article,[18] Harvey®:

> '…provides a comprehensive cardiology curriculum by realistically simulating 27 cardiac conditions. The physical findings programmed in the simulator for each disease include blood pressure, bilateral jugular venous, carotid and peripheral arterial pulses, precordial impulses in six different areas… auscultatory events… that are synchronised with the pulses and vary with respiration.'

However, although Harvey® has impressive engineering fidelity and simulates 30 different heart sounds, it contains only one example of 25 of the sounds, two examples of three sounds (mitral regurgitation, aortic regurgitation, acute myocardial infarction) and three examples of mitral stenosis. Thus, the opportunity to hear a variety of examples of each condition is severely constrained.

The originators of Harvey® and others have conducted a number of studies[6, 7, 19] to demonstrate the validity of Harvey® in learning auscultation skills. However, with only one exception, outcomes using Harvey®[21] have been compared with those in a no-intervention control group that received the usual haphazard bedside teaching of heart sounds.[19]

Perhaps the most widely cited of these studies is that by Issenberg *et al.*,[20] in which students in an internal medicine rotation were randomly assigned (by rotation) to learn auscultation skills either in formal training sessions using Harvey® or by the usual ward-based instruction, which was neither standardised nor documented. At the end of the rotation, students in both groups were tested using heart sounds broadcast from the Harvey® simulator. The Harvey®-trained group showed a 33% gain in performance; the 'usual care' group showed a change of 5%.

Another study by Butter *et al.*[7] used a similar design, in which outcomes in Year 3 students who had undertaken a simulation-based course were compared with those in untrained Year 4 students. The Year 3 group showed a large gain in accuracy (94% versus 74%) when tested with the same simulator.

Although these studies show impressive gains, they have two serious limitations to their ability to address our questions. Firstly, the control intervention is unspecified and is likely to be minimal. Although it might be argued that this represents the usual approach to auscultation instruction, nevertheless it is not useful in addressing the comparison between HFS and LFS. Secondly, the outcome was assessed on Harvey® and thus there is no evidence from this study that the skills are transferable.

Some reviews have claimed that Harvey® training does enhance transfer. As Issenberg *et al.*[18] state:

> 'Harvey has been rigorously tested to establish its educational efficacy… Students who used CPS [cardiology patient simulator] performed significantly better than the non-CPS group… not only on the CPS skills post-test ($p < 0.001$) but also on the patient skills post-test ($p < 0.03$).'

A study of transfer was conducted by Ewy *et al.*[6] using final-year medical students. On real patients, the Harvey®-trained group achieved a mean score of 54.9 and the control group a mean score of 52.1. The difference of 2.8%, although statistically significant, was of little educational significance. Butter *et al.*[7] also assessed transfer to a real patient, with a significant but small gain in the simulator group compared with the no-intervention control group (82% versus 75%).

All of these studies compared outcomes of training on a simulator with those of no formal training and such comparisons are not directly relevant to our review. However, despite this strong contrast in intensity of instruction, which is mirrored in large gains when trainees are tested on the simulator, the few reports of transfer to real patients show only modest improvement. Another study reveals one reason for this lack of transfer. Fraser *et al.*[21] trained students to recognise a case of either symptomatic aortic stenosis or asthma on Harvey® and then tested them on both scenarios. Perhaps not surprisingly, the students showed large improvements in their ability to recognise and diagnose the training problem, but showed no improvement on either the alternative problem or a new problem in the same system (mitral regurgitation or pulmonary fibrosis). In short, transfer to a new problem, even when tested on the same simulator, did not occur.

One recent study by de Giovanni *et al.*[19] made a specific comparison between HFS and LFS by comparing outcomes in a group of 37 Year 2 students who received 4 hours of instruction on Harvey® with those in a second group who received the same amount of training listening to a CD that included multiple sounds for each condition. Both groups were tested 6 weeks later in an objective structured clinical examination (OSCE) format with real patients with mitral stenosis, mitral regurgitation, aortic stenosis, aortic regurgitation or normal heart sounds. The Harvey®-trained group showed a small advantage in the detection of specific sounds (62% versus 49%), but no differences emerged in diagnostic accuracy (59% versus 57%), clinical skills (76% versus 79%) or communication skills (70% versus 70%) (the latter two categories were rated by a doctor-examiner). Hence, the opportunity to practise auscultation skills and learn heart sounds on the HFS did not translate into greater improvement in clinical or detection skills.

There is also correlational evidence for HFS versus LFS. Hatala *et al.*[22] studied 28 internists in an OSCE with 12 stations. Heart sounds were portrayed by Harvey® in four stations, by an SP with heart sounds delivered by a laptop computer placed beside the patient (what Kneebone *et al.*[9] call a 'hybrid' simulation) in four stations, and by real patients in four stations. The same four conditions (aortic stenosis, mitral stenosis, mitral regurgitation, normal) were present in all modalities. Overall diagnostic accuracy was 7.7 on Harvey® stations and 8.0 on SP-based stations, but only 6.7 on real patient-based stations (the differences between outcomes on real examples and those on simulation were significant, whereas the differences between outcomes on the two simulations were not). Furthermore, an examination of correlations between modalities, corrected for reliability, showed that the disattenuated correlation between the Harvey® and SP-based stations was 0.99, but correlations between each of these and real patient-based stations were 0.51 and 0.54, respectively. Thus, the HFS and LFS stations both appear to have been easier than real patient-based stations, and both appear to have tested equivalent skills, but the relationship between performance on either simulator and performance with real patients was moderate and equivalent.

Training in basic motor skills

A number of studies have examined the relative performance gains to be derived from HFS versus LFS in the development of basic motor, usually surgical, skills. As indicated earlier, although the HFS is not computer-controlled in many of these studies[23–27] (the exception is virtual reality [VR] simulations for some skills), it does clearly differentiate from LFS in both realism and cost.

There is considerable uniformity in the approach; this is unsurprising as all the studies originated in the same centre in Toronto. All the studies used either two or three groups so that outcomes were compared in HFS versus LFS groups and occasionally in a control group of participants who read a text or watched a video, and outcomes were measured either on the HFS or on a more realistic object. Results are summarised in Table 1.

Anastakis *et al.*[23] compared outcomes of a 4-hour training session on a cadaver with outcomes achieved by simple bench models (e.g. a coconut lined with plastic and a water-filled balloon for burr hole insertion) with outcomes achieved by reading a text for six different basic skills (Burr hole insertion, chest tube insertion, bowel anastamosis, wound closure, tendon repair, metacarpal fixation). Residents were tested on a cadaver in a realistic operating room (OR) environment a week later. Both simulations achieved significantly better outcomes than did the text, but there was no difference between the simulations.

Matsumoto *et al.*[24] taught residents ureteroscopy using a didactic session, an LFS consisting of a Styrofoam coffee cup and plastic straws (based on task analysis) and a commercially available HFS costing US$3700. Testing was conducted using the

Table 1 Characteristics of studies comparing outcomes of training using high-fidelity simulation with outcomes of training using low-fidelity simulation or other intervention

Reference	Domain	Treatment: high fidelity	Control (1): low fidelity	Control (2)	Outcome	Treatment mean	Control 1 mean	Control 2 mean	Significance
Issenberg *et al.*[19]	Heart sounds	Harvey®	N/A	Nothing	Simulator diagnosis	80%	47%		Yes
Butter *et al.*[7]	Heart sounds	Harvey®	N/A	Nothing	Simulator diagnosis	94%	74%		Yes
					Patient diagnosis	82%	75%		Yes
Ewy *et al.*[6]	Heart sounds	Harvey®	N/A	Nothing	Simulator diagnosis	68.2%	58.6%		Yes
					Patient diagnosis	54.9%	52.1%		Yes
de Giovanni *et al.*[21]	Heart sounds	Harvey®	CD-ROM	N/A	Patient diagnosis	59%	57%		No
Hatala *et al.*[22]*	Heart sounds	Harvey®	SP + laptop	Real patient	Harvey®	77%	N/A	N/A	N/A
					SP + laptop	80%			
					Patient	67%			
Anastakis *et al.*[23]	Basic skills	Cadaver	Bench model	N/A	Cadaver checklist	70%†	68%†	61%†	No
Matsumoto *et al.*[24]	Ureteroscopy	Limbs and things	Coffee cup, plastic straw	Didactic	High-fidelity model checklist	95%†	90%†	77%†	No
Grober *et al.*[25]	Microsurgery anastomosis	Rat vas deferens	Surgical tubing	Didactic	Rat vas deferens	88%†	84%†	73%†	No
Sidhu *et al.*[27]	Vascular anastomosis	Human cadaver arm	Plastic model		Live pig†	69%†	81%†	N/A	Yes/No‡
Chandra *et al.*[26]	Bronchoscopy	VR simulator	Wood box	N/A	Real patient	74%†	71%†	N/A	No
Munz *et al.*[28]	Laparoscopy	VR (LapSim®)	Box trainer	None	Moves§	500	600	1500	No¶
					Path§	3500 cm	4000 cm	7000	No¶
					Economy§	0.80	0.85	1.60	No¶
Brydges *et al.*[29]	i.v. insertion	SimMan®	Static plastic	VR	SP checklist**	72%†	68%†	64%	Yes
Morgan *et al.*[30]	Critical care	HF simulator	Video demo	N/A	HF simulator	85%	85%	N/A	No
Nyssen *et al.*[31]	Critical care anaesthesia	Computer + manikin	Computer	N/A	Checklist	76%	81%	N/A	Yes
Baxter & Norman[32]	Critical care	SimMan®	Video demo	N/A	Global rating	72%	67%	52%	
Wenk *et al.*[33]	Induction	METI simulator	Problem-based discussion	N/A	METI simulator	71%	65%		No
Bruppacher *et al.*[34]	Cardiac weaning	Simulator	Interactive seminar		Real patient	89%	73%		Yes

* This was a comparative study in an examination setting, not an intervention study. See text
† Read from figure
‡ Significant difference for junior residents ($p = 0.05$), but not for global ratings or senior residents
§ Lower score is better
¶ Test of high- versus low-fidelity
** Raw scores arbitrarily divided by 25
N/A = not applicable; VR = virtual reality; SP = standardised patient; HF = high-fidelity

HFS by expert observers using a global rating and checklist. Again, the LFS and HFS were superior to the didactic training, but there was no difference in outcomes between the LFS and HFS.

Grober *et al.*[25] taught residents microsurgery of the vas deferens using either a live rat vas deferens or a surgical tubing bench model and then tested the residents using both modalities. Checklist and global scores for the live rat outcomes did not differ between the HFS and LFS groups, but both were higher than those achieved by a didactically taught group. This study also involved a 'patient' outcome: the patency of the rat vas deferens at 30 days. Again, outcomes in the LFS and HFS groups were superior to those in the didactic training group but did not differ between type of simulation.

Chandra *et al.*[26] compared a VR bronchoscopy simulator (AccuTouch®; Immersion Medical, Inc., Gaithersburg, MD, USA) with a simple 'choose the hole' model, based on a careful task analysis. Outcomes were assessed on a real patient 1 week later. No control group was used as prior research by the group had shown little training benefit from didactic instruction. The results were consistent with those of other studies; no significant difference between the LFS and HFS groups emerged.

Sidhu *et al.*[27] conducted a similar study of training in vascular anastomosis, comparing an LFS (plastic tube model) and an HFS (a human cadaver arm brachial artery). There was no didactic control group. Testing was carried out in a live pig a week later. By contrast with the previous studies, these authors[27] found a difference between outcomes in the LFS and HFS groups; however, although this difference was significant for checklist scores, junior residents and 'final product' scores, it was not significant for global ratings or senior residents.

Finally, two studies directly manipulated engineering and psychological fidelity. Munz *et al.*[28] compared two groups, one of which was trained with a box trainer using real laparoscopic instruments (costing about US$2000) and one of which was trained using a VR trainer (LapSim®; Surgical Science, Gothenburg, Sweden) costing about US$20 000 that gave realistic visual information but no haptic feedback. Training was given on three laparoscopic tasks (navigation, coordination, grasping) in three 30-minute instructional periods delivered over 3 weeks. Outcome was measured on the same trainer using objective measures of the number of movements, total distance, total time and economy, and changes in outcomes pre- and post-training were computed. Both groups showed improvement over a control group, but there was no difference between the two simulation groups in the number of movements, distance or economy of movement. However, this study did not include a transfer task, its instructional time was short (90 minutes) and its sample sizes were small (eight per group).

Brydges *et al.*[29] compared three conditions in teaching intravenous (i.v.) catheterisation: LFS (a VR simulator costing US$12 000); HFS (SimMan® [Laerdal Medical Corp.], costing US$35 000), and 'progressive' fidelity, in which students progressed from low- to high-fidelity conditions. However, it is notable that the 'low-fidelity' simulator was an expensive single-purpose VR simulator, and the 'medium-fidelity' simulator (which was not studied separately) was an inexpensive static simulator costing US$600. The ordering was explicitly based on psychological fidelity; although the VR simulator provided better visual information, it lacked a realistic arm and so movements were not accurately simulated. In contrast with the other studies, the authors[29] found a clear gradient with fidelity, with progressive fidelity showing the best performance on a global rating and checklist, and LFS showing the worst. These two studies[28,29] provide some evidence that simulations that focus on psychological fidelity can provide performance gains equal to or greater than those of HFS at costs at least an order of magnitude lower.

Thus, of the studies examined, all but one showed that LFS resulted in performance gains, assessed on either an HFS or a real patient (animal or human), that are approximately equivalent to those observed with HFS, although outcomes of LFS appear to be consistently a little lower than those of HFS. The one exception showed a difference on checklist but not global scores.

Training in critical care and crisis management skills

It is often assumed that although LFS may be adequate for the development of simple skills, as skills become more complex, the authenticity and fidelity of the simulation must increase in a manner commensurate with the skill.[14] If this is true, studies of LFS versus HFS training in complex tasks such as advanced trauma life support or other crisis management situations may well show a clear superiority of HFS.

The design of such studies must, however, result in compromises. Firstly, it is not exactly clear what an

LFS of a crisis management situation would look like. Perhaps for that reason, many of the comparative studies use a passive learning strategy, such as watching a video, as a control. Further, it would be extremely demanding to attempt to validate the training outcomes with real patients because the management of a real heart attack victim is not the best context in which to conduct educational research. One study that did look at formal ACLS training with simulators used a case–control design and measures of response to actual clinical events.[14] However, the comparison was with 'usual care', not a formal LFS instruction. As a result, all the studies of complex skills that compare outcomes of HFS with those of some type of control actually test outcomes on the HFS. This leads to the obvious bias that transfer is not being examined for the HFS group and hence we would expect a bias toward a larger effect of HFS. Surprisingly, this was generally not observed.

Morgan *et al.*[30] trained medical students in an anaesthesia rotation to manage myocardial ischaemia, anaphylaxis or hypoxaemia in a 1.5-hour session using either a full HFS with an instructor or a video demonstration by an instructor on the HFS. All students were then tested on the same scenario using the HFS. There were no significant overall differences between the two groups (mean score in the HFS group: 10.27; mean score in the video group: 10.21 [our calculation]).

Nyssen *et al.*[31] trained students on an anaesthesia rotation on two simulators, comprising, respectively, a full HFS or a computer screen-based simulation. Performance was assessed on the HFS on the same or a different scenario. Strangely, these authors[31] did not perform a statistical test of the difference between the two groups; however, on their 'same scenario' test the mean score of the computer-trained group (81.4) was significantly higher than that of the HFS group (75.8) ($t = 2.35$, $p < 0.05$ [our calculation]). On the 'different scenario' test, scores did not differ between the training groups (computer-trained group, mean score: 68.8; HFS group, mean score: 70.8).

Baxter and Norman[32] examined students' ability to respond in a critical care situation. After an initial orientation, they randomised 27 final-year nursing students to receive one of: no instruction; a video demonstration, and practice on an HFS (SimMan®). All students were then tested with a new simulation on the simulator and evaluated by a blinded assessor using a 7-point global rating scale. Both instructional groups significantly outperformed the control group (simulator group, 5.04; seminar group, 4.74; control group 3.64), but there was no difference between the instruction groups.

Wenk *et al.*[33] compared a problem-based discussion with practice on an HFS in emergency induction with 32 students. The outcome was assessed 10 days later on a simulator. Although self-assessment scores (confidence) were higher in the HFS group (21.0 versus 19.4), an objective task-specific score was not (111 versus 123), although the effect size was reported at 0.52.

Bruppacher *et al.*[34] conducted a randomised trial in which anaesthesiology residents were instructed in cardiac 'weaning' from the bypass and compared the outcomes of trainees instructed using an HFS with those of residents instructed in a 2-hour seminar. Twenty trainees were involved. Two weeks and 5 weeks after instruction, they were observed performing a weaning on an actual patient, using a global rating and a checklist. The simulation group performed significantly higher on both the post-test and the retention test (post-test global scores: 14.3 versus 11.8; retention test global scores: 14.1 versus 11.7; post-test checklist scores: 89.9 versus 75.4; retention test checklist scores: 93.2 versus 77.0).

Thus, in the learning of complex skills, four of five studies[30–33] showed no significant advantage to the complex HFS. However, by contrast with the studies of basic motor skills, four of the five studies[30, 32–34] used the passive viewing of a criterion performance on discussion as the LFS. Clearly, this involves no practice whatsoever. In addition, in four of the five studies, performance was assessed on the simulator[30–33]. Thus these studies have two potential biases in favour of the HFS group. Despite this, only one of the studies[34] showed a significant effect of HFS over LFS, although this was also the single study that clearly assessed transfer to real patients.

The results are summarised in Table 1, in which we have abstracted the scores from each study and converted them to a percentage (except those in the study by Munz *et al.*[28]). These are summarised here.

Studies of training in heart sounds ([6,7,19,21])

The average effect of training on the Harvey® simulator compared with no treatment when tested on the same simulator[6,7,19] was 21%. On real patients, the effect of simulator-based training compared with no treatment[6,7] was 4.5%; the effect of simulator-based training compared with LFS training was 2%.

Studies of training in basic surgical skills ([23–29])

High-fidelity simulator training produced a gain of 12% compared with no active treatment[23–25,29] and a gain of 1% compared with LFS training.[23–27,29]

Studies of training in critical care ([30–34])

High-fidelity simulator training produced a gain of 4.4% compared with a passive demonstration.[30–34] However, one study of weaning from bypass showed HFS training to have a positive effect of 16%. When this study was omitted, the average difference was 1.5%.

DISCUSSION

Despite a few anomalies, the results of this review appear quite consistent. HFS shows clear gains in performance and transfer to the real patient setting compared with typical opportunistic instruction. However, when the outcomes afforded by HFS are compared with those of LFS, the gains of HFS are more modest and, almost without exception, are not statistically significant.

The review may be faulted because the search was not based on explicit criteria and computer databases; it is not a 'systematic review'. Although we cannot claim to have identified all the relevant literature, the results are sufficiently consistent that we have confidence in our conclusions.

The decision to choose HFS over LFS is not simply economic; if it were, one might well decide that the additional benefit, although small, can justify the additional cost, however large. However, although we have not conducted an exhaustive economic analysis, our impression is that any simulator that involves a dynamic computer typically costs an order of magnitude (factor of 10) more than an equivalent static simulator. No institution can afford an unlimited number of HFSs, yet furnishing even one simulator for each special purpose can become astronomically expensive.

The cost issue may ultimately translate into an educational issue. Expertise comes with deliberate practice; indeed, this axiom is the mantra of the simulation movement. However, the cost of the simulator can restrict the number of simulators available, which, in turn, can seriously limit the number of hours each student can access the simulator. It may be more effective, not just in terms of cost, to provide each student with unlimited access to an LFS rather than an hour or two on an HFS.

Why is it that increases in fidelity are not matched by equivalent gains in performance? There are a number of possible explanations.

The role of context

Clearly, HFS expends substantial resources on being as realistic as possible. Harvey® has a realistic face and extremities, lies on a real hospital gurney and has a chest that moves up and down to simulate breath sounds. SimMan® has built-in speakers so that the operator can communicate like a real patient. All of these strategies are aimed at making the context as realistic as possible. But is context that important? As we indicated in the introduction, studies of context effects in psychology are equivocal. It may be that we are more capable of moving from one context to another than has been thought.

Complexity and cognitive load

One assumption underlying the attraction of HFS is that practice in a realistic environment will lead to better transfer as a result of context similarity. This may not be the case for at least two reasons. Firstly, the relationship may not be linear. Using the reference standard of simulators, the cockpit simulator, a trainee must have considerable prerequisite skill to profit from the experience of a simulated Boeing 747 cockpit. If someone cannot fly a 747, placing him in a simulated situation in which he must try to land one with two engines out is unlikely to accomplish much. A related issue is cognitive complexity. Cognitive load theory[35] demonstrates that many additions to the learning task may detract from learning because of our limited ability to process incoming information.

There is evidence in the motor skills literature[36] that novices may well be better off with simpler models and should gradually move to more complex models as their skills improve, a strategy known as 'progressive fidelity'. This was the thesis of the study by Brydges *et al.*[29] described earlier. Dubrowski *et al.*[36] showed that a group learning laparoscopic knot-tying under conditions of progressive fidelity did as well as a group of learners who spent equivalent time on the full complexity model. Notably, however, the progressive fidelity group did no better and therefore this study does not, of itself, provide evidence of the advantage of progressive fidelity. By contrast, the study by Brydges *et al.*[29] does show superior performance in a progressive fidelity group; however, this may have resulted from mixed practice, not progression, which has been shown to have advantages.[37]

Fidelity

A number of authors have borrowed a page from aviation psychology and suggested that fidelity can be assessed on two levels: 'engineering fidelity', which refers to whether the simulation looks realistic, and 'psychological fidelity', which concerns whether the simulator contains accurate simulations of the critical elements that will demand specific behaviours to complete the task.[12] The two types of fidelity can be dissociated. Moreover, there is some evidence from this review that psychological fidelity may be a more critical determinant of learning and transfer than engineering fidelity. The one study that specifically manipulated psychological fidelity[29] showed a clear advantage for greater fidelity. Two studies that contrasted expensive commercial simulators with 'kitchen table' simulators[24,26] (in both of the latter were based on careful task analysis and thus had high psychological fidelity) showed equivalent performance.

The nature of the task: 'sensory' versus 'simple' motor skills versus 'complex management'

The studies we have reviewed illustrate the difficulty of generalisation. Although the findings appear quite uniform, closer inspection indicates that training in each group of studies is directed at the acquisition of very different skills. The studies of training in auscultation skills exemplify the use of simulation to acquire perceptual skills. Other examples are reading electrocardiograms, interpreting skin lesions and synthesising laboratory test values. From the perspective of psychological fidelity, practice must include multiple different examples of confusable categories. Yet, as we indicated earlier, Harvey® contains only one example of most conditions.

Training in so-called 'simple' motor skills generally requires considerable practice. The development of skills for suturing, for example, requires many hours of practice and psychological fidelity demands that the critical elements – the resistance of the tissue, the 'feel' of the instruments – must closely resemble these elements of the task performed with real tissue. It is probably less important that the simulated tissue has the right colour or shape.

Conversely, the development of 'complex' skills may actually require relatively little practice. If each element of an action is relatively straightforward or has been previously learned (e.g. starting an i.v. line, taking a blood pressure), mastering the correct approach to crisis management may be mainly a matter of remembering the exact sequence in which actions must be taken. Although it is often assumed that HFS has a central role in this situation, the opposite may be true. Practice may involve nothing more than mental rehearsal of the steps.

The interaction with expertise

Although previous authors have indicated that there may well be an interaction with the level of the learner,[36] it is generally assumed that increasing expertise should be accompanied by increasing fidelity. However, this may not be the case at all. Two potential uses of HFS may lie in the teaching of an approach to rare and difficult problems, and in the teaching of the management of complex problems involving multiple health professional roles. Paradoxically, neither may require any sophistication of simulation. In the context of a rare problem, experts may regard the issue as a cognitive problem and simply 'think' their way to a solution. Their experience in the real setting is likely to be sufficiently extensive to make practice of the motor skills involved superfluous. In the context of a complex problem that requires input from multiple health professionals, the focus of activity is between professionals; the features of the simulator may be irrelevant.

Although the role played by each of these classes of variables is necessarily speculative, it does illustrate, as does the evidence reviewed in this paper, that the relationship between simulation fidelity and learning is not unidimensional and linear. It should come as no surprise, therefore, that the large investments required to acquire more advanced (higher-fidelity) simulators are not accompanied by commensurate increases in learning.

Sadly, this lesson was learned in aviation many years ago and we are now relearning it. As Salas *et al.* wrote with reference to aviation simulators:

> 'In sum, more is not necessarily better. Generally, the level of simulation fidelity does not translate to learning. High-fidelity simulations… should be used as determined by training and task requirements, cost and learning objectives.'[15]

Contributors: GN conceptualised and drafted the article, and finalised the submission. KD and LG contributed to the conceptualisation of the article and critiqued the

drafts. All authors approved the final manuscript for submission.

Acknowledgements: the authors acknowledge the continuing support of the Medical Council of Canada and Canadian Institutes of Health Research (GRN Holds a Canada Research Chair).
Funding: none.
Conflicts of interest: none.
Ethical approval: not applicable.

REFERENCES

1 Teteris E, Fraser K, Wright B, McLaughlin K. Does training learners on simulators benefit real patients? *Adv Health Sci Educ Theory Pract* 2012;**7** (1):137–44.

2 Cook DA, Hatala R, Brydges R, Zendejus B, Szostek JH, Wang AT, Erwin PJ, Hamstra SJ. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA* 2011;**306**:978–88.

3 Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM. Virtual reality training improves operating room performance: results of a randomised, double-blinded study. *Ann Surg* 2002;**236** (4):458–63; discussion 63–4.

4 Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB. Use of simulation-based education to reduce catheter-related bloodstream infections. *Arch Intern Med* 2009;**169** (15):1420–3.

5 Wayne DB, Didwania A, Feinglass J, Fudala MJ, Barsuk JH, McGaghie WC. Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case–control study. *Chest* 2008;**133** (1):56–61.

6 Ewy GA, Felner JM, Juul D, Mayer JW, Sajid AW, Waugh RA. Test of a cardiology patient simulator with students in fourth-year electives. *J Med Educ* 1987;**62** (9):738–43.

7 Butter J, McGaghie WC, Cohen ER, Kaye M, Wayne DB. Simulation-based mastery learning improves cardiac auscultation skills in medical students. *J Gen Intern Med* 2010;**25** (8):780–5.

8 Schuwirth LW, van der Vleuten CP. The use of clinical simulations in assessment. *Med Educ* 2003;**37** (Suppl 1):65–71.

9 Kneebone RL, Kidd J, Nestel D, Barnet A, Lo B, King R, Yang GZ, Brown R. Blurring the boundaries: scenario-based simulation in a clinical setting. *Med Educ* 2005;**39** (6):580–7.

10 Godden DR, Baddeley AD. Context dependent memory in two natural environments: on land and underwater. *Brit J Psychol* 1975;**66**:325–31.

11 Hockley WE. The effects of environmental context on recognition memory and claims of remembering. *J Exp Psychol Learn Mem Cogn* 2008;**34** (6):1412–29.

12 Maran NJ, Glavin RJ. Low- to high-fidelity simulation – a continuum of medical education? *Med Educ* 2003;**37** (Suppl 1):22–8.

13 Miller RB. Psychological Considerations in the Design of Training Equipment 1953. Wright-Patterson Air Force Base (OH): Wright Air Development Research Center (US); 2003. Report No. WADC-TR-54-563, AD 71202.

14 McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003–2009. *Med Educ* 2010;**44** (1):50–63.

15 Salas E, Bowers CA, Rhodenizer L. It is not how much you have but how you use it: toward a rational use of simulation to support aviation training. *Int J Aviat Psychol* 1998;**8** (3):197–208.

16 Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005;**27** (1):10–28.

17 Eva KW, Regehr G. Exploring the divergence between self-assessment and self-monitoring. *Adv Health Sci Educ Theory Pract* 2011;**16** (3):311–29.

18 Issenberg SB, McGaghie WC, Hart IR *et al.* Simulation technology for health care professional skills training and assessment. *JAMA* 1999;**282** (9):861–6.

19 de Giovanni D, Roberts T, Norman G. Relative effectiveness of high- versus low-fidelity simulation in learning heart sounds. *Med Educ* 2009;**43** (7):661–8.

20 Issenberg SB, Petrusa ER, McGaghie WC, Felner JM, Waugh RA, Nash IS, Hart IR. Effectiveness of a computer-based system to teach bedside cardiology. *Acad Med* 1999;**10** (Suppl):93–5.

21 Fraser K, Peets A, Walker I, Tworek J, Paget M, Wright B, McLaughlin K. The effect of simulator training on clinical skills acquisition, retention and transfer. *Med Educ* 2009;**43** (8):784–9.

22 Hatala R, Issenberg SB, Kassen B, Cole G, Bacchus CM, Scalese RJ. Assessing cardiac physical examination skills using simulation technology and real patients: a comparison study. *Med Educ* 2008;**42** (6):628–36.

23 Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, Hutchinson C. Assessment of technical skills transfer from the bench training model to the human model. *Am J Surg* 1999;**177** (2):167–70.

24 Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD. The effect of bench model fidelity on endourological skills: a randomised controlled study. *J Urol* 2002;**167** (3):1243–7.

25 Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu, Jarvi KA. The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. *Ann Surg* 2004;**240** (2):374–81.

26 Chandra DB, Savoldelli GL, Joo HS, Weiss ID, Naik VN. Fibreoptic oral intubation: the effect of model fidelity on training for transfer to patient care. *Anesthesiology* 2008;**109** (6):1007–13.

27 Sidhu RS, Park J, Brydges R, MacRae HM, Dubrowski A. Laboratory-based vascular anastomosis training: a randomised controlled trial evaluating the effects of bench model fidelity and level of training on skill acquisition. *J Vasc Surg* 2007;**45** (2):343–9.
28 Munz V, Kumar BD, Moorthy K, Bann S, Darzi A. Laparoscopic virtual reality and box trainers. *Surg Endosc* 2004;**18**:484–94.
29 Brydges R, Carnahan H, Rose D, Rose L, Dubrowski A. Coordinating progressive levels of simulation fidelity to maximise educational benefit. *Acad Med* 2010;**85** (5):806–12.
30 Morgan PJ, Cleave-Hogg D, McIlroy J, Devitt JH. Simulation technology: a comparison of experiential and visual learning for undergraduate medical students. *Anesthesiology* 2002;**96** (1):10–6.
31 Nyssen AS, Larbuisson R, Janssens M, Pendeville P, Mayne A. A comparison of the training value of two types of anaesthesia simulators: computer screen-based and mannequin-based simulators. *Anesth Analg* 2002; **94** (6):1560–5.
32 Baxter P, Norman G. Teaching critical management skills to senior nursing students: videotaped or interactive 'hands-on' instruction? *Nurs Educ Perspect* 2011; **67**:2406–13.
33 Wenk M, Waurick R, Schotes D, Gerdes C, van Aken HK, Popping DM. Simulation-based medical education is no better than problem-based discussions and induces misjudgement in self-assessment. *Adv Health Sci Educ Theory Pract* 2009;**14**(2):159–71.
34 Bruppacher HR, Alam SK, LeBlanc VR, Latter D, Naik VN, Savoldelli GL, Mazer CD, Kurrek MM, Jog HS. Simulation-based training improves physicians' performance in patient care in high-stakes clinical setting of cardiac surgery. *Anesthesiology* 2010;**112** (4):985–92.
35 van Merrienboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ* 2010;**44** (1):85–93.
36 Dubrowski A, Park J, Moulton CA, Larmer J, MacRae H. A comparison of single- and multiple-stage approaches to teaching laparoscopic suturing. *Am J Surg* 2007;**193** (2):269–73.
37 Hatala RM, Brooks LR, Norman GR. Practice makes perfect: the critical role of mixed practice in the acquisition of ECG interpretation skills. *Adv Health Sci Educ Theory Pract* 2003;**8** (1):17–26.

Received 28 September 2011; editorial comments to authors 11 November 2011; accepted for publication 5 January 2012