Technology-enhanced Simulation in Emergency Medicine: A Systematic Review and Meta-Analysis

Jonathan S. Ilgen, MD, MCR, Jonathan Sherbino, MD, MEd, and David A. Cook, MD, MHPE

Abstract

Objectives: Technology-enhanced simulation is used frequently in emergency medicine (EM) training programs. Evidence for its effectiveness, however, remains unclear. The objective of this study was to evaluate the effectiveness of technology-enhanced simulation for training in EM and identify instructional design features associated with improved outcomes by conducting a systematic review.

Methods: The authors systematically searched MEDLINE, EMBASE, CINAHL, ERIC, PsychINFO, Scopus, key journals, and previous review bibliographies through May 2011. Original research articles in any language were selected if they compared simulation to no intervention or another educational activity for the purposes of training EM health professionals (including student and practicing physicians, midlevel providers, nurses, and prehospital providers). Reviewers evaluated study quality and abstracted information on learners, instructional design (curricular integration, feedback, repetitive practice, mastery learning), and outcomes.

Results: From a collection of 10,903 articles, 85 eligible studies enrolling 6,099 EM learners were identified. Of these, 56 studies compared simulation to no intervention, 12 compared simulation with another form of instruction, and 19 compared two forms of simulation. Effect sizes were pooled using a random-effects model. Heterogeneity among these studies was large ($I^2 > 50\%$). Among studies comparing simulation to no intervention, pooled effect sizes were large (range $= 1.13$ to $1.48$) for knowledge, time, and skills and small to moderate for behaviors with patients (0.62) and patient effects (0.43; all $p < 0.02$ except patient effects $p = 0.12$). Among comparisons between simulation and other forms of instruction, the pooled effect sizes were small ($< 0.33$) for knowledge, time, and process skills (all $p > 0.1$). Qualitative comparisons of different simulation curricula are limited, although feedback, mastery learning, and higher fidelity were associated with improved learning outcomes.

Conclusions: Technology-enhanced simulation for EM learners is associated with moderate or large favorable effects in comparison with no intervention and generally small and nonsignificant benefits in comparison with other instruction. Future research should investigate the features that lead to effective simulation-based instructional design.

Technology-enhanced simulation has emerged as a cornerstone of many emergency medicine (EM) training programs. Simulation allows complex tasks to be deconstructed into more manageable learning objectives and fosters repeated practice on specific tasks to be deconstructed into more manageable learning objectives and fosters repeated practice on specific tasks that occur infrequently or expose patients to risk when performed by novice learners. Computer-based virtual reality simulators, robotic and static mannequins, artificial models, live animals, inert animal products, and human cadavers have all been employed as educational tools for these purposes. At the 2008 Academic Emergency Medicine (AEM) consensus conference entitled “The Science of Simulation in Healthcare: Defining and Developing Clinical Expertise,” participants identified future research directions for simulation technologies with the aim of promoting learning across the education spectrum (i.e., undergraduate to postgraduate to physicians in practice). These topics included effective instructional design for teaching individual EM expertise, teamwork and communication, procedural and surgical skills, and systems-based issues (e.g., patient safety, disaster management).

Two recent meta-analyses across all clinical disciplines (including EM) have demonstrated that when compared with no intervention or baseline performance, simulation has large effects on outcomes of knowledge, skills, and behavior, and moderate effects on patient-related outcomes. However, the educational effect varied according to clinical topics and learners, among other factors. Despite a significant number of EM-specific studies involving simulation, the instructional effect specific to this discipline remains unknown. If simulation is to continue to play a dominant and growing role in EM teaching, evidence for its effectiveness and insight into appropriate instructional designs (i.e., the systematic and theory-based arrangement of features that comprise and influence a learning activity) is needed.

While several review articles have proposed principles for the design of EM simulation curricula, these reviews were limited by incomplete accounting of existing studies, limited assessment of study quality, or no quantitative pooling.

A systematic synthesis of evidence regarding simulation-based instruction that addresses the unique learning issues specific to EM (e.g., variety of learners, team-based multidisciplinary care, systems of emergent care) would be of great value to the education community. The goal of this study, therefore, was to evaluate the effectiveness of technology-enhanced simulation training in EM and to identify the instructional design features associated with improved outcomes.

**METHODS**

This review was planned, conducted, and reported in adherence to PRISMA standards of quality for reporting meta-analyses.

**Questions**

We sought to answer two questions:

1. What is the effectiveness of technology-enhanced simulation for teaching EM in comparison with no intervention or nonsimulation instruction?

2. Which instructional design features are associated with improved learning outcomes in simulation-based EM instruction?

**Study Eligibility**

We defined technology-enhanced simulation as an educational tool or device with which the learner physically interacts to mimic an aspect of clinical care. Such tools include robotic and static mannequins, partial task trainers, cadavers, live animals or animal parts, and computer-based virtual reality simulators. We included comparative studies that evaluated technology-enhanced simulation for training EM health professionals at any stage of training or practice, including physicians, midlevel providers, nurses, paramedics, and emergency medical technicians (including those serving in the military). We included studies making comparison with no intervention (i.e., a control arm or preintervention assessment), an alternate simulation, or a nonsimulation instructional activity. We made no exclusions based on language or year of publication. Human standardized patients were not included as technology-enhanced simulation.

**Study Identification**

With the assistance of a research librarian, we searched MEDLINE, EMBASE, CINAHL, PsychINFO, ERIC, Web of Science, and Scopus, with a last search date of May 11, 2011. Search terms included “simulation,” “computer simulation,” “learning,” “training,” “skill,” “mannequin” or “manikin,” and “assessment,” among others; the complete search strategy has been published previously. We also reviewed all articles published in two journals devoted to health professions simulation (Simulation in Healthcare and Clinical Simulation in Nursing), and the references of several key review articles.

**Study Selection**

Two reviewers independently screened for inclusion the titles and abstracts of all potentially eligible studies. We then obtained the full text of all articles that could not be confidently excluded and reviewed these for definitive inclusion or exclusion, again independently and in duplicate, with conflicts resolved by consensus. The chance-adjusted interrater agreement for these two steps, determined using the intraclass correlation coefficient (ICC), was 0.69 (95% confidence interval [CI] = 0.67 to 0.71). As part of the data extraction (below), we subsequently identified all studies containing EM learners for inclusion in the present review (ICC = 0.94; 95% CI = 0.92 to 0.95).

**Data Extraction**

We collected information from each study using an electronic data abstraction form. Two independent reviewers abstracted all information for which reviewer judgment was required, with conflicts resolved by consensus. We abstracted information on the training level of learners and the clinical topic being taught. We also coded features of the simulation training, including instructional design features of team training, feedback, mastery learning, curricular integration, and repetitive practice. We evaluated study methods including study...
design, method of group assignment, and blinding of assessments using the Medical Education Research Study Quality Instrument (MERSQI)\(^15\) and an adaptation of the Newcastle-Ottawa Scale (NOS) for cohort studies.\(^16\) The psychometric properties of these two instruments have been described previously.\(^7, 15, 16\)

We abstracted information separately for learning outcomes of knowledge, skills, behaviors with patients, and direct effects on patients. We further classified skills as time (to complete the task), process (e.g., observed proficiency or economy of movements), and product (e.g., successful task completion or major errors).\(^7\) We similarly classified behaviors with real patients as time and process (see Data Supplement S1, available as supporting information in the online version of this paper, for definitions of instructional design key features and outcomes).

Data Synthesis
We planned for both quantitative and qualitative evidence synthesis. Because there was high between-study inconsistency in prior analyses,\(^7\) we planned to use random effects meta-analysis to quantitatively pool results, organized by comparison (comparison with no intervention, another form of technology-enhanced simulation, or another form of instruction). Additionally, we planned subgroup analyses based on trainee level, topic, and the key instructional design features noted above. Finally, we conducted subgroup analyses based on key study design elements (randomization and blinding of assessment) and sensitivity analyses, excluding the results of studies with imprecise effect size calculations.\(^7\)

For each comparison, we calculated a standardized mean difference (Hedges’ \(g\) effect size) using standard techniques.\(^17–19\) For articles reporting insufficient information to calculate an effect size, we requested additional information from the authors. We quantified between-study inconsistency (heterogeneity) using the \(I^2\) statistic,\(^20\) which estimates the percentage of variability not due to chance. \(I^2\) values > 50% indicate large inconsistency. We used SAS 9.1.3 (SAS Institute, Cary, NC) for all analyses. Statistical significance was defined by a two-sided alpha of 0.05, with no adjustment for multiple comparisons. Interpretations of clinical significance emphasized CIs in relation to Cohen’s effect size classifications (>0.8 = large, 0.5 to 0.8 = moderate, 0.2 to 0.5 = small, and <0.2 = negligible).\(^21\) Studies that could not be combined in a quantitative synthesis of results were analyzed repeatedly to identify key themes, which were then summarized in a narrative synthesis.

RESULTS

Trial Flow
The trial flow is shown in Figure 1. We identified 10,903 potentially relevant articles. From these, we identified 85 studies with eligible EM learners, reflecting data from 6,099 trainees. Fifty-six of these studies compared technology-enhanced simulation with no intervention (control group or baseline assessment), 12 made comparison with other active instruction (four comparisons with standardized patients and eight with non-simulation instruction), and 19 compared two different forms of technology-enhanced simulation. Several studies had multiple comparison arms. Table 1 summarizes the key features of these studies, and Data Supplement S2 (available as supporting information in the online version of this paper) provides detailed information for each included study.

Study Characteristics
Studies evaluating the effect of technology-enhanced simulation in EM date from 1985. Two studies were published in a non-English language (one in Norwegian, one in French). Most (\(n = 71; 84\%)\) were published since 2002. Learners in the included studies reflect the complex fabric of health professionals involved in emergency care, including physicians, midlevel providers, nurses, paramedics, respiratory therapists, emergency medical technicians, and military medics. Studies included both students and in-practice professionals in most disciplines.

More than half of the studies (45 of 85, \(n = 2,501\)) had a focus on procedural training, and nearly half of these studies (21 of 45, \(n = 1,438\)) involved airway management training. Twenty-eight studies had a focus on resuscitation or trauma training, and 11 addressed systems of emergent care. Resident physicians (1,615 participants in 39 studies) and prehospital providers (1,615 participants in 32 studies) were the dominant learners represented in the studies.

Of the 85 studies, 27 provided high-intensity feedback, and five involved multiple (more than 10) repetitions. Six studies used a mastery learning model, in which completion of simulation training required
achievement of preestablished criterion-based objectives. Nearly half of the studies (40 of 85, n = 3,362) involved learning in a team and approximately half of these studies (22 of 40, n = 2,009) involved teams that were multidisciplinary. Twenty-three studies (n = 2,210) integrated their simulation training as part of a larger curriculum.

Study Quality
The methodologic quality of the included studies is summarized in Table 2. The number of participants ranged from 7 to 497, with a median of 52. Over half the studies employed a single-group pre-post design, while 25 were randomized. Twenty-eight studies used objective assessment measures, and approximately two-thirds of the studies (58 of 85) had response rates ≥75%. Nineteen of the studies were multi-institutional.
Synthesis: Comparison of Simulation Versus No Intervention

Figures 2 and 3 summarize the meta-analysis of 56 studies comparing technology-enhanced simulation versus no intervention for outcomes of knowledge, process skills, products, time, learner behavior, and patient outcomes. Overall, these analyses demonstrate that simulation training was associated with moderate to large effect sizes, but with substantial inconsistency from study to study. There were no significant interactions.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. studies (No. trainees)</th>
<th>Favors No Instruction</th>
<th>Favors Simulation</th>
<th>Pooled Effect Size (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>14 (1384)</td>
<td></td>
<td></td>
<td>1.48 (0.89, 2.08)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Time</td>
<td>9 (257)</td>
<td></td>
<td></td>
<td>1.22 (0.25, 2.19)</td>
<td>0.01</td>
</tr>
<tr>
<td>Skill-Process</td>
<td>37 (2377)</td>
<td></td>
<td></td>
<td>1.13 (0.92, 1.35)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Skill-Product</td>
<td>5 (300)</td>
<td></td>
<td></td>
<td>2.09 (0.90, 3.27)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Behavior-process</td>
<td>6 (308)</td>
<td></td>
<td></td>
<td>0.62 (0.22, 1.02)</td>
<td>0.002</td>
</tr>
<tr>
<td>Patient effect</td>
<td>5 (482)</td>
<td></td>
<td></td>
<td>0.44 (-0.12, 0.99)</td>
<td>0.12</td>
</tr>
</tbody>
</table>

**Figure 2.** Random-effects meta-analysis of simulation training compared with no intervention: products, time, learner behavior, and patient outcomes. Simulation compared with no intervention; positive numbers favor the simulation intervention. P-values reflect statistical tests exploring the effect of simulation training compared with no intervention.

**Figure 3.** Random-effects meta-analysis of simulation training: knowledge and process. (A) Knowledge outcomes. (B) Skill process outcomes. Simulation compared with no intervention; positive numbers favor the simulation intervention. P-values reflect statistical tests exploring the differential effect of simulation training (i.e., interaction) for study subgroups. Participant groups are not mutually exclusive; thus, no statistical comparison is made and the number of trainees is not reported. Some features could not be discerned for all studies; hence, some subgroups do not sum to the total number of studies. MERSQI = Medical Education Research Study Quality Instrument; NOS = Newcastle-Ottawa Scale.
demonstrated in subgroup analyses exploring curricular integration, feedback, and repetition. Sensitivity analyses did not alter study conclusions (data not shown).

**Knowledge.** Fourteen studies (1,384 participants) reported comparisons to a no-intervention control group or a preintervention assessment of knowledge as an outcome (Figure 3A). The pooled effect size of 1.48 (95% CI = 0.89 to 2.08; \( p < 0.001 \)) suggests a large favorable association.\(^{21}\) However, individual effect sizes ranged from –0.19 to 3.38, and significant statistical heterogeneity (\( I^2 = 98\% \)) suggests high inconsistency between studies. To explore that inconsistency, we subanalyzed these studies based on instructional design features (team training, curricular integration, high-quality feedback, and multiple repetitions). In each case, the interaction of a subgroup with the overall effect was not statistically significant, suggesting that these instructional design features were not associated with the overall outcomes. The one study with randomized group assignment demonstrated a large effect size of 1.70 (95% CI = 1.09 to 2.31). Higher quality studies (i.e., blinded assessment and higher NOS and MERSQI scores) had larger effect sizes, although this interaction did not reach statistical significance. Sensitivity analyses excluding three studies with imprecise effect size calculations showed an effect size of 1.69 (95% CI = 0.81 to 2.57).

**Process.** Thirty-seven no-intervention comparison studies (2,377 participants) reported measures of process skills (i.e., how a clinical task is completed; Figure 3B). Effect sizes were all positive, ranging from 0.16 to 2.97. The pooled effect size of 1.13 (95% CI = 0.92 to 1.35; \( p < 0.001 \)) suggests large benefits, but there was again a high degree of inconsistency between studies (\( I^2 = 92\% \)). There were no statistically significant interactions in planned subgroup analyses. The lone randomized trial found a small effect of 0.16 (95% CI = –0.32 to 0.63), while three nonrandomized two-group studies (183 participants) demonstrated a large effect size of 1.22 (95% CI = 0.25 to 2.18). Sensitivity analyses excluding 13 studies with imprecise effect size calculations showed an effect size of 1.30 (95% CI = 1.01 to 1.60).

**Skill Product, Time, Learner Behavior, and Patient Effect Outcomes.** We performed analyses pooling effect sizes for outcomes concerning products of a learner’s performance in a simulation setting (e.g., procedural success or quality of the completed product), time to complete required tasks (e.g., procedure time, test ordering), behaviors with real patients, and effects on patients (see Figure 2). Because we found few studies, we limited subgroup analyses to key methodologic variations, namely group allocation (randomized/nonrandomized) and blinding of assessment. These subgroup analyses, and the sensitivity analyses, did not appreciably alter study conclusions (data not shown).

**Synthesis: Comparative Effectiveness Research**

**Comparison of Simulation Versus Other Forms of Instruction.** Twelve studies evaluated the impact of technology-enhanced simulation on learning when compared with another form of instruction distinct from technology-enhanced simulation, including four studies making comparison with standardized patients and eight studies making comparison with nonsimulation instruction. The pooled effect sizes were small for all outcomes except product skills, and none were statistically significant (see Figure 4). Planned subgroup and sensitivity analyses based on key design features did not appreciably alter conclusions (data not shown).

**Comparison between alternate types of technology-enhanced simulation.** Direct comparisons between alternate types of technology-enhanced simulation can illustrate how instructional design features or different types of educational media affect learning outcomes. There were 19 studies that used designs in which one approach to simulation-based education was compared with another (see Table 3).\(^{23–41}\) Twelve of the 19 studies included prehospital providers as participants, and nine of these studies focused on airway management training. Because the research questions in these 19 studies varied substantially, we were unable to provide a quantitative synthesis of results. Instead, we report a qualitative synthesis of common themes.

Two studies evaluated the effect of different types of feedback and found that this intervention generally led to improved performance and improved self-efficacy.\(^{23, 24}\) Lammers\(^{23}\) used a “pause and perfect” model for teaching posterior epistaxis management, with faculty supervising participants during each step of a procedure, and pausing to correct mistakes when they were evident. When compared with trainees who received traditional observation, intervention subjects were more efficient and had better technique; this effect, however, was not sustained when subjects were retested at 3 months.

![Figure 4](image-url) Random-effects meta-analysis of simulation training versus other forms of instruction. Simulation compared with other forms of instruction; positive numbers favor the simulation intervention. \( p \)-values reflect statistical tests exploring the effect of simulation training compared with another form of instruction for each of the learning outcomes listed.
### Table 3
Studies That Compared Alternate Types of Technology-enhanced Simulation

<table>
<thead>
<tr>
<th>First Author, Year</th>
<th>Participants: N; Type</th>
<th>Clinical topic</th>
<th>RCT?</th>
<th>Intervention 1</th>
<th>Intervention 2</th>
<th>Design Features</th>
<th>Outcomes and Effect Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lammers, 200823</td>
<td>28; PG</td>
<td>Epistaxis</td>
<td>Yes</td>
<td>“Pause and perfect”</td>
<td>Traditional teaching</td>
<td>F, M</td>
<td>ST: 0.59; SPc: 0.40</td>
</tr>
<tr>
<td>Auerbach, 201124</td>
<td>151; PG</td>
<td>Resuscitation</td>
<td>No</td>
<td>Repetitive simulation</td>
<td>Traditional simulation</td>
<td>F, T</td>
<td>R: 0.38</td>
</tr>
<tr>
<td>Agazio, 200225</td>
<td>60; MD, RN, PP</td>
<td>Venous access</td>
<td>Yes</td>
<td>Virtual reality</td>
<td>Plastic IV task trainer</td>
<td>F</td>
<td>R: 0.68; ST: 0.63; SPc: -0.63</td>
</tr>
<tr>
<td>Davis, 200726</td>
<td>120; RN, PP</td>
<td>Airway management</td>
<td>No</td>
<td>Mannequin, novel curriculum</td>
<td>Cadaver, traditional curriculum</td>
<td>C, T</td>
<td>SPc: 0.28; P: 0.57</td>
</tr>
<tr>
<td>Hoadley, 200927</td>
<td>53; MD, RN, PP, O</td>
<td>Resuscitation</td>
<td>Yes</td>
<td>High fidelity</td>
<td>Low fidelity</td>
<td>M, F, T</td>
<td>R: -0.01; K: 0.01; SPc: 0.45</td>
</tr>
<tr>
<td>Thomas, 201028</td>
<td>100; PG</td>
<td>Resuscitation</td>
<td>Yes</td>
<td>High fidelity</td>
<td>Low fidelity</td>
<td>F, T</td>
<td>ST: 0.31; SPc: 0.70</td>
</tr>
<tr>
<td>Lee, 201033</td>
<td>30; PP</td>
<td>Resuscitation</td>
<td>Yes</td>
<td>Resuscitation labs</td>
<td>Traditional ACLS</td>
<td>F, T</td>
<td>ST: 1.36; SPc: 0.57</td>
</tr>
<tr>
<td>Stratton, 199130</td>
<td>125; PP</td>
<td>Airway management</td>
<td>Yes</td>
<td>Mannequin plus cadaver</td>
<td>Mannequin</td>
<td>C, F, M, R</td>
<td>P: 0.03</td>
</tr>
<tr>
<td>Trooskin, 199231</td>
<td>26; PP</td>
<td>Airway management</td>
<td>Yes</td>
<td>Mannequin (A)</td>
<td>Animal (B), Cadaver (C)</td>
<td>SPd: AB 1.31, AC 1.1, BC 0.46</td>
<td></td>
</tr>
<tr>
<td>Cummings, 200632</td>
<td>140; MS, PG, MD, RN, PP, O</td>
<td>Airway management</td>
<td>Yes</td>
<td>Animal</td>
<td>Mannequin</td>
<td>R: 0.47; ST: 0.47</td>
<td></td>
</tr>
<tr>
<td>Cho, 200933</td>
<td>49; PG, MD, RN, PP</td>
<td>Airway management</td>
<td>Yes</td>
<td>Animal</td>
<td>Plastic model</td>
<td>R: 1.1</td>
<td></td>
</tr>
<tr>
<td>Yang, 201034</td>
<td>56; MD, RN</td>
<td>Airway management</td>
<td>Yes</td>
<td>Cadaver</td>
<td>Mannequin (A, B, C)</td>
<td>R: A 0.79, B 0.76, C 1.77</td>
<td></td>
</tr>
<tr>
<td>van Stralen, 199535</td>
<td>88; RN, PP, O</td>
<td>Airway management</td>
<td>No</td>
<td>Mannequin</td>
<td></td>
<td>ST: 1.21</td>
<td></td>
</tr>
<tr>
<td>Bond, 200636</td>
<td>62; PG</td>
<td>Resuscitation</td>
<td>Yes</td>
<td>Cognitive debriefing</td>
<td>Technical debriefing</td>
<td>T</td>
<td>R: -0.55</td>
</tr>
<tr>
<td>Low, 200837</td>
<td>49; MS, PP</td>
<td>Airway management</td>
<td>Yes</td>
<td>Video laryngoscope</td>
<td>Macintosh laryngoscope</td>
<td>ST: -1.6; SPd: 0.56</td>
<td></td>
</tr>
<tr>
<td>Youngquist, 200838</td>
<td>245; PP</td>
<td>Airway management</td>
<td>No</td>
<td>Lecture, facilitated practice with instructor</td>
<td>Self-directed practice</td>
<td>F</td>
<td>SPd: 0.56</td>
</tr>
<tr>
<td>Girzadas, 200939</td>
<td>45; PG</td>
<td>Ultrasound, OB/GYN</td>
<td>Yes</td>
<td>Dynamic images, pelvic ultrasound trainer</td>
<td>Static images</td>
<td>R: 0.95; ST: 0.60</td>
<td></td>
</tr>
<tr>
<td>Hein, 201040</td>
<td>55; PP</td>
<td>Airway management</td>
<td>Yes</td>
<td>Just-in-time retraining</td>
<td>Mannequin, no retraining</td>
<td>ST: 0.64; SPc: 0.69</td>
<td></td>
</tr>
<tr>
<td>Orde, 201041</td>
<td>120; MS, RN</td>
<td>Airway management</td>
<td>Yes</td>
<td>4-stage LMA insertion, Two-stage LMA insertion</td>
<td></td>
<td>ST: 0.29; SPc: 0.19; SPd: 0.15</td>
<td></td>
</tr>
</tbody>
</table>

Participants: MD = practicing physicians; MS = medical students; PG = post-graduate resident physicians; RN = registered nurse; PP = prehospital provider; O = other
ACLs = Advanced Cardiac Life Support; LMA = laryngeal mask airway; OB/GYN = obstetrics and gynecology; RCT = randomized controlled trial.
Design features: C = curricular integration; F = feedback; M = mastery learning; R = repetition; T = team training.
Outcomes: K = knowledge; P = patient effect; R = reaction; ST = skill time; SPc = skill process; SPd = skill product. For example, the study by Stratton compared a mannequin plus cadaver training versus cadaver alone, using outcomes of patient effects, and found a negligible effect (0.03) favoring the mannequin-cadaver approach.
Auerbach et al.\textsuperscript{24} studied the effect of feedback (i.e., debriefing) after a simulated pediatric resuscitation, followed by immediate repetition of a second simulated pediatric resuscitation to reinforce learning, in comparison with feedback alone (no repetition). Trainees who repeated the scenario reported greater improvements in knowledge and skills.

One study evaluated a mastery learning model, in which learners needed to meet a clearly defined standard of performance before qualifying or advancing to the next task and compared this to a model of non-mastery learning.\textsuperscript{23} This study found that those trained using the mastery model completed the test task faster (effect size $= 0.59$, $p = 0.003$) and with better technique (effect size $= 0.40$, $p < 0.001$) compared with the non-mastery group. As above, these positive effects were not seen when subjects were tested 3 months later.

Ten studies explored issues of fidelity and realism between simulation-based curricula. Five of these compared “low-fidelity” to “high-fidelity” scenario-based teaching.\textsuperscript{25–29} While none clearly described the features that defined fidelity as “low” or “high,” it was implicit that high-fidelity simulation more closely resembled a living human patient or actual clinical setting. For example, Hoardle\textsuperscript{27} compared knowledge and reaction outcomes between participants in an Advanced Cardiac Life Support course who were divided into two groups, one that was asked to gather all data from a mannequin and its monitoring devices, and another that was provided with all of this information by course instructors while using an inanimate mannequin. All but one study demonstrated a positive association between higher fidelity and process skills (effect size $= −0.63$ to 0.70, five studies), and the single study with negative effect sizes had positive associations with both learner reactions and time skill outcomes.\textsuperscript{25} Likewise, higher fidelity was associated with uniformly better time outcomes (effect size $= 0.31$ to 1.36, three studies). Only one study measured patient effects,\textsuperscript{26} and although it reported moderate benefits to learning (effect size $= 0.57$), the comparison curriculum differed in more than just the level of fidelity. The other five studies compared different simulation modalities for procedural instruction purposes (i.e., mannequin vs. animal or cadaver models).\textsuperscript{30–34}

Learners generally preferred animal or cadaver models to mannequin models for teaching intubation or cricothyroidotomy skills (reaction effect size $= 0.47$ to 1.77, three studies).\textsuperscript{32–34} Only two studies measured patient effects when comparing cadaver or animal models to mannequin-based instruction and found these to be similar\textsuperscript{35} or in favor of the nonmannequin modalities.\textsuperscript{31} The latter study also measured skill product outcomes, and these favored the mannequin format (effect size $= 1.31$ in comparison with animal; 1.1 in comparison with cadaver).\textsuperscript{31} Finally, a study comparing a white tail deer model to a mannequin model for teaching intubation found a moderate effect in favor of the animal model (effect size $= 0.47$) concerning time to complete this task.\textsuperscript{32}

**DISCUSSION**

In comparison with no intervention, technology-enhanced simulation for EM learners is associated with large favorable effects for knowledge and skills and small to moderate benefits for patient-related outcomes. Although between-study inconsistency is high, subgroup analyses exploring the possible influence of study methods and instructional design features reveals similarly positive associations.

In contrast, when compared to other forms of instruction, the effects of technology-enhanced simulation are smaller and not statistically significant. Among the few studies comparing different approaches to simulation-based education, feedback and mastery learning appear to improve learning. Higher fidelity appears to confer small to moderate benefits in skill process and time when compared with programs using lower fidelity, but studies are few and the defining features of “high” and “low” fidelity are vague and vary from study to study.

**Integration With Prior Work**

Our study supports the assertions of the 2008 AEM consensus conference, namely that simulation can help EM learners achieve individual expertise, foster teamwork and communication, strengthen procedural and surgical skills, and promote systems-based issues.\textsuperscript{3} Our results also align with prior meta-analyses of simulation in health professions education in showing large effects when compared with a placebo or no intervention.\textsuperscript{7, 16, 42, 43}

These findings confirm that simulation works, in the sense that it is better than no intervention or baseline assessments.

Our results concerning studies that compared two forms of technology-enhanced simulation seem to suggest that higher-fidelity simulation offers benefit over lower-fidelity simulation. However, these findings should be interpreted with caution. Comparisons between high and low fidelity have generally failed to demonstrate significant differences in learning outcomes between the two, as noted in two recent reviews,\textsuperscript{44, 45} and recent studies suggest that the degree of realism required of a simulation is a function of the learning task and context. For example, teaching ureteroscopy using a plastic straw may confer the same degree of performance improvement as an expensive task simulator.\textsuperscript{44, 45}

Thus, given the variety of clinical topics explored in our data set, and the challenges in clearly defining the salient features of “fidelity,” it is difficult to draw clear conclusions regarding whether higher fidelity or enhanced realism truly translates to greater learning gains.

**Implications for Current Practice and Future Research**

The consistent results among studies comparing technology-enhanced simulation to no intervention suggest that future studies with similar designs are likely to show similar findings. Further comparisons with no intervention will contribute little to the advancement of EM education. Conversely, the number of EM studies exploring simulation versus other teaching techniques was small, and pooled effect sizes failed to reach statistically significant in any of the predefined analyses. This may suggest that multiple forms of active instruction—whether simulation-based or not—similarly affect learning outcomes. Alternatively, given the small number of
studies and heterogeneity of design, it is possible that the presence of an effect went undetected. Pending further research, available evidence supports the use of simulation for the training of procedural tasks.

The substantial variation among studies comparing different forms of simulation precluded a quantitative synthesis. The few themes that emerged support no clear recommendations for the optimal design of EM simulation-based instruction, although they suggest several directions for future research. However, principles derived from examination of studies outside of EM do support several design principles. The 2008 AEM consensus conference focused on “The Science of Simulation in Healthcare.” If we truly hope to succeed in understanding the science of simulation, then we must approach research in this field as we do other scientific fields—including building on prior work, testing theory-informed hypotheses, and seeking to identify generalizable lessons that extend beyond the immediate research context. When framed by established scales of education research quality (MERSQI, NOS), the methodologic rigor of the included studies is relatively low, a finding that is neither surprising nor unique to EM education or education generally. Our findings suggest that future work might be strengthened by design features such as randomization and blinded outcome assessment, collecting validity evidence for outcome measures, and focusing on measurements of learner behaviors and patient effects. However, rigorous research methods will not compensate for failure to focus on a research question grounded in prior empiric work and robust conceptual frameworks.

We identify at least two themes for future research. First, when comparing simulation to other forms of instruction or to other types of simulation, work should be directed toward the instructional design features that lead to the greatest improvements in learning and performance. For example, to study the design effect of repetition in a simulation-based program of a particular procedure, investigators could randomize subjects into different levels of repetition (leaving the remainder of the curriculum identical between these arms) and then collect blinded outcomes such as process, time, task products, and behaviors with actual patients. Second, in light of the comparisons between “high”- and “low”-fidelity simulation, the substantial costs associated with technology-enhanced simulation need to be justified if cheaper alternate modes of instruction can achieve similar results. Researchers should evaluate costs of both the simulation and the comparison training.

LIMITATIONS

There was a large degree of between-study variation across all analyses. Such inconsistency likely stems from differences in the instructional modalities and instructional designs and the type of learners enrolled. However, we enhanced conceptual similarities within analyses by clustering studies by comparison intervention. Moreover, subgroup analyses exploring this heterogeneity demonstrated consistent findings for different study designs and instructional designs. Many of the studies we evaluated had methodologic limitations or failed to clearly describe the context, instructional design, or outcomes. This limits the inferences that we can draw from these studies.

The subgroup analyses should be interpreted with caution in light of the lack of a priori hypotheses for many analyses, the inconsistent findings across some outcomes, and the limitations of between-study comparisons. For example, when examining skill process outcomes, we found (contrary to expectation) that studies with lower repetition had higher effect sizes. These and other counterintuitive findings could be due to confounding (features related to repetition that consistently affect the outcomes), chance, bias, or a true effect.

The paucity of common conceptual frameworks among studies comparing technology-enhanced simulation to other forms of instruction, or between two forms of simulation, precluded our ability to perform meta-analyses for these studies. This limits our ability to draw firm conclusions regarding ways in which EM simulation educators can design curricula to maximize learning.

CONCLUSIONS

Technology-enhanced simulation for EM learners is associated with moderate or large favorable effects in comparison with no intervention and generally small and nonsignificant benefits in comparison with other instruction. Yet, current evidence does little to clarify how to optimally design simulation-based instruction to maximize learning efficiency or minimize costs. Such evidence will be vitally important as educators integrate simulation-based and other educational approaches to improve provider competence and patient safety.

The authors thank Ryan Brydges, Stan Hamstra, Rose Hatala, Jason Szostek, Amy Wang, Benjamin Zendejas, and Pat Erwin for their efforts in initial study selection and data abstraction.

References


Supporting Information

The following supporting information is available in the online version of this paper:

Data Supplement S1. Instructional design key features.

Data Supplement S2. List of all included studies.

VIRTUAL ISSUES

"Virtual issues" now are a key feature of the journal’s new home page on our publisher’s recently implemented platform, Wiley Online Library (WOL). A virtual issue is basically just a collection of articles on a given topic - so the EMS virtual issue, for example, will be a running compilation of all EMS articles that we publish. The idea is that a reader will go there to look for a particular article, but then will see our other offerings on that topic as well - increasing our full-text download numbers and helping ensure the broadest dissemination of our authors' work.

The first Geriatrics Virtual Issue is online. Go to the journal’s home page on the WOL platform, see “Special Features” on the left-hand side and click on the feature. Three additional virtual issues on ultrasound, toxicology and injury prevention are up and running. Again, consult the "Special Features" area and click on the desired issue.


Stay tuned for updates!