Virtual Simulation and Serious Games for Medical Education: A Review of the Literature and Development of a Virtual Peritoneal Dialysis Simulator

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:27007751">http://nrs.harvard.edu/urn-3:HUL.InstRepos:27007751</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>


Table 1. The serious games found in our review, organized according to the author and publication year, game title, game design, study design, outcome assessed, subjects/intended audience, outcome category, and study findings.

<table>
<thead>
<tr>
<th>Authors and publication year</th>
<th>Serious Title</th>
<th>Game Design</th>
<th>Study Design</th>
<th>Subjects / Intended audience</th>
<th>Outcomes</th>
<th>Outcome Category</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aki et al.22 2008</td>
<td>Guide-O-Game©</td>
<td>Following timed format of TV game shows, teams compete to answer questions based on Clinical Practice Guidelines.</td>
<td>Survey</td>
<td>Internal Medicine residents in USA; n=30.</td>
<td>Feasibility, curriculum integration, survey.</td>
<td>1, 2, 3, 6</td>
<td>Integrating four weekly 60-minute sessions within the curriculum was feasible. Participants found the game to be fun, engaging, and raised interest in guidelines recommendations. Criticism included some recommendations being irrelevant to practice, lack of rationale for each recommendation, and lack of a review period at the end of each session.</td>
</tr>
<tr>
<td>Al-Dahr et al.16 2015</td>
<td>Decision Simulation</td>
<td>VE using looped, case-based, branch-learing of pharmacotherapy decision-making for patients with atrial fibrillation.</td>
<td>RCT: SG vs. C (small-group discussion)</td>
<td>Pharmacy students; n=119.</td>
<td>Pre/post MCQ, survey.</td>
<td>1, 2, 3, 5</td>
<td>Both groups showed a significant increase in post-test vs. pre-test scores (SG: 77.7±12.6, 66.5±13.6, p&lt;0.001; C: 61.1±12.9, 74.8±11.7, p=0.001). C post-test scores were significantly higher than SG post-test scores (66.5±13.6 vs. 74.8±11.7, p&lt;0.001). More C group testers had a score increase (83.3% vs 69.5%, respectively, p&lt;0.013). Both groups would recommend the method, said that the method reinforced knowledge (4.0 for SG vs. 4.4 for C, p=0.034) and added new knowledge (4.0 for SG vs. 4.4 for C, p&lt;0.01).</td>
</tr>
<tr>
<td>Amer et al.10 2011</td>
<td>Interactive dental video game</td>
<td>VE of three-step resin bonding procedure. Level-based.</td>
<td>RCT: SG vs. C (clinical video)</td>
<td>Dental students in USA; n=80.</td>
<td>Pre/post MCQ, dental exercise, satisfaction survey.</td>
<td>1, 2, 3, 5, 6</td>
<td>No significant difference for SG vs. C in dentin bonding exercise (median bond strength 12.8 MPa vs. 12.9 MPa, p=0.97), or in MCQ (median score 8, 8, p=0.27). 97% enjoyed the SG. 95% felt the SG improved understanding. Compared to conventional lecture, 77% found the SG more informative, 86% preferred it, 26% thought it should replace lecture.</td>
</tr>
<tr>
<td>Andreotta et al.16 2010</td>
<td>CAVE Triage Training</td>
<td>VE where users must assess safety of scene, seek resources, assess victim status and priorities, and tag victims in a timely fashion.</td>
<td>RCT: SG vs. C (live simulation with same SG scenarios)</td>
<td>Emergency residents in USA; n=15.</td>
<td>Mean triage score, correct triage; pre/post MCQ at 2 weeks.</td>
<td>5</td>
<td>No significant differences across groups. Mean scores and total of correct triages showed small effect size favoring SG [Cohen’s d=0.25, 3.47 (C =0.41) vs. 3.55 SG (C =0.17); t=0.27, 11.38/61% CI(1.92) vs. 11.86/85% SG (± 1.57)]. Post-test MCQ scores were higher and had a greater increase in scores compared to pre-test scores with a large effect size favoring C [Cohen’s d=0.63, 18.50/74% C (+2.62) vs. 16.71/67% SG (+3.04), p=0.48].</td>
</tr>
<tr>
<td>Antoniou et al.24 2014</td>
<td>Second Life (SL) Periodontology Virtual Patient</td>
<td>VE with embedded MCQ about clinical/surgical, critical management of periodontology cases. Two versions of SG: Web vs. SL.</td>
<td>Survey: Web vs. SL-based versions of the SG</td>
<td>Dental students; n=9.</td>
<td>Survey.</td>
<td>1, 2, 3</td>
<td>Participants reported easy-to-use interface, easy access with Web SG. Most described SL as more interactive, immersive, and less challenging, due to prompting visual cues, but complained about access, un-optimized graphics and slow functionality. Instructions/content reported as clear, but some requested a user-driven, non-linear format, vs. MCQ format.</td>
</tr>
<tr>
<td>Alack et al.10 2009</td>
<td>The Disaster Management Course</td>
<td>Multi-player VE of mass casualty exercise teaching disaster management and inter-professional skills, integrated within online course. Embedded interactive components.</td>
<td>Quasi-experimental</td>
<td>Several inter-professional groups in Canada, n=74.</td>
<td>Pre/post MCQ, self-described competency survey.</td>
<td>1, 2, 3, 5</td>
<td>36% of participants did not complete the course, citing technical problems (50%) and lack of time (40%) as reasons. 77% completed pre- and post-competency survey (pre mean 2.72 (SD 0.79), post mean 4.08 (SD 0.72)). Participants made significant gains in all 10 tested competencies (mean score increase 1.11-1.47; p&lt;0.0001) and overall score (n=10, 12; df=34; p=0.0001). 41.6% participants completed the inter-professional competency survey. Post-intervention scores were significantly higher than pre-intervention scores (p&lt;0.009).</td>
</tr>
<tr>
<td>Bergeron et al.22 2008</td>
<td>Radiation Hazards Assessment Challenge</td>
<td>VE incorporating a simulated Geger counter to assess a variety of patients exposed to radiation.</td>
<td>RCT: SG vs. C (didactic lecture)</td>
<td>Paramedics in USA; n=89.</td>
<td>Pre/post MCQ.</td>
<td>5</td>
<td>Significantly higher post-test scores for SG compared to C group (p&lt;0.01).</td>
</tr>
<tr>
<td>Bergeron et al.22 2008</td>
<td>Nuclear Event Triage Challenge</td>
<td>VE cases with augmented media and interactive testing of triage decision-making.</td>
<td>RCT: SG vs. C (didactic lecture)</td>
<td>Paramedics in USA; n=89.</td>
<td>Pre/post MCQ.</td>
<td>5</td>
<td>Significantly higher post-test scores for SG compared to C group (p&lt;0.01).</td>
</tr>
<tr>
<td>Bindoff et al.12 2014</td>
<td>Computer Simulation of Community Pharmacy Practice</td>
<td>3D VE with VP cases teaching principles of community pharmacy.</td>
<td>RCT: SG vs. C (reading)</td>
<td>Pharmacy students in Australia; n=33.</td>
<td>Pre/post MCQ, self-assessment, survey.</td>
<td>1, 2, 3, 5</td>
<td>The difference between SG vs. C post MCQ scores was approaching significance (p=0.059). SG achieved a mean change in pre/post score of 0.6 (standard error (sd)=-0.3), while C had a mean change of -0.2 (sd=-0.3). In the self-assessment, there were no significant differences. Comments included difficulty using the SG, overly simplistic dialog options, and the value of being able to perform the whole dispensing process.</td>
</tr>
<tr>
<td>Boada et al.10 2015</td>
<td>LISSA</td>
<td>3D VE where learner performs CPR to save victims.</td>
<td>RCT: SG vs. C (self-directed learning, live simulation)</td>
<td>Nursing students in Spain; n=109.</td>
<td>Post-intervention live simulation testing.</td>
<td>1, 2, 3, 5</td>
<td>SG groups 2 and 3 had significantly better learning acquisition scores than C group (1). Paired samples t-test between group 1 and 2 (1 = 35, 67, 2 = 47, 50 and p &lt; 0.05) and between group 1 and 3 (1 = 35, 67, 3 = 50, 58 and p &lt; 0.05) showed significant differences in both cases.</td>
</tr>
<tr>
<td>Boeker et al.17 2013</td>
<td>Uro-Island</td>
<td>VE adventure game with interactive tasks related to urine pathology and urinalysis.</td>
<td>RCT: SG vs. C (self-study)</td>
<td>Medical students in Germany; n=145.</td>
<td>Post-intervention MCQ.</td>
<td>5, 6</td>
<td>Higher post-test scores in SG group compared to C group [mean 28.63/34.0 vs. 28.03/34.0, Cohen’s d=0.71 (ITT analysis)].</td>
</tr>
<tr>
<td>Buttussi et al.18 2013</td>
<td>EMSAVE</td>
<td>VE where learners interact with, examine, and treat patients</td>
<td>Quasi-experimental</td>
<td>ALS providers in Italy; n=40.</td>
<td>Pre/post MCQ, re-testing at 3 months, survey.</td>
<td>1, 2, 3, 5</td>
<td>Significantly higher MCQ scores after EMSAVE training (+4.8 points; 95% CI +3.4, +6.2, p&lt;0.001). Significantly higher MCQ scores at 3 months vs. pre-test (+2.9 points, 95% CI &lt;0.01).</td>
</tr>
</tbody>
</table>
Diehl et al. 2013  
Interactive Storytelling  
VP cases to teach physician-patient interaction, communication and empathy.  
Descriptive  
Medical students in USA.  
None. None.  
Development of SG described.

Chan et al. 2012  
Ultrasound-guided needle placement  
Technical trainer with virtual tools and tissues to teach ultrasound-guided needle placement.  
RCT: SG vs. C (self-study)  
Radiology residents in China; n=21.  
Post testing via live simulation.  
4, 5  
Increased success rate for SG vs. C [mean 33.3% to 68.2% (p<0.001)] vs. 20.2% to 37.8% (+17.6%) Reduced completion time for SG vs. C [51.5-34.8 s (16.9 s) vs. 54.3-42.8 s (-11.5 s)]. Validation with only inexperienced users’ post-test scores improved vs. pre-test.

Off-pump Coronary Artery Bypass  
Multi-player VE, teaching the steps of off-pump coronary artery bypass.  
Descriptive  
Cardiac surgeons in Canada.  
None. None.  
Development of SG described.

Cowan et al. 2010  
Total Knee Arthroscopy (TKA) Game  
VE teaching the steps of a TKA, with MCQ sub-games interspersed.  
Survey  
Medical students in Canada.  
Survey.  
1, 2, 3  
Participants rated the SG easy to use, intuitive, stimulating.

Craig et al. 2014  
Virtual Pathology Instructor (V-PIN)  
Interactive CBS with MCQ: branched narratives with virtual pathology materials that guide learners through diagnosis.  
Quasi-experimental (SG at various points in curriculum).  
Medical students in USA; n=185.  
Post-intervention MCQ, SG performance, survey.  
1, 2, 3, 5  
Significant pre-test to post-test improvement in scores for two cases (mean 3.8 vs. 4.2, p=0.0007; 1.9 vs. 3.0, p=0.0001), but not the third (mean 2.5 vs. 2.3, p=0.12). Maximum score for traditional teaching with SG vs. SG alone (4.2 vs. 3.0, p=0.0001). No significant difference in time on task, total steps, score, and requests for help when SG placed before vs. after testing, 66% felt that “V-PIN cases helped me better understand the concepts.”

Multiplayer Virtual World-CPR Game  
Multiplayer VE with 4 short patient-based scenarios where learners care for cardiac arrest victims in teams of three, communicating via microphone. Learners approach and examine victim, and perform resuscitation in communication with 911 dispatcher.  
Four studies: Quasi-experimental.  
RCT: CPR training followed by SG refresher 6 vs. 18mo prior to assessment vs. control (no refresher).  
Medical students in Sweden; n=12, 13, 15, 56 n=30.  
Post testing via live simulation, time and knowledge gain testing within SG.  
1, 76, 85, 85  
Students reported enjoyment and learning with SG. Self-efficacy increased after first [5.9 to 6.57 (p=0.01)] and second SG session [8.0 to 6.77 (p=0.03)]. Mean perceived concentration increased [54.2/100 to 66.6/100 (p=0.006)], with low to moderate mental strain (mean=2.6/10). Significant increase in self-reported situational awareness (SA) [3.4 (2.4-1) to 3.9 (3.1-4.5), p<0.001], and correlation between SA and concentration (p=0.61) (Spearman rank order correlation, p<0.001). Decreased mean elapsed time between first and last scenarios until start of examination (42.8 to 16.2s), ventilation (66.8 to 44s), and chest compressions (68 to 49.8 s). No decrease in mean time to call for help. Mean compression fraction increased (42% to 52%) and mean occurrence of protocol violations decreased (5.5 to 1.8). Higher scores in 6mo vs. control [93 (±11)% to 65 (±28)% (p<0.05), 18mo scores: 73 (±23%), Mean guideline violations 0.2 (±0.5), 1.5 (±1.0), 4.1 (±1.0) for 6mo, 18mo and control, respectively. Incorrectly delivered chest compressions 54 (±44%). 44 (±48%), 0% for control, 18mo, and 6mo (p<0.001 control) or 18mo vs. 6mo. 32.

Davidis et al. 2011  
Electrolyte Workshop  
VP cases simulating electrolyte abnormalities, augmented with multimedia. Physiology modeled.  
Survey  
Residents, physicians in South Africa; n=16.  
Usability survey.  
1, 2, 3  
Mean SUS 78.4 +/- 13.8. 16/16 would use the SG often, 15/16 would recommend to others. 13/16 found it easy to use and well-integrated, while 2/16 found it cumbersome and 3/16 found navigation difficult. 15/16 found the content scientifically sound and interesting. 9/16 found it realistic: 14/16 reported concept clarity and increased understanding.

De Leo et al. 2014  
Game Medical Team Training (GaMaTT)  
Multi-player VE teaching interprofessional acute care (microphone communication). Learners manage virtual military and civilian casualties. VP operated remotely by live person. Physiology mapped.  
Survey  
Air Force National Guard Medical Service in USA; n=29.  
Survey.  
1, 2, 3  
Comparing user characteristics via Independent Television Commission Sense of Presence Inventory using Kendall T test showed: characteristics “how often you play video games” (T[26]=0.458, p>0.01) and “television/film production knowledge” (T[27]=0.516, p<0.01) were significantly related to negative effects (dizziness, nausea, headache, and eyestrain during SG experience). The user characteristic “knowledge of virtual reality” was significantly related to engagement (T[26]=0.463, p<0.01) and negative effects (T[26]=0.404, p<0.05).

Delassobra et al. 2010  
Cardiac Arrest!  
VE where users manage patients requiring advanced cardiac life support skills.  
RCT: reading (R) vs. live simulation (S) vs. SG/multimedia group (M)  
Paramedic students in India; n=117.  
Pre/post MCQ, cardiac arrest live simulation, re-testing at 3 weeks, survey.  
1, 2, 3, 5  
Significantly higher increases in live simulation post-test scores for S vs. R and M groups compared to pre-test (+9% S vs. +5% R vs. +2% M, p=0.05). Significantly higher MCQ post-test scores for M vs. R, no difference compared to S (+5% M, +3% S, +2% R, p=0.05 for R only). Significantly higher increases in live simulation post-test scores at 3 weeks for M vs. R, no difference compared to S (+6% S, +1% M, -1% R, p=0.05 for R only). Significantly higher increases in MCQ post-test scores at 3 weeks for M vs. R and vs. S (+5% vs. 0% vs. 0%, p=0.05). 95% of S and 84% of M felt their modality should be incorporated into the curriculum. 95% of S and 85% of M reported enjoyment.

Dev et al. 2011  
CliniSpace  
VE of emergency department. Physiology modeled.  
Quasi-experimental  
Physicians, students.  
Survey.  
1, 2, 3  
Although numbers not reported, users rated the game as easy to use, realistic, and stimulating.

Diehl et al. 2013, 2015  
InsuOnline  
VE with diabetic VP. Embedded quizzing and multimedia assets.  
Comparison trial: Lecture (L) vs. SG.  
Medical students and residents in USA; n=20, n=41;  
Pre/post MCQ, re-testing at 3 months.  
1, 2, 3, 5  
Development of SG and future studies described.  
Mean SUS score of 86/100 for prototype, 92.5 for beta version. Users found the SG to be fun, engaging, challenging, relevant, realistic, and preferable to a lecture. Users said the SG increased knowledge of and confidence with diabetes and insulin, that it would impact how they treated patients with diabetes. Mean knowledge/skills score improved from 68% to 89% in L group (n=23; p<0.001), from 61% to 90% in SG group (n=18; p<0.001). At 3 months, mean score decreased (80% in L group, 78% in SG group; p=0.001 for both), but was significantly higher than baseline (p=0.001 for both). No difference between SG and L groups immediately or 3 months post. Score increment was better for SG (29%) than L (21%; p=0.04). Benefits improved in SG group only.
Dugue et al. 2008 The Virtual Home Visit VE of a patient’s home, teaching about geriatric home visits. Timed, with distractors. Learners identify risk factors for falls/harm. Quasi-experimental Medical students in Australia; n=56. Pre/post MCQ, survey. 1, 2, 3, 5 Significantly improved post-test (90 ±5%) vs. pre-test scores (42 ±5%) (p<0.001). 77% of participants played twice, obtaining a change in scores from 250 ±20 points to 400 ±50 points (p<0.001). 100% of participants described home visit as important, but 58% felt they had not had enough training. 78% of participants would recommend the SG, 77% felt it improved knowledge and confidence, 85% considered the experience good or excellent.

Evans et al. 2014 Sepris CBS about sepsis and shock. Used triage test and treat several patients at once, timed. Physiology modeled. Quasi-experimental Medical students; residents in USA; n=156. Pre/post MCQ, satisfaction and survey. 1, 2, 3, 5 Significantly improved post-test score vs. pre-test (5.5, SD = 2.31, to 6.94, SD = 2.88, p<0.001). Self-reported confidence managing sepsis improved (p<0.001) following SG. Over 85% of subjects reported they would or maybe would recommend the SG.

Fonseca et al. 2015 e-Baby VE of an incubator, teaching clinical evaluation of preterm infants’ respiratory process. Embedded multimedia assets to share within a Moodle course. Survey Nurses, nursing students in Australia; n=141. Satisfaction survey. 1, 2, 3, 4 Development of SG described. 100% found the tool easy to use, enjoyed the didactic component and felt motivated to use it, felt the tool enabled learning preterm oxygenation needs, requested such technologies to be applied to other topics and felt that it helped in their learning. 83% enjoyed learner autonomy, 86% felt feedback was immediate, 7% felt technologies like this could replace teachers.


Graafland et al. 2014 Medialis VE of mixed CBS with multimedia. Learners diagnose and manage patients with biliary tract disease. Validation study: performance vs. skill level Surgeons, trainees in Netherlands; n=41. Performance during SG: Number of solved cases. 4 Surgeons solved more cases correctly (mean 77%) vs. residents (67%), interns (60%), master-degree students (50%), and bachelor-degree students (39% (p<0.01). Trainees performed significantly better in second session (median 72 vs. 48%, p=0.00).

Graafland et al. 2013 Situational Awareness in Surgical Training Game CBS with text questions augmented with multimedia, teaching management of equipment-related errors. Quasi-experimental Surgeons, trainees, n=45. Usability and satisfaction survey. 1, 2, 3, 4 Majority found SG realistic (64.4%-88.9%), useful (53%), positive (78%) and challenging (60%). 66% would play the SG in their leisure time. Surgeons more likely to perceive the SG as boring than residents and students (23.5% vs. 6.7% and 8.3%; p=0.045).

Graafland et al. 2014 Laparoscopic Equipment Failure Serious Game VE simulating a minimally invasive surgery (MIS) unit, testing learners’ assessment and management of equipment settings and displays. Validation study: comparison of performance vs. skill level Surgeons, trainees, MIS equipment specialists in Netherlands; n=45. Performance during SG: Number of solved cases. 4 Equipment specialists (ES) solved significantly more equipment problems than medical students (MS), residents (R), and surgeons (S) (68.9 vs. 51.0%, 51.4%, and 45.0%, respectively, p=0.01). ES required a median of 1.00 problem-solving steps (IQR 1.00–4.00), MS 2.00 (IQR 1.00–4.00), Kruskal-Wallis, p=0.03. ES had higher proportion of correct steps (median of 1.00 (IQR 0.50–1.00) vs. S 0.50 (IQR 0.00–1.00) vs. R 0.50 (IQR 0.00–1.00) vs. MS 0.50 (IQR 0.00–1.00). Mann–Whitney U for ES vs. other groups: p=0.01 vs. S, p=0.05 vs. R, and p=0.02 vs. MS.


Hashimoto et al. 2015 LapMentor VR VE simulating laparoscopic cholecystectomies. RCT: competitive SG vs. control SG (C). Medical students in UK; n=18. Time, movements, instrument path length, global rating scale. 1, 2, 3, 5 Time and global rating scale score were not significantly different between groups. Competitive SG group was significantly more Dexterous than C and had significantly lower variance in number of movements and instrument path length (p<0.015).

Jalink et al. 2014 Underground world VE: learner demonstrates and reconstructs mine using 2 Wi Remote controllers in laparoscopic tool shells. Trans basic laparoscopic skills, not medical knowledge. Quasi-experimental Laparoscopic surgeons at a global conference; n=72. Speed within SG, satisfaction survey. 1, 2, 3, 4, 5 Experts were 111% faster (p<0.001) than novices. Also, scores of the FLS Peg Transfer test and the Wi Laparoscopy showed a significant, high correlation (r = 0.812, p<0.001). On a 1-to-10 scale, mean score for hardware realism was 7.2, mean score for usefulness as a training tool was 6.4.

Janssen et al. 2015 They Know: Anatomy Multi-player adventure competition, advancement with correctly answering questions about human anatomy. Survey Medical students in Australia; n=16. Satisfaction survey, open-ended interviews. 1, 2, 3 89% found the SG engaging. 93% described the SG as challenging, 74% would like to play the SG again if given the opportunity. Participants found the competitive aspects positive, and described subjective improvement in knowledge of anatomy, as well as knowledge of personal strengths and weaknesses.

Kaczmargczyk et al. 2015 Video-based VE teaching and testing acute management of tachycardias Survey Medical students in UK; n=47. Satisfaction survey, open-ended interviews. 1, 2, 3 98% agreed that the use of SGs would support the teaching of acute medicine. Participants enjoyed the interactivity and found the SG a useful aid for knowledge consolidation. Suggested areas for improvement were shortening video length, fixing technical glitches and increasing the detail of the end-game feedback.

Kanthan, Senger et al. 2011 The Path is Right Testing VE with multimedia for teaching clinical diagnosis. MCQ, fill-in the blank, and extended matching options. Quasi-experimental Medical students in Canada; n=77. Pre/post test: midterm vs. final exam scores. 5, 6 Students given access to game midterm and before final exam. No significant difference in midterm vs. final scores (midterm = 74.31%, range 53.06% - 85.50%, final = 75.52%, range 57.84% - 89.22%).

Katz, Central Venous VE for performing central venous Descriptive Residents, None. None. Development of SG and future studies described.
Demaria
et al. 2013

Catheter Game
catheter placement.
survey of patients in the USA.

Kerfoot et al.
2012

SE Urology
Content Game

Competitive MCQ game on core urology content. 2 MCQs with feedback emailed daily. MCQ re-
sent in 2 or 6 weeks if answered incorrectly and correctly, respectively. Competition fostered by posting relative performance.

Quasi-

1, 2, 3

Baseline scores (median 62%) improved by year (p<0.001, dmax = 2.08), school (p<0.001, dmax = 0.75), and gender (p=0.01, d = 0.38). Median completion score was 93% (95% CI 12) and varied significantly by year (p<0.001, dmax = 1.12), school (p<0.001, dmax = 0.34), and age (p=0.019, dmax = 0.43). Scores did not differ significantly between years 3 and 4. 70% of participants requested to participate in future SE games.

Kerfoot et al.
2012

SE Pre-Clinical
Content Game

Competitive MCQ game on pre-
clinical content, 2 MCQs with feedback emailed daily. MCQ re-
sent over time and relative scores posted to foster competition.

Quasi-

1, 2, 3

Median baseline score was 53% (43% 16) and varied significantly by year (p<0.001, dmax = 2.08), school (p<0.001, dmax = 0.75), and gender (p=0.01, d = 0.38). Median completion score was 93% (95% CI 12) and varied significantly by year (p<0.001, dmax = 1.12), school (p<0.001, dmax = 0.34), and age (p=0.019, dmax = 0.43). Scores did not differ significantly between years 3 and 4. 70% of participants requested to participate in future SE games.

Kerfoot et al.
2012

SE Urology
Continuing Medical
Education Guidelines
Game

Competitive MCQ game on urology guidelines. 2 MCQ with feedback emailed daily. MCQ re-
sent over time and relative scores posted to foster competition.

RCT: 2 MCQs every 2 days, vs. 4 MCQs every 4 days

Physicians worldwide; n=1470.

Baseline scores and completion rates within SE.

1, 2, 3

Median baseline score was 48% (95% CI 17), and median completion score was 98% (95% CI 15) for SE group. 4 MCQ every 4 day group performed worse than 2 MCQ every 2 days (d = 0.43, p<0.001). 76% of participants requested to participate in future SE games.

Kizakiewicz et al.
2007

Sim-Patient

VE with single and multiple-
casualty cases (bioterrorism, trauma, improvised explosive device). Learners apply triage

tags and manage cases. Physiology modeled.

Survey

Military physicians in the USA; n=31.

Survey.

1, 2, 3

Surveyed participants rated simulation realism and navigation a mean of 4.40 (out of 5, SD 0.43), simulation content and responsiveness 4.42 (SD 0.04), simulation learning content 4.41 (SD 0.20).

Knight et al.
2010

Triage Trainer

VE of major incident scenario where users perform medical checks, tag and assess multiple patients.

RCT: SG vs. C (card sort exercise)

Physicians, nurses, paramedics in the UK; n=91.

Post testing with live simulation.

5

Significantly improved tagging and step accuracy for SG vs. C group (Chi2 = 13.12, p=0.02; Chi2 = 5.45, p=0.0196). No difference in time to triage between the two groups (456 ± 62 vs. 435 ± 74 s, p=0.155).

Kupnov et al.
2009

Burn Center

Mass burn casualty disaster VE. Learners stabilize, sort, tag and transport victims, care for patients in intensive care unit.

Descriptive

Trauma surgeons, nurses, therapists in the USA.

None.

None.

Positive correlation between training with Burn Center and performance in a traditional lectured course.

Lagro et al.
2014

GeriatrX

Geriatric clinic VE. Users care for VIP with fatigue due to iron-
deficiency anemia. Users scored on appropriateness and cost-
efficiency of care.

RCT: SG vs. C (self-study)

Medical students in the Netherlands; n=134.

Pre/post MCQ, confidence survey.

1, 2, 3

Significantly higher post-test scores for SG vs. C in cost consciousness (cost deviation: 13.5 +/- 3.6 vs. 15.4 +/- 3.6, p<0.05). Significantly higher self-perceived knowledge in SG vs. C group for appropriateness/ cost of care (efffect sizes 1.0, and 1.2). No difference for geriatric content. Most felt SG was clear (3.9/5±0.8), enjoyed it (4.0/5±0.8), wanted to finish the SG (4.0/5±0.8), considered it a safe environment (4.1/5±0.8), and felt it increased knowledge (3.9/5±0.8). Some users were not satisfied with clarity (2.3/5±1.0) and feedback (2.6/5±1.0).

LeBoy et al.
2008

Virtual ED II

Multiplexer VE with cases simulating SARin exposure or radioactive bomb blast. Communication through microphone. Physiology modeled.

Survey

Physicians, nurses in the USA; n=13.

Satisfaction survey.

1, 2, 3

62% reported a change in their feelings about working in an ED team. Mean ratings showed users felt immersed (3.47/5) and the SG increased confidence in ability to respond to trauma (2.00/5 before SG; 3.085/5 after SG). Most thought the SG would be useful for learning teamwork (mean = 3.7/5) and clinical skills (mean = 3.15/5). Comments indicate users perceived the SG as realistic.

LeBoy et al.
2008

Pennsylvania City

Multiplexer VE of a triage area after an explosion. Communication through microphone.

Survey

Paramedics, physicians, nurses in the USA; n=16.

Satisfaction survey.

1, 2, 3

75% and 56% of participants thought the SG was useful for refresher training and initial training, respectively. 62% thought the SG was as/more effective vs. traditional methods.

Lim et al.
2008

eNLG (online Neurological Localisation Game)

Modified essay and MCQ of patients with neurological problems, with learners examining and managing patients. Embedded videos.

Survey

Medical students in Singapore; n=76.

Satisfaction survey.

1, 2, 3

93.4% of participants felt that the eNLG helped them better understand neurological localization principles, 90.7% believed the question-based format was suitable, 96.1% liked the use of videotaped vignettes, and 98.6% requested more eNLG scenarios in the future. 42.1% felt the eNLG could replace regular bedside teaching.
<table>
<thead>
<tr>
<th>Study</th>
<th>Title</th>
<th>Year</th>
<th>Methodology</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qin et al.</td>
<td>Stopping the Fountains pre-game + Virtual Orthopedic-Surgery Game</td>
<td>2010</td>
<td>Operating room VE, learners manipulate surgical tools via haptic device. VP present with varying degrees of blood loss, users assess and manage. Physiology modeled.</td>
<td>SG with pre-training group outperformed SG without pre-training group in both completion time (p=0.01 by Wilcoxon-Mann-Whitney test), and p=0.006 by Kruskal-Wallis test). Performance score (p=0.001 by Wilcoxon-Mann-Whitney test and p=0.001 by Kruskal-Wallis test).</td>
</tr>
<tr>
<td>Rondon et al.</td>
<td>Anatesse 2.0 Quiz</td>
<td>2013</td>
<td>Interactive MCQ with multimedia, teaching anatomy and physiology of the speech, language, hearing, and swallowing mechanisms.</td>
<td>Pre-test scores were significantly lower vs. post-test and long-term post-test scores. No significant differences between the groups’ post-test scores (p=0.176). When isolating specific topics within the assessment, SG group had significantly higher Anatomy (p=0.001) post-test scores. Pre- vs. long-term post-test scores indicated significant differences only for control group (p=0.042). No significant difference in post-test vs. long-term post-test scores.</td>
</tr>
<tr>
<td>Mohamed et al.</td>
<td>ED Physician Decision Making</td>
<td>2014</td>
<td>VE, users evaluate and manage cases simulating a busy ED shift. One version contained distracting cases and other audiovisual distractors. Timed, physiology modeled.</td>
<td>VE: SG with distractors (cognitive load, CL) vs. SG without distractors (control, C). Performance during game: transfer to trauma center, number of orders entered, time spent, CT scan rates.</td>
</tr>
<tr>
<td>Meule et al.</td>
<td>Virtual Emergency Telemedicine (VETM) Game</td>
<td>2015</td>
<td>VE simulating cardiac emergencies through a telemedicine system.</td>
<td>The majority of respondents reported an increase in knowledge and suggested that they would recommend the resource to others.</td>
</tr>
<tr>
<td>Nicolaidou et al.</td>
<td>GRAPHC (Games Research Applied to Public Health with Innovative collaboration)</td>
<td>2012</td>
<td>CBS, learners work collaboratively and individually to explore population oral health and the evidence for community initiatives, care for simulated 5-year-old patient in an inner city environment, Moodle.</td>
<td>Satisfaction survey, SG completion.</td>
</tr>
<tr>
<td>Vankipuram et al.</td>
<td>ACLS Virtual Reality Simulator</td>
<td>2014, 2013</td>
<td>Multiplayer hospital VE, learners communicate via microphone with members of a team (randomly assigned to role) and perform ACLS on VP. CPR simulated via joystick. Two games with differences in amount of feedback (persuasion) given.</td>
<td>No difference in post-intervention performance control vs. MP (p=0.37 for pulseless electric activity (PEA) and p=0.10 for ventricular fibrillation &amp; ventricular tachycardia (VFib/VTach)). No difference in performance in P vs. MP (p=0.1 for PEA; p=0.63 for VFib/VTach). Significant difference in performance in control vs. MP (p=0.05 for PEA and p=0.02 for VFib/VTach). The post-test comparison of performances of the groups showed that control (p=0.017 for PEA, p=0.01 for VFib/VTach) and P (p=0.02 for PEA, p=0.048 for VFib/VTach) groups improved their performances significantly, whereas MP group did not (p=0.45 for PEA, p=0.46 for VFib/VTach).</td>
</tr>
<tr>
<td>Qin et al.</td>
<td>SICKO (Surgical Improvement of Clinical Knowledge Ops)</td>
<td>2015</td>
<td>Same game format as Septin (above), but with timed surgical CBS. Learner triages and prioritizes care, decides whether or not to operate, makes intraoperative decisions.</td>
<td>Mean total SG scores for the novice, junior resident, senior resident, and expert groups were 5,461; 5,819; 11,404; and 13,913, respectively (p = 0.001). Usability survey results were positive, with mean scores ranging from 3.52 to 4.28 across ten questions.</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Description</td>
<td>Design/Methodology</td>
<td>Survey</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Roy et al. 2006</td>
<td></td>
<td>Virtual Standardized Patients</td>
<td>VE where learners examine and care for VP: Augmented multimedia. Communication via microphone.</td>
<td></td>
</tr>
<tr>
<td>Schwab et al. 2011</td>
<td></td>
<td>Second Life (SL) Case-Based Mock Emergency Medicine Board Examination</td>
<td>Hospital and examination room VE, learners manage cases, communicate via microphone with a faculty examiner at a remote computer (virtual patient). Mock oral American Board of Emergency Medicine examination.</td>
<td>Survey</td>
</tr>
<tr>
<td>Schwaz et al. 2013</td>
<td></td>
<td>Medical Facilities Network (MEFANET)</td>
<td>Timed CBS in basic life support, emergency, anesthesiology, pain, and critical care, with embedded multimedia. Physiology modeled.</td>
<td>Survey</td>
</tr>
<tr>
<td>Schwid et al. 2001</td>
<td></td>
<td>Anesthesia Simulator 3.0</td>
<td>VE, learners examine and treat patients during anesthetize emergencies. Physiology modeled.</td>
<td>RCT: SG vs. C (self-study)</td>
</tr>
<tr>
<td>Stredney et al. 1996</td>
<td></td>
<td>Virtual Simulated Epidural Anesthesia</td>
<td>Hospital VE of a simulated patient’s back, for practicing epidural placement, with computer-generated needle.</td>
<td>Survey</td>
</tr>
<tr>
<td>Sward et al. 2008</td>
<td></td>
<td>Pediatric Rotation Game</td>
<td>Testing environment of a web-based, interactive adaptation of an existing board game for general pediatric content.</td>
<td>RCT: SG vs. C (self-study)</td>
</tr>
<tr>
<td>Tubik et al. 2015</td>
<td></td>
<td>Zinc Phosphate Cement Virtual Learning Object</td>
<td>Moodle-based VE teaching and testing handling of zinc phosphate cement for dental prostheses.</td>
<td>RCT: Longitudinal SG (15 days) vs. SG (20 minutes) vs. longitudinal reading vs. reading (C)</td>
</tr>
<tr>
<td>Youngblood et al. 2008</td>
<td></td>
<td>Virtual ED</td>
<td>Multiplayer emergency room VE, users examine and treat patients with a team. Users communicate through microphone. Physiology modeled.</td>
<td>RCT: SG vs. C (live simulation)</td>
</tr>
</tbody>
</table>

Abbreviations: SG = serious game, C= control, VE = virtual environment, VP = virtual patients, CBS = case-based simulation, RCT= randomized controlled trial, MCQ= multiple choice questions, SUS= System Usability Scale, s= seconds.