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 Game 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. 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=20</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. 2015</td>
<td>Decision Simulation</td>
<td>Using ve with looped, case-based, branching-learning 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 57.7±12.6, 66.5±13.6, p=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=0.001). More C group testers had a score increase (83.3% vs 69.5%, respectively, p=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=0.01).</td>
</tr>
<tr>
<td>Amer et al. 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=0.07), 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. 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 triages; 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 (±0.17); d=2.07, 11.38/81%. C (2.9)/ versus 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=0.048).</td>
</tr>
<tr>
<td>Antoniou et al. 2014</td>
<td>Second Life (SL) Periodontology Virtual Patients</td>
<td>VE with embedded MCQ about clinical/surgical, diagnostic, and treatment of periodontitis cases. Two versions of SG: Web vs. SL.</td>
<td>Survey: Web- vs. SL-based version 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>Atack et al. 2009</td>
<td>The Disaster Management Course</td>
<td>Multi-player VE of mass casualty exercise teaching disaster management and interprofessional skills, integrated within online course. Embedded interactive components.</td>
<td>Quasi-experimental</td>
<td>Several interprofessional 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 (30%) and lack of time (40%) as reasons. 77% completed pre- and post-competency survey (pre mean 2.72 (SG 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% participants completed the inter-professional competency survey. Post-intervention scores were significantly higher than pre-intervention scores (p=0.009).</td>
</tr>
<tr>
<td>Bergeron et al. 2008</td>
<td>Radiation Hazards Assessment Challenge</td>
<td>VE incorporating a simulated Giger 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. 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. 2014</td>
<td>Computer Simulation of Community Pharmacy Practice</td>
<td>3D VE with VE 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: 0.03), while C had a mean change of -0.2 (standard error: 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>Boasta et al. 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 (t = 35.67, 2 = 47.50 and p &lt; 0.05) and between group 1 and 3 (t = 35.67, 3 = 50, 58 and p &lt; 0.05) showed significant differences in both cases.</td>
</tr>
<tr>
<td>Boeker et al. 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 26.8/34.0 vs. 26.0/34.0, Cohen’s d=0.71 (ITF analysis)).</td>
</tr>
<tr>
<td>Buttussi et al. 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...</td>
</tr>
</tbody>
</table>
Cavazza et al. 2010
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% (+34.9%); 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.

Coyne et al. 2011
Off-pump Coronary 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 South Africa; n=117.
Post-intervention MCQ, SG performance, survey.
1, 2, 3
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 teaching, 66% felt that “VP-PIN cases helped me better understand the concepts.”

Craige 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 teaching, 66% felt that “VP-PIN cases helped me better understand the concepts.”

Multiples Player Virtual World CPR Game
Multiples Player 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.

Quart studies: Quasi-experimental. (T(26)=
RCT: CPR training followed by SG refresher 6 vs. 18mo prior to assessment vs. control (no refresher)).
Medical students in Sweden; n=12, 15, 56
n=30, 34.
Post testing via live simulation, time and knowledge gain testing within SG. *p=0.05
Survey.
1, 1.05, 0.97
1.05, 0.97
Survey.
70, 85, 85
3, 70, 85, 85
8, 76
8, 76
5
5
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).11 Significant increase in self-reported situational awareness (SA) [3.4 (2.4-1.3) to 3.9 (1.4-5), p<0.001], and correlation between SA and concentration (p=0.61) (Spearman rank order correlation, p<0.001).36 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 (6.5 to 1.8).36 Higher scores in 6mo vs. control [93% (±11%)] to 65% (±28%) (p=0.05), 18mo scores: 73% (±23%), 15. Mean guideline violations 0.2 (±0.5), 1.5 (±1.0), 4.1 (±0.4) 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).

David 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. 16/18 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=0.64) 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).

Delasserra 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)
Paramedics 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 increases in 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 3 weeks for M vs. R, no difference compared to S (+4% S, +1% M, -1% R, p=0.05 for R only). Significantly higher increases in MCQ post-test scores 3 weeks for M vs. R and vs. S (+5% vs. 0% vs. p=0.05). 95% of S and 84% of M felt their modality should be incorporated into the curriculum. 95% of S and 89% 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.

Dietl 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.20 Mean SUS score of 88/100 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 88% to 89% in L group; (p=0.03), 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.

requiring advanced life support skills.

+1.5, +4.2, p<0.001). 39/40 participants regarded EMSAVE as a valuable ALS refresher tool, 85% of participants were willing to devote 1 hour/month to retrain with the game.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Title</th>
<th>Methodology</th>
<th>Subject</th>
<th>Training</th>
<th>Test</th>
<th>Scores</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duque et al.</td>
<td>2008</td>
<td>The Virtual Home Visit</td>
<td>Quasi-experimental</td>
<td>Medical students in Australia; n=56</td>
<td>Pre/post MCQ, survey.</td>
<td>1, 2, 3, 5</td>
<td>Significantly improved post-test (90 ±5%) vs. pre-test scores (42 ±5%) (p&lt;0.001). 77% of participants played twice, obtaining a change in scores from 250 ±20 points to 400 ±50 points (p&lt;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.</td>
<td></td>
</tr>
<tr>
<td>Evans et al.</td>
<td>2014</td>
<td>Septis</td>
<td>Quasi-experimental</td>
<td>Medical students; residents in USA; n=156</td>
<td>Pre/post MCQ, satisfaction and survey.</td>
<td>1, 2, 3, 5</td>
<td>Significantly improved post-test vs. pre-test (5.85, SD = 2.31, to 6.94, SD = 2.88, p&lt;0.001). Self-reported confidence managing sepsis improved (p&lt;0.001) following SG. Over 85% of subjects reported they would or maybe would recommend the SG.</td>
<td></td>
</tr>
<tr>
<td>Forseca et al.</td>
<td>2014</td>
<td>e-Baby</td>
<td>Survey</td>
<td>Nurses, nursing students in Australia; n=14</td>
<td>Satisfaction survey.</td>
<td>1, 2, 3, 4</td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>Foster et al.</td>
<td>2009</td>
<td>3D Renal Care Learning Environment</td>
<td>Descriptive</td>
<td>Nurses in Australia.</td>
<td>None</td>
<td>None</td>
<td>Development of SG described.</td>
<td></td>
</tr>
<tr>
<td>Gleason et al.</td>
<td>2015</td>
<td>Research &amp; Evidence Learning in Medicine</td>
<td>Descriptive</td>
<td>Medical students in USA.</td>
<td>None</td>
<td>None</td>
<td>Development of SG described.</td>
<td></td>
</tr>
<tr>
<td>Graafland et al.</td>
<td>2014</td>
<td>Medialis</td>
<td>Validation study</td>
<td>Surgeons, trainees in Netherlands; n=41.</td>
<td>Performance during SG: No user of solved cases.</td>
<td>4</td>
<td>Surgeons solved more cases correctly (mean 77%) vs. residents (67%), internes (60%), master-degree students (50%), and bachelor-degree students (39% (p&lt;0.01). Trainees performed significantly better in second session (median 72 vs. 48%, p&lt;0.00).</td>
<td></td>
</tr>
<tr>
<td>Graafland et al.</td>
<td>2013</td>
<td>Situational Awareness in Surgical training Game</td>
<td>Quasi-experimental</td>
<td>Surgeons, trainees, n=45.</td>
<td>Usability and satisfaction survey.</td>
<td>1, 2, 3, 4</td>
<td>Majority found SG realistic (64.4%–88.9%), useful (53%), positive (78%) and challenging (65%). 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).</td>
<td></td>
</tr>
<tr>
<td>Graafland et al.</td>
<td>2014</td>
<td>Laparoscopic Equipment Failure Serious Game</td>
<td>Validation study</td>
<td>Surgeons, trainees, MIS equipment specialists, n=45.</td>
<td>Performance during SG: Number of solved cases.</td>
<td>4</td>
<td>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 1.00–3.00), S 2.00 (IQR 1.00–4.00), R 2.00 (IQR 1.00–4.00), Kruskal–Wallis χ²=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.</td>
<td></td>
</tr>
<tr>
<td>Hashimoto et al.</td>
<td>2015</td>
<td>LapMentor VR</td>
<td>Quasi-experimental</td>
<td>Medical students in UK; n=18.</td>
<td>Time, movements, instrument path length, global rating scale.</td>
<td>1, 2, 3, 5</td>
<td>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=0.019).</td>
<td></td>
</tr>
<tr>
<td>Jatink et al.</td>
<td>2014</td>
<td>Underground world</td>
<td>Quasi-experimental</td>
<td>Laparoscopic surgeons at a global conference; n=72.</td>
<td>Speed within SG, satisfaction survey.</td>
<td>1, 2, 3, 4</td>
<td>Experts were 111% faster (p&lt;0.001) than novices. Also, scores of the FLS Peg Transfer test and the Wii Laparoscope showed a significant, high correlation (r = 0.812, p&lt;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.</td>
<td></td>
</tr>
<tr>
<td>Janssen et al.</td>
<td>2015</td>
<td>They Know: Anatomy</td>
<td>Survey</td>
<td>Medical students in Australia; n=16.</td>
<td>Satisfaction survey, open-ended interviews.</td>
<td>1, 2, 3</td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>Kaczmierzczak et al.</td>
<td>2015</td>
<td>Video-based VE teaching and testing acute management of tachyarrhythmias</td>
<td>Survey</td>
<td>Medical students in UK; n=47.</td>
<td>Satisfaction survey, open-ended interviews.</td>
<td>1, 2, 3</td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>Kanthan, Senger et al.</td>
<td>2011</td>
<td>The Path is Right</td>
<td>Quasi-experimental</td>
<td>Medical students in Canada; n=77.</td>
<td>Pre/post test: midterm vs. final exam scores.</td>
<td>5, 6</td>
<td>Students given access to game midterm and before final exam. No significant difference in midterm vs. final scores (midterm = 74.31%, range 53.06% - 88.50%, final = 75.52%, range 57.84% - 89.22%).</td>
<td></td>
</tr>
<tr>
<td>Katz,</td>
<td></td>
<td>Central Venous</td>
<td>Descriptive</td>
<td>Residents, None.</td>
<td>None</td>
<td>None</td>
<td>Development of SG and future studies described.</td>
<td></td>
</tr>
</tbody>
</table>
Demarsa et al. 2013 Catheter Game catheter placement. students in USA.

Kerfoot et al.2012 SE Urology Content Game Competitive MCQ game on core urology content, 2 MCQs with feedback emailed daily. MCQs re- sent in 2 or 6 weeks if answered incorrectly and correctly, respectively. Competition fostered by posting relative performance.

Kerfoot et al.2012 SE Pre-Clinical Content Game Competitive MCQ game on pre-clinical content, 2 MCQ with feedback emailed daily. MCQ re- sent over time and relative scores posted to foster competition.

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.

Kizakevich et al.2007 Sim-PatientVE VE with single and multiple-casualty cases (bioterrorism, trauma, improvised explosive device). Learners apply triage tags and manage cases. Physiology modeled.

Knight et al.2010 Triage Trainer VE of major incident scenario where users perform medical checks, tag and assess multiple patients.

Kurenov et al.2009 Burn Center Mass bomb blast casualty disaster VE. Learners stabilize, sort, tag and transport victims, care for patients in intensive care unit.


LeBoy et al.2008 Virtual ED II Multiplayer VE with cases simulating SARAH exposure or radioactive bomb blast. Communication through microphone. Physiology modeled.

LeBoy et al.2008 Peninsula City Multiplayer VE of a triage area after an explosion. Communication through microphone.

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.

Mean baseline score was 48% (IQ 17), and median completion score was 98% (IQ 25).

Baseline scores (median 62%, interquartile range [IQR] 17%) correlated with scores on 2008 American Urological Association in-service examination (ISE08, ISE09, 2009 American Board of Urology qualifying examination (ISE209) [r = 0.76, 0.46, and 0.64, respectively; all p < 0.01]. Baseline scores varied by gender, country, degree, and training year (all p < 0.05). Completion scores (median 100%, IQR 2%) correlated with ISE08 and ISE09 scores (<0.35, p < 0.001 for both). Median completion score was 100% (IQR 2%). Completion scores varied by training, ranging from a median 99% (IQR 4%) for year 1 to 100% (IQR 1%) for year 4 residents (p = 0.01). 72% surveyed requested future SE games.

Kerfoot, Baker et al 2012 RCT: SG vs. control, n=134. School (p<0.001, dmax = 0.28), school (p<0.001, dma = 0.75), gender (p<0.001, d = 0.38). Median completion score was 93% (IQ 12) and varied significantly by year (p<0.001, dma = 0.12), school (p<0.001, dma = 0.34), and age (p<0.019, dma = 0.43). Scores did not differ significantly between years 3 and 4. 70% surveyed requested future SE games.

Kizakevich et al.2007 Sim-PatientVE 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 USA; n=31. Survey.

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 group (Chi2 = 13.126, p<0.02; Chi2 = 5.45, p=0.019). No difference in time to triage between the two groups (456 ± 62 vs. 435 ± 74 s, p=0.155).

Kurenov et al.2009 Burn Center Mass bomb blast casualty disaster VE. Learners stabilize, sort, tag and transport victims, care for patients in intensive care unit.

Descriptive Trauma surgeons, nurses, therapists in USA. None. None. Positive correlation between training with Burn Center and performance in a traditional lectured course.


RCT: SG vs. C (self-study) Medical students in Netherlands; n=134. Pre/post MCQ, confidence survey. 1, 2, 3, 5, 4, 2.5, 3, 3, 4.6, 0.05. Significantly higher self-perceived knowledge in SG vs. C group for appropriateness/ cost of care (eff size 1.0 and 1.2). No difference for geriatric content. Most felt SG was clear (3.9 ± 0.4), enjoyed it (4.0 ± 0.4), wanted to finish the SG (4.3 ± 0.4), considered it a safe environment (4.1 ± 0.9), and felt it increased knowledge (3.8 ± 0.5). Some users were not satisfied with clarity (2.3 ± 1.0) and feedback (2.4 ± 1.0).

LeBoy et al.2008 Virtual ED II Multiplayer VE with cases simulating SARAH exposure or radioactive bomb blast. Communication through microphone. Physiology modeled.

Survey Physicians, nurses in USA; n=13. Satisfaction survey. 1, 2, 3. 0.62 reported a change in their feelings about working in an ED team. Mean ratings showed users felt immersed (4.5 ± 0.5) and the SG increased confidence in ability to respond to trauma (2.0 ± 0.5 after SG). Most thought the SG would be useful for learning teamwork (mean = 3.775) and clinical skills (mean = 3.155). Comments indicate users perceived the SG as realistic.

LeBoy et al.2008 Peninsula City Multiplayer VE of a triage area after an explosion. Communication through microphone.

Survey Paramedics, physicians, nurses in USA; n=16. Satisfaction survey. 1, 2, 3. 75% and 96% of participants thought the SG was useful for refresher training and initial training, respectively. 62% thought the SG was as/more effective than 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, 99.7% 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.
Lin et al. 2015  
SICKO (Surgical Improvement of Clinical Knowledge Ops)  
Same game format as Scrivna (above), but with timed surgical CBS. Learner triages and prioritizes care, decides whether or not to operate, makes intraoperative decisions.  
Validation study: comparison of performance vs. skill level  
Medical students, residents and surgeons in USA; n=49.  
Performance during gameplay compared across skill level, survey.  
1, 2, 3, 4  
Mean total SG scores for the novice, junior resident, senior resident, and expert groups were 5,461; 8,519; 11,404; and 13,913, respectively (p=.001). Usability survey results were positive, with mean scores ranging from 3.52 to 4.28 across ten questions.

McKenzie et al. 2013  
Game Informed Online Learning About Managing Aggression in Health Settings  
Informed learning activity with VP scenarios involving aggression and balance of tasks in a healthcare setting. Embedded multimedia tools including video clips and external websites.  
Comparison trial, not randomized: SG vs. control (didactic 90-minute teaching session)  
Clinical psychology students in UK; n=68.  
Case-based pre/post testing, self-reported confidence, usability survey.  
1, 2, 3, 5, 6  
Knowledge scores increased significantly following training in SG (t=5.564, df=34, p=0.001) and control group (t=.913, df=30, p=0.001). A split plot ANOVA illustrated a significant interaction effect between group and time: F(1,64)=11.11, p=0.001, g2=0.148, large effect size. Post-test scores vs. pre-test scores significantly higher for control vs. SG group. Confidence scores increased significantly following training in both SG (t=4.345, df=34, p=0.001) and control group (t=4.842, df=30, p=0.001). Mean educational activity rating scores for both groups were positive, but the control had significantly more positive scores than SG group in perception of learning tool as: interesting, (t=3.654, df=41.88, p=0.001) easy to follow (t=4.08, df=31.14, p=0.001) and helpful (t=3.34, df=37.12, p=0.003).  

Mohan et al. 2014  
ED Physician Decision Making  
VE, users evaluate and manage cases simulating a busy ED shift. One version contained distracting cases and other audiovisual distractors. Timed, physiology modeled.  
RCT: SG with distractors (cognitive load, CL) vs. SG without distractors (control, C)  
Emergency physicians in USA; n=209.  
Performance during game: transfer to trauma center, number of orders entered, time spent, CT scan rates.  
4, 5  
C was significantly more likely to appropriately transfer severely injured, hemodynamically unstable, and younger patients to trauma centers than CL group (40% vs 25% CL, p=0.01; 49% vs. 28% CL, p=0.04; 44% vs. 27% CL, p=0.01). Same average number orders entered (10.9 C [SD 4.8] vs. 10.7 CL [SD 5.6], p=.074), but less time spent per case in C group (9.7 C [SD 7.1] vs. 11.7 C [SD 6.7], min, p=0.01). No difference in transfer of non-representative cases reliant on heuristics (45% C vs. 34% CL, p=0.20). Higher transfer of non-representative cases for C group (38% C vs. 26% CL, p=0.03). Overall physician transfer rates (31%) and CT scan rates (62%) were consistent with rates reported in literature for actual clinical practice (30%, 57-67%, respectively).

Mujle et al. 2015  
VPs for cancer nursing education  
VE simulation scenarios to teach nursing topics in caring for men with prostate cancer.  
Survey  
Nurses, students  
Survey.  
1, 2, 3  
The majority of respondents reported an increase in knowledge and suggested that they would recommend the resource to others.

Nicoladou et al. 2015  
Virtual Emergency Telemedicine (VETM) Game  
VE simulating cardiac emergencies through a telemetry system.  
Survey  
Ambulance crew nursing personnel in Cyprus, n=90.  
Survey.  
1, 2, 3  
Along tested domains (user interface, difficulty level, feedback, educational value, user engagement, terminology), means ranged from 3.25 to 3.95. Analysis of log files showed a low success rate (20.6%). Participants described educational value and usefulness of the SG for pre-hospital emergency training (mean 3.93, SD 1.05), but identified confusing features and provided input for improving them.

Nosek et al. 2007  
Cancer Genetics Tower  
VE, users progress through levels and complete patient-based tasks  
Survey  
Medical students in USA; n=17.  
Survey.  
1, 2, 3  
Participants rated the SG: excellent (10/17), good (9/17), or average (1/7). Participants found the quality of the graphics acceptable, and felt the SG held their interest. 14/17 felt the SG was helpful in learning the content and would be interested in using it more if content were added. Some found the speed of download too slow.

O’Neill et al. 2012  
GRAPHIC (Games Research Applied to Public Health with Innovative collaboration)  
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.  
Survey  
Dental students in UK; n=117.  
Survey.  
1, 2, 3, 6  
Positive feedback, not described. Learning outcomes achieved with all participants successfully progressing the game. The software used in the construction of the game limited certain aspects of the interaction of the participants with the game.

Vankpuram et al. 2014, 2015  
ACLS Virtual Reality Simulator  
Multiplayer hospital VE, learners communicate via microphone with members (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.  
Two RCTs: Persuasive (P) vs. minimally persuasive (MP) SG. Control (didactic) vs. P vs. MP.  
ACLS-certified physicians in US; n=984; n=148.  
Useability survey.  
1st, 2, 3, 5th  
Higher mean usability (p=0.0944) and ease-of-use (p=0.0813) scores in MP compared to P.  
No difference in post-intervention performance control vs. P (p=0.37) for pulseless electric activity (PEA) and p=.01 for ventricular fibrillation & 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 pre-post 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).  

Qin et al. 2010  
Stopping the Fountains pre-game + Virtual Orthopedic-Surgery Game  
Operating room VE, learners manipulate surgical tools via haptic device. VP present with varying degrees of blood loss, users assess and manage. Physiology modeled.  
RCT: SG with and without pre-training elements  
Orthopedic surgeons, students in China; n=21.  
Performance within SG: Time to complete blood management tasks.  
5  
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) and performance score (p=0.001 by Wilcoxon-Mann-Whitney test and p=0.001 by Kruskal-Wallis test).

Rondon et al. 2013  
Anatessis 2.0 Quiz  
Interactive MCQ with multimedia, teaching anatomy and physiology of the spine, language, hearing, and swallowing mechanisms.  
RCT: SG group vs. control (self-study)  
Speech Language & Hearing students in Brazil; n=29.  
Pre/post MCQ, re-test at 6-months.  
5, 6  
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.
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Type</th>
<th>Description</th>
<th>Research Design</th>
<th>Outcome Measures</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roy et al. (^1) 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>Descriptive</td>
<td>Medical students, physicians in USA</td>
<td>None</td>
</tr>
<tr>
<td>Schwaab et al. (^2) 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>
<td>Emergency residents in USA, n=27.</td>
<td>Satisfaction survey, comparing user experience with SL examination and oral examination.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Participants rated SL as: easy to log into (92.6%) and navigate (96.3%), fair (100%), objective (100%), conducted efficiently (100%), and realistic (92.6%). 70.3% found it more realistic than the traditional oral examination (70.3%). Participants felt comfortable communication with examiner via remote computer (100%). A majority preferred oral examinations via SL over oral examination and expressed interest in using SL for other educational experiences (66.6 and 92.6%, respectively).</td>
</tr>
<tr>
<td>Schwaz et al. (^3) 2013</td>
<td></td>
<td>Medical Faculties 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>
<td>Medical students in Czech Republic, Slovakia, n=62.</td>
<td>Satisfaction survey.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>87% of participants agreed or strongly agreed that the SG was an effective tool for learning. 81.4% agreed or strongly agreed that the SG increased knowledge in acute medicine, 74% agreed or strongly agreed that the SG was a better way to study than textbooks, 87% agreed or strongly agreed that the embedded multimedia along with timed stress factor evoked a clinical atmosphere. 85.2% agreed or strongly agreed that they liked using the interactive algorithms at home and in a classroom setting.</td>
</tr>
<tr>
<td>Schwaid et al. (^4) 2001</td>
<td></td>
<td>Anesthesia Simulator 3.0</td>
<td>VE, learners examine and treat patients during anesthetic emergencies. Physiology modeled.</td>
<td>RCT: SG vs. C (self-study) Anesthesiologist residents in USA; n=31.</td>
<td>Post testing with live simulation.</td>
<td>Significantly higher post-test score for SG vs. C group (52.6 ± 9.9 vs. 43.3 ± 5.9, p&lt;0.004).</td>
</tr>
<tr>
<td>Shawagha et al. (^5) 2013</td>
<td></td>
<td>Z-DOC</td>
<td>Technical touch-based trainer for z-plasty procedure.</td>
<td>Descriptive</td>
<td>Residents, physicians in Canada.</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Development of SG described.</td>
</tr>
<tr>
<td>Strednay et al. (^6) 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>
<td>Anesthesiologist residents in USA; n=3.</td>
<td>Survey/questionnaire.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>56% of SG and 78% of C group rated themselves as “confident” in leading an ED team. Most subjects found their modality “useful” or “very useful.”</td>
</tr>
<tr>
<td>Sward et al. (^7) 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>
<td>Medical students in USA; n=100.</td>
<td>Post-rotation NBME Pediatric Clerkship Exam scores, satisfaction survey.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No difference in NBME exam scores, comfort with content (p=0.68), readiness for exam (p=0.52), content presented with enough detail (p=0.17), content relevant to patients (p=0.77), or user questions answered (p=0.47) between groups. 94% felt that SG visuals were attractive. Mean overall reaction to the SG score was 6.59 (SD=0.97). SG favored vs. self-study in understanding content (p=0.001), perceived help with learning (p=0.05), and enjoyment of learning (p=0.008).</td>
</tr>
<tr>
<td>Tuley et al. (^8) 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>
<td>Dental students in Brazil, n=46.</td>
<td>Post-rotation MCQ and laboratory skill tests.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1, 2, 3, 5, 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1, 2, 3, 5, 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1, 2, 3, 5, 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1, 2, 3, 5, 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1, 2, 3, 5, 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1, 2, 3, 5, 6</td>
</tr>
<tr>
<td>Youngblood et al. (^9) 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>
<td>Medical students, residents in USA; n=30.</td>
<td>Pre/post testing with live cases, satisfaction survey.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No significant difference in post-test scores between groups (p=0.40). 75% of SG and 86% of C group reported a change in their feelings about ED teams. 88% of SG and 93% of C group felt immersed. 56% of SG and 78% of C group rated themselves as “confident or very confident” in leading an ED team. Most subjects found their modality “useful” or “very useful” for learning clinical skills and teamwork (100%, 94% SG, 100% C).</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.