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>Accessed</td>
<td>November 29, 2017 1:55:35 PM EST</td>
</tr>
<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>

(Article begins on next page)
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Year</th>
<th>Methodology</th>
<th>Design</th>
<th>Intervention</th>
<th>Outcome Measures</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavazza et al.</td>
<td>2013</td>
<td>Interactive Storytelling</td>
<td>VP cases</td>
<td>Medical students in USA</td>
<td>None</td>
<td>Development of SG described.</td>
</tr>
<tr>
<td>Chan et al.</td>
<td>2012</td>
<td>Ultrasound-guided needle placement</td>
<td>Technical trainer</td>
<td>Radiology residents in China: n=21</td>
<td>Post testing via live simulation.</td>
<td>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.</td>
</tr>
<tr>
<td>Coyan et al.</td>
<td>2011</td>
<td>Off-pump Coronary Artery Bypass</td>
<td>Multi-player VE, teaching the steps of off-pump coronary artery bypass</td>
<td>Cardiac surgeons in Canada</td>
<td>None</td>
<td>Development of SG described.</td>
</tr>
<tr>
<td>Coyan et al.</td>
<td>2010</td>
<td>Total Knee Arthroscopy (TKA) Game</td>
<td>VE teaching the steps of a TKA, with MCQ sub-games interspersed</td>
<td>Medical students in Canada</td>
<td>Survey</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>Craig et al.</td>
<td>2014</td>
<td>Virtual Pathology Instructor (V-PIN)</td>
<td>Interactive CBS with MCQ: branched narratives with virtual pathology materials that guide learners through diagnosis.</td>
<td>Medical students in USA: n=185</td>
<td>Post-intervention MCQ, SG performance survey.</td>
<td>1, 2, 3, 5</td>
</tr>
<tr>
<td>Creutzfeldt et al.</td>
<td>2006, 2010, 2012, 2014</td>
<td>Multiplayer Virtual World CPR Game</td>
<td>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.</td>
<td>Medical students in Sweden: n=12, 15, 55, 30</td>
<td>Post testing via live simulation, time and knowledge gain testing within SG, survey.</td>
<td>1, 76.0+5, 76.0+5, 76.0+5, 76.0+5</td>
</tr>
<tr>
<td>Davids et al.</td>
<td>2011</td>
<td>Electrolyte Workshop</td>
<td>VP cases simulating electrolyte abnormalities, augmented with multimedia. Physiology modeled.</td>
<td>Residents, physicians in South Africa; n=16.</td>
<td>Usability survey.</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>De Leo et al.</td>
<td>2014</td>
<td>Game Medical Team Training (GaMeTT)</td>
<td>Multi-player VE teaching interprofessional acute care (microphone communication). Learners manage virtual military and civilian casualties. VP operated remotely by live person. Physiology mapped.</td>
<td>Air Force National Guard Medical Service in USA, n=29.</td>
<td>Survey</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>Delasobra et al.</td>
<td>2010</td>
<td>Cardiac Arrest!</td>
<td>VE where users manage patients requiring advanced cardiac life support skills.</td>
<td>Paramedics in India; n=117.</td>
<td>Pre/post MCQ, cardiac arrest live simulation, re-testing at 3 weeks.</td>
<td>1, 2, 3, 5</td>
</tr>
<tr>
<td>Dev et al.</td>
<td>2011</td>
<td>Clinispace</td>
<td>VE of emergency department. Physiology modeled.</td>
<td>Physicians, students.</td>
<td>Survey.</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>Dietl et al.</td>
<td>2013, 2015, 2015</td>
<td>InsuOnline</td>
<td>VE with diabetic VP. Embedded quizzing and multimedia assets.</td>
<td>Medical students and residents in USA: n=20, n=41.</td>
<td>Pre/post MCQ, re-testing at 3 months.</td>
<td>1, 2, 3, 5</td>
</tr>
</tbody>
</table>
Katz, Kangt, 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 Septis 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 (5.85, 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 might would recommend the SG.

Forseca et al. 2014, 2015 e-Baby VE of an incubator, teaching clinical evaluation of preterm infants’ respiratory system. Embedded multimedia assets to share within a Moodle course Survey Nurses, nursing students in Australia; n=14.8 Satisfaction survey. 1, 2, 3, 4 Development of SG described. 74% 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 training. 80% enjoyed learner autonomy, 86% felt feedback was immediate, 7% felt technologies like this could replace teachers.


Graafland et al. 2014 Medialis Timed CBS with MCQ augmented 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. 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 74 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 Majorly found SG realistic (84.4%-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 Games VE simulating a minimally invasive surgery (MIS) unit, testing learner’s assessment and performance 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. 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; 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=0.019).

Jalink et al. 2014 Underground world VE: learner demolishes and rebuilds mine using 2 Wii 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 Experts were 111% faster (p<0.001) than novices. Also, scores of the FLS Peg Transfer test and the Wii 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.

Kaczmingczyk et al. 2015 Video-based VE teaching and testing acute management of tachyarrhythmias 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 pathology. 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 after 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%).

Katz, Central Venous VE for performing central venous Descriptive Residents, None. None. Development of SG and future studies described.
Data from various studies and methodologies are presented in the table below:

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting</th>
<th>Description</th>
<th>Sample</th>
<th>Methods</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>LeRoy et al. (2014)</td>
<td>Spaced retrieval (SE)</td>
<td>Blood pressure (BP) control game</td>
<td>Competitive MCQ game, learners aim to maintain BP at regular intervals.</td>
<td>RCT: SG vs. control.</td>
<td>Higher post-test scores in SE group vs. control group (90% vs. 78%; SD 19, p&lt;0.001). Median time to BP target in 17.86 hypertensive patients was decreased in SE vs. control groups (142 vs. 148 days, p=0.018). Hazard ratio for time to BP target in SE group was 1.043 (95% CI, 1.007–1.081; PH&lt;0.018). The number of hypertensive episodes needed to treat to normalize one additional patient’s BP was 67.8. The number of clinicians needed to teach to achieve this was 0.43.</td>
</tr>
<tr>
<td>Kerfoot et al. (2012)</td>
<td>SE Urology Content Game</td>
<td>Competitive MCQ game on core urology content, 2 MCQs with feedback emailed daily.</td>
<td>Baseline scores and completion rates were higher in SE compared with standard exam scores.</td>
<td>1, 2, 3</td>
<td>Baseline scores (median 62%, interquartile range [IQR] 17%) correlated with scores on 2008 American Urological Association in-service examination (ISEQ08, IS09, 2009 American Board of Urology qualifying examination (ISE09) (r=0.76, 0.46, and 0.64, respectively; all p&lt;0.001). Baseline scores varied by gender, country, degree, and training year (all p&lt;0.001). Completion scores (median 100%, IQR 2%) correlated with ISEQ08 and ISEQ09 scores (r &lt;0.35, p&lt;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=0.001). 72% surveyed requested future SE games.</td>
</tr>
<tr>
<td>Kerfoot et al. (2012)</td>
<td>SE Pre-Clinical Content Game</td>
<td>Competitive MCQ game on pre-clinical content, 2 MCQ with feedback emailed daily.</td>
<td>Baseline scores and completion rates were higher in SE compared with standard exam scores.</td>
<td>1, 2, 3</td>
<td>Baseline median score was 53% (IQR 16) and varied significantly by year (p&lt;0.001, max +2.08), school (p&lt;0.001, dmax = 0.75), and gender (p&lt;0.001, d = 0.38). Median completion score was 93% (IQR 12) and varied significantly by year (p&lt;0.001, dmax = 1.12), school (p&lt;0.001, dmax = 0.34), and age (p&lt;0.019, dmax = 0.43). Scores did not differ significantly between years 3 and 4. 70% surveyed requested future SE games.</td>
</tr>
<tr>
<td>Kerfoot et al. (2012)</td>
<td>SE Urology Continuing Medical Education Guidelines Game</td>
<td>Competitive MCQ game on urology guidelines, 2 MCQ with feedback emailed daily.</td>
<td>Baseline scores and completion rates were higher in SE compared with standard exam scores.</td>
<td>1, 2, 3</td>
<td>Median baseline score was 48% (IQR 17), and median completion score was 98% (IQR 25) for SE group vs. control group performed worse than 2 MCQ every 2 days (d = 0.43, p&lt;0.001). 76% of participants requested to participate in future SE games.</td>
</tr>
<tr>
<td>Kizakevich et al. (2007)</td>
<td>Sim-Patient™</td>
<td>VE with single and multiple-casualty cases (bioterrorism, trauma, improvised explosive device).</td>
<td>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).</td>
<td>1, 2, 3</td>
<td>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).</td>
</tr>
<tr>
<td>Knight et al. (2010)</td>
<td>Traige Trainer</td>
<td>VE of major incident scenario where users perform medical checks, tag and assess multiple patients.</td>
<td>Post testing with live simulation.</td>
<td>1</td>
<td>Significantly improved tagging and step accuracy for SG vs. C group (p=0.001, 0.05, 0.43). No difference in time to triage between the two groups (5 ± 0.42 vs. 4.35 ± 0.74 s, p=0.155).</td>
</tr>
<tr>
<td>Kureev et al. (2009)</td>
<td>Burn Center</td>
<td>Mass burn casualty disaster VE.</td>
<td>Descriptive.</td>
<td>None.</td>
<td>None.</td>
</tr>
<tr>
<td>Lagro et al. (2014)</td>
<td>GeriatrX</td>
<td>Geriatric clinical VE.</td>
<td>RCT: SG vs. C (self-study).</td>
<td>1, 2, 3</td>
<td>Significantly higher post-test scores for SG vs. C in cost consciousness (cost deviation: 13.5 +/- 3.6 vs. 15.4 +/- 3.6, p&lt;0.05). Significantly higher self-perceived knowledge in SG vs. C group for appropriateness/ cost of care (effed sizes 1.0, and 1.2). None difference for geriatric content. Most felt SG was clear (3.85±0.8), enjoyed it (4.0±0.7), wanted to finish the SG (4.0±0.9), considered it a safe environment (4.1±0.8), and felt increased knowledge (3.9±0.8). Some users were not satisfied with clarity (2.3±1.0) and feedback (2.6±1.0).</td>
</tr>
<tr>
<td>LeBoy et al. (2008)</td>
<td>Virtual ED II</td>
<td>Multiplayer VE with cases simulating SARin exposure or radioactive bomb blast.</td>
<td>Survey.</td>
<td>No.</td>
<td>62% reported a change in their feelings about working in an ED team. Mean ratings showed users felt immersed (4.47/6) and the SG increased confidence in ability to respond to trauma (2.00/5 before SG; 3.08/5 after SG). Most thought the SG would be useful for learning teamwork (mean = 3.77/5) and clinical skills (mean = 3.15/5). Comments indicate users perceived the SG as realistic.</td>
</tr>
<tr>
<td>LeBoy et al. (2008)</td>
<td>Peninsula City</td>
<td>Multiplayer VE of a triage area after an explosion.</td>
<td>Survey.</td>
<td>75% and 96% of participants thought the SG was useful for refresher training and initial training.</td>
<td>62% thought the SG was as/more effective than traditional methods.</td>
</tr>
<tr>
<td>Lim et al. (2008)</td>
<td>eNLG (online Neurological Localization Game)</td>
<td>Modified essay and MCQ of patients with neurological problems, with learners examining and managing patients.</td>
<td>Survey.</td>
<td>93.4% of participants felt that the eNLG helped them better understand neurological localization principles, 90.7% believed the question-based format was suitable, 98.1% 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.</td>
<td></td>
</tr>
</tbody>
</table>
Lin et al. 2015  
SICKO (Surgical Improvement of Clinical Knowledge Ops)  
Same game format as Sepriva (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 = 0.01). 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=-5.66, df=34, p<0.001) and control group (t=-9.13, 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, g=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=-4.34, df=34, p<0.001) and control group (t=-4.84, 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=3.14, df=37.12, p<0.003).

Mohed 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 appropriate transfer severely injured, hemodynamically unstable, and younger patients to trauma centers than CL group (40% vs 25% CL, p=0.01; 49% C vs. 28% CL, p=0.01; 44% C 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 = 0.74), 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 relying 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).

Moyle et al. 2015  
VPs for cancer nursing education  
VP 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.

Nicolaidou et al. 2015  
Virtual Emergency Telemedicine (VETM) Game  
VE simulating cardiac emergencies through a telemedicine system.  
Survey Ambulance crew nurses 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 3.2/5 to 3.95/5. Analysis of log files showed a low success rate (20.6%). Participants described educational value and usefulness of the SG for pre-emergency training (mean 3.93, SD 1.06), but identified confusing features and provided input for improving them.

Nosok 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 (6/17), or average (1/17). Participants found the quality of the graphics acceptable, and felt the SG held their attention. 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. not provided.  
Survey, SG completion.  
1, 2, 3, 6  
Positive feedback, not described. Learning outcomes achieved with all participants successfully passing the game. The software used in the construction of the game limited certain aspects of the interaction of the participants with the game.

Vankjipuran et al. 2014  
ACLS Virtual Reality Simulator  
Multiplayer hospital VE, learners communicate via microphone with mentor (a team of experts 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.  
ACLS-certified physicians in US; n=96;148.  
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=0.10 for ventilator fibrillation & ventilator 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  
Anatees 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.
| Roy et al.\(^2\) 2006 | Virtual Standardized Patients | VE where learners examine and care for VP: Augmented multimedia. Communication via microphone. | Descriptive | Medical students, physicians in USA. | None. | None | Development of SG described. |
| Schwaab et al.\(^{17}\) 2011 | Second Life (SL) Case-Based Mock Emergency Medicine Board Examination | 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. | Survey, participants given traditional oral exam/1mo prior to SG exam | Emergency residents in USA, n=27. | Satisfaction survey, comparing user experience with SL examination and oral examination. | 1, 2, 3 | Participants rated SL as: easy to log into (92.6%), navigate (96.3%), far (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 (86.6% and 92.6%, respectively). |
| Schwid et al.\(^{17}\) 2013 | Medical Faculties Network (MEFANET) | Timed CBS in basic life support, emergency, anaesthesiology, pain, and critical care, with embedded multimedia. Physiology modeled. | Survey | Medical students in Czech Republic, Slovakia, n=62. | Satisfaction survey. | 1, 2, 3 | 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. |
| Schwiid et al.\(^{17}\) 2001 | Anesthesia Simulator 3.0 | VE, learners examine and treat patients during anesthetic emergencies. Physiology modeled. | RCT: SG vs. C (self-study) | Anesthesiologists residents in USA, n=31. | Post testing with live simulation. | 5 | Significantly higher post-test score for SG vs. C group (52.6 ± 9.9 vs. 43.3 ± 5.9, p=0.004). |
| Stredney et al.\(^{16}\) 1996 | Virtual Simulated Epidural Anesthesia | Hospital VE of a simulated patient’s back, for practicing epidural placement, with computer-generated needle. | Survey | Anesthesiologists residents in USA, n=3. | Survey/questionnaire. | 1, 2, 3 | Participants felt that the interface was comfortable to very comfortable to use, interface was straightforward to use. Participants reported the SG was “too mechanical,” movements were “stiff.” Comments for improvement included: need for additional visual cues for orientation, introduction of a method to angle the needle. |
| Sward et al.\(^{15}\) 2008 | Pediatric Rotation Game | Testing environment of a web-based, interactive adaptation of an existing board game for general pediatric content. | RCT: SG vs. C (self-study) | Medical students in USA, n=100. | Post-rotation NBME Pediatric Clerkship Exam scores, satisfaction survey. | 1, 2, 3, 5, 6 | 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 6.5/9 (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). |
| Tuleja et al.\(^{12}\) 2015 | Zinc Phosphate Cement Virtual Learning Object | Needle-based VE teaching and testing handling of zinc phosphate cement for dental prostheses. | RCT: Longitudinal SG (15 days) vs. SG (20 minutes) vs. longitudinal reading vs. reading (C) | Dental students in Brazil, n=46. | Pre/post MCQ and laboratory skill tests. | 1, 2, 3, 5 | MCQ post-test showed a significant difference between the longitudinal groups (C 6.0 ± 1.15 and SG 7.33 ± 1.43, p<0.05), but no significant differences between immediate groups. The immediate and longitudinal SG groups finished with lower film thickness than the controls (C 25 ± 9.3 and SG 16.24 ± 5.17, longitudinal C 50 ± 27.08 and longitudinal SG 22.5 ± 9.65, p=0.05 for all groups). SG groups had higher setting times, and the immediate group showed a significant difference (C 896 ± 218.90 and SG 1138.5 ± 177.95, p<0.05 for immediate group). |
| Youngblood et al.\(^{11}\) 2008 | Virtual ED | Multiplayer emergency room VE, users examine and treat patients with a team. Users communicate through microphone. Physiology modeled. | RCT: SG vs. C (live simulation) | Medical students, residents in USA, n=30. | Pre/post testing with live cases, satisfaction survey. | 1, 2, 3, 5 | 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). |

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.