DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

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

Citation
Olszewski, Aleksandra E. 2016. Virtual Simulation and Serious Games for Medical Education: A Review of the Literature and Development of a Virtual Peritoneal Dialysis Simulator. Doctoral dissertation, Harvard Medical School.

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:27007751

Terms of Use
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 http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

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

Accessibility

		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.
Cavazza et al.[50] 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.[18] 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 [51.5-34.6 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.
Cowan et al.[33] 2011	Off-pump Coronary Artery Bypass Game	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.[35] 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.[59] 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 "V-PIN cases helped me better understand the concepts."
Creutzfeldt et al.[77,86,100,105] 2008, 2010, 2012, 2014	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.[75,85,95] RCT: CPR training followed by SG refresher 6 vs.18mo prior to assessment vs. control (no refresher).[92]	Medical students in Sweden; n=12,[76,85,95] n=30.[92]	Post testing via live simulation, time and knowledge gain testing within SG,[92,85] survey.[76,95]	1,[76,85,94] 2,[76,85,95] 3,[76,85,95] 5[85,92]	Students reported enjoyment and learning with SG. Self-efficacy increased after first [5.9 to 6.5/7 (p=0.01)] and second SG session [6.0 to 6.7/7 (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).[76] 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 (ρ=0.61) (Spearman rank order correlation, p<0.001).[95] 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).[85] 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), 4.5 (±1.0) for 6mo, 18mo and control, respectively. Incorrectly delivered chest compressions 54 (±44)%, 44 (±49)%, 0% for control, 18mo, and 6mo (p<0.001 control or 18mo vs. 6mo).[92]
Davids et al.[17] 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.[66] 2014	Game Medical Team Training (GaMeTT)	Multi-player VE teaching inter-professional 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).
Delasobera et al.[44] 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 vs. 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 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 89% of M reported enjoyment.
Dev et al.[16] 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.[19,94,101] 2013, 2015, 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.[19] 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 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). Beliefs improved in SG group only.

Duque et al.[78] 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.[69] 2014	Septris	CBS about sepsis and shock. Users 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 maybe would recommend the SG.
Fonseca et al.[54,79] 2014, 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[78]	Nurses, nursing students in Australia; n=14.[78]	Satisfaction survey.[78]	1,[78] 2,[78] 3[78]	Development of SG described.[54] 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. 93% enjoyed learner autonomy, 86% felt feedback was immediate, 7% felt technologies like this could replace teachers.[78]
Foster et al. 2009.[55]	3D Renal Care Learning Environment	VE with MCQ, learner diagnoses and manages renal patients. 'Ask the Expert' button guides learners.	Descriptive	Nurses in Australia.	None	None.	Development of SG described.
Gleason et al.[51] 2015	Research & Evidence Learning in Medicine	Timed CBS with text-based PICO exercises.	Descriptive	Medical students in USA.	None.	None.	Development of SG described.
Graafland et al.[99] 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.	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.[70] 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.[98] 2014	Laparoscopic Equipment Failure Serious Game	VE simulating a minimally invasive surgery (MIS) unit, testing learner's 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 1.00–3.00), S 3.00 (1.00–4.00), R 2.00 (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.
Hannig et al.[95] 2012	eMedOffice	VE simulating design of a medical office for optimizing patient care.	Quasi-experimental	Medical students in UK; n=41.	Pre/post self assessment, survey.	1, 2, 3	Overall usability of SG ranked highly (4.07/5). Significant post-test difference in self-assessment scores.
Hashimoto et al.[65] 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.[74,96] 2014	Underground world	VE: learner demolishes and rebuilds mine using 2 Wii Remote controllers in laparoscopic tool shells. Trains basic laparoscopic skills, not medical knowledge.	Quasi-experimental	Laparoscopic surgeons at a global conference; n=72.	Speed within SG, satisfaction survey.	1,[73] 2,[73] 3,[73] 4[99]	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).[99] On a 1-to-10 scale, mean score for hardware realism was 7.2, mean score for usefulness as a training tool was 8.4.[73]
Janssen et al.[92] 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.
Kaczmarczyk et al.[89] 2015		Video-based VE teaching and testing acute management of tachyarrhythmias	Survey	Medical students in UK; n=47.	Satisfaction survey, open-ended questions.	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[102] 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.

Demaria[52] 2013	Catheter Game	catheter placement.		students in USA.			
Kerfoot et al.[107] 2014	Spaced-education (SE) blood pressure (BP) control game	Competitive MCQ game, learners emailed MCQs at regular intervals and relative scores posted to foster competition among peers. BP recordings of hypertensive patients were evaluated.	RCT: SG users vs. control (identical content online)	Primary care clinicians in USA; n=111.	Pre/post MCQ at 14 weeks, time to BP target in hypertensive patients.	5, 7	Higher post-test scores in SE group vs. control group (90% [SD 8] vs. 78% [SD 19], respectively; Cohen d 0.8, p<0.001). Median time to BP target in 17,866 hypertensive periods 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; P=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.
Kerfoot, Baker[60] 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-experimental	Urology residents in USA and Canada; n=931.	Baseline scores and completion rates within SE compared with standardized exam scores.	1, 2, 3	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 (QE09) (r<0.76, 0.46, and 0.64, respectively; all p<0.001). Baseline scores varied by gender, country, degree, and training year (all p<0.001). Completion scores (median 100%, IQR 2%) correlated with ISE08 and ISE09 scores (r <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<0.001). 72% surveyed requested future SE games.
Kerfoot et al.[75] 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.	Quasi-experimental	Medical students in USA; n=731.	Baseline scores and completion rates within SE.	1, 2, 3	Median baseline score was 53% (IQR 16) and varied significantly by year (p<0.001, dmax =2.08), school (p<0.001, dmax= 0.75), and gender (p< .001, d = 0.38). Median completion score was 93% (IQR 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% surveyed requested future SE games.
Kerfoot, Baker[60] 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 world-wide; n=1470.	Baseline scores and completion rates within SE.	1, 2, 3	Median baseline score was 48% (IQR 17), and median completion score was 98% (IQR 25) 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.
Kizakevich et al.[84] 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 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.[37] 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 UK; n=91.	Post testing with live simulation.	5	Significantly improved tagging and step accuracy for SG vs. C group (Chi2 = 13.126, p=0.02; Chi2 = 5.45, p=0.0196). No difference in time to triage between the two groups (456 ± 62 s vs. 435 ± 74 s, p=0.155).
Kurenov et al.[56] 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.
Lagro et al.[71lim] 2014	GeriatriX	Geriatric clinic VE. Users care for VP 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 Netherlands; n=134.	Pre/post MCQ, confidence survey.	1, 2, 3, 5	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 (effect sizes 1.0, and 1.2). No difference for geriatric content. Most felt SG was clear (3.8/5±0/8), enjoyed it (4.0/5±0.9), wanted to finish the SG (4.0/5±0.9), 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±1.0) and feedback (2.6/5±1.0).
LeRoy et al.[36] 2008	Virtual ED II	Multiplayer VE with cases simulating Sarin exposure or radioactive bomb blast. Communication through microphone. Physiology modeled.	Survey	Physicians, nurses in 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; 3.08/5 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.
LeRoy et al.[36] 2008	Penninsula 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 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.[80] 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.

Lin et al.[97] 2015	SICKO (Surgical Improvement of Clinical Knowledge Ops)	Same game format as Septris (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[83] 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=66.	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=-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, gq2=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=-4.346, df=34, p<0.001) and control group (t=-4.642, 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).
Mohan et al.[58] 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% C 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 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).
Moule et al.[90] 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.[88] 2015	Virtual Emergency TeleMedicine (VETM) Game	VE simulating cardiac emergencies through a telemedicine 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 3.2/5 to 3.99/5. 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 3.83, SD 1.05), but identified confusing features and provided input for improving them.
Nosek et al.[46] 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 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.[82] 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.	Satisfaction 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.
Vankipuram et al.[61] 2014, Khanal et al.[63] 2014	ACLS Virtual Reality Simulator	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.	Two RCTs: Persuasive (P) vs. minimally persuasive (MP) SG.[61] Control (didactic) vs. P vs. MP.[63]	ACLS-certified physicians in US; n=96[61], n=148.[63]	Usability survey.[60] Pre/post performance on simulated tasks.[63]	1[81], 2, 3, 5[83]	Higher mean usability (p=0.0944) and ease-of-use (p=0.0813) scores in MP compared to P.[61] No difference in post-intervention performance control vs. P (p=0.37 for pulseless electric activity (PEA) and p=0.1 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).[63]
Qin et al.[62] 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.[109] 2013	Anatesse 2.0 Quiz	Interactive MCQ with multimedia, teaching anatomy and physiology of the speech, 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.[57] 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.[67] 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 exam1mo 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%) 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).
Schwarz et al.[21] 2013	Medical Faculties Network (MEFANET)	Timed CBS in basic life support, emergency, anesthesiology, 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.
Schwid et al.[101] 2001	Anesthesia Simulator 3.0	VE, learners examine and treat patients during anesthetic emergencies. Physiology modeled.	RCT: SG vs. C (self-study)	Anesthesiology 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).
Shewaga et al.[53] 2013	Z-DOC	Technical touch-based trainer for z-plasty procedure.	Descriptive	Residents, physicians in Canada.	None.	None.	Development of SG described.
Stredney et al.[81] 1996	Virtual Simulated Epidural Anesthesia	Hospital VE of a simulated patient's back, for practicing epidural placement, with computer-generated needle.	Survey	Anesthesiology 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.[72] 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).
Tubelo et al.[103] 2015	Zinc Phosphate Cement Virtual Learning Object	Moodle-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.[48] 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%, 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.