**Data extraction**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Study question** | **Study Design** | **Setting/date** | **Participants** | **Intervention** | **Control** | **Outcomes** | **Results** | **Conclusion and limitations** |
| **RANDOMISED CONTROLLED TRIALS** | | | | | | | | | |
| Cheng 2013 | To determine whether use of a scripted de- briefing by novice instructors and/or simulator physical realism affects knowledge and performance in simulated cardiopulmonary arrests. | RCT | North America  2008-2011 | 387 Interprofessional (PALS instructors) | Scripted and non-scripted debriefing with high realism | Scripted and non-scripted debriefing with low realism | Knowledge  Technical skills  Teamwork | There was no difference between scripted or non-scripted groups' mean knowledge Baseline (no sig difference in MEAN scripted vs non-scripted): (69.3% [67.6-71.1] vs 69.1% [67.4-70.8]; P = .87)  -post debriefing (no sig difference in MEAN scripted vs non-scripted): (74.6% [73.1-76.3] vs 72.7% [71.1-74.3]; P = .09). - Scripted vs non-scripted overall improvement in MEAN (scripted higher): (5.3% [4.1%-6.5%] vs 3.6% [2.3%-4.7%]; P = .04). In the high vs low realism groups, no difference in knowledge scores baseline (p= .24) or after the second sim (4.9% [3.7%-6.1%] vs 4.0% [2.8%-5.2%]; P = .29).  Behavioral Assessment Tool (BAT) - team leader performance: Baseline (no sig difference in MEDIAN scores scripted vs non-scripted): (52% [38%-71%] vs 54% [40%-67%]; P = .99); post debriefing (no sig difference in MEDIAN scores scripted vs non-scripted): (82% [62.5%- 90%] vs 74.6% [54.5%-88%]; P = .24). Scripted vs non-scripted overall improvement in MEDIAN (scripted higher): 16% [7.4%-28.5%] vs 8% [0.2%-31.6%]; P = .03). Baseline: No difference in scores low realism vs high for BAT (P = .82); 2nd sim: No difference in scores low vs high for BAT scores (12.0% [6.4%-32.7%] vs 12.7% [0.4%-26.5%]; P = .28)  Clinical Performance Tool (CPT) - team performance: Baseline (no sig difference in MEDIAN scores scripted vs non-scripted): (73% [68.2%-79.3%] vs 74.6% [69.8%-76.6%]; P = .95) ; post debriefing (no sig difference in MEDIAN scores scripted vs non-scripted): (82.5%[79.3%-87.3%]vs82.5%[77.7%-85.7%];P = .38). Scripted vs non-scripted overall improvement in MEDIAN (scripted higher but not sig.): 7.9% [4.8%-15.1%] vs 6.7% [2.8%-12.7%]) (P = .18). Baseline: No difference in scores low realism vs high for CPT (P = .34); 2nd sim: No difference in scores low vs high for CPT (7.9% [4.8%-14.3%] vs 6.4% [3.2%-12.7%]; P = .23) | The use of a standardized script by novice instructors to facilitate team de- briefings improves acquisition of knowledge and team leader behavioral performance during subsequent simulated cardiopulmonary arrests. Implementation of de- briefing scripts in resuscitation courses may help to im- prove learning outcomes and standardize delivery of debriefing, particularly for novice instructors.  Limitations  \*One scenario type  \*Learners exposed to only one scenario and one debriefing prior to assessment  \*Debriefing script didn't have information on how to utilize it (additional training on debriefing script may have helped)  \*Debriefing times didn't vary (limited to 20 minutes)  \*Variable adherence to debriefing script |
| Curran 2015 | The effect of low and high-fidelity mannequin simulators for NRP learning outcomes, and more specifically on teamwork performance and confidence | RCT | Canada  2012 | 66 Medical Students | High fidelity simulator | Low fidelity simulator | Satisfaction  Confidence  Technical skills  Teamwork | The high-fidelity mannequin group had higher mean satisfaction scores than the low fidelity group (62.87 vs. 48.43, p<0.01)  Students who interacted with high fidelity mannequins reported higher overall confidence levels than those who worked with low fidelity mannequins (1,723.61 vs.1,489.77, p<0.01)  Megacode skills: no difference between groups (M = 10.3 vs. 10.0, p = 0.45)  No difference in mean "observability" (p = .144) or "Frequency" (p = .446) of teamwork behaviors between groups | Medical students reported greater satisfaction and confidence with high- fidelity mannequin simulators but did not demonstrate overall significantly improved team- work or integrated skills station performance. Low and high-fidelity mannequin simulators facilitate similar levels of objectively measured NRP outcomes for integrated skills station and teamwork performance.  Limitations  \* Variability in instructor background  \*Differences in mannequins (different infant ages)  \*Sim to observe teamwork behaviors was brief  \* Participants all undergrad med students |
| Finan 2012 | To compare the effects of low fidelity simulator (LFS) versus high fidelity simulator (HFS) technology on performance levels, objective and subjective measures of stress in neonatal trainees. | RCT | Canada  2007 | 16 Neonatal trainees | High fidelity simulator | Low fidelity simulator | Technical skills  Indicators of stress  Teamwork | Intra-scenario: NRP score- adherence to newborn resus guidelines: When comparing LFS versus HFS, there was no significant difference in performance, as measured by NRP score (78.2%±11.7 versus 72.7%±9, P = 0.17)  Objective measures of stress measured through buccal cortisol levels increased significantly (P= 0.001) over the duration of the simulated experience. Subjective stress also increased significantly (P<0.001) from baseline to completion of scenarios. However, no difference in stress, either objective or subjective was noted between high and low fidelity groups.  Intra-scenario: Anesthesia non-technical skills (ANTS)- task management, teamwork, situation awareness, decision making. When comparing LFS versus HFS, there was no significant difference in median ANTS score between LFS and HFS groups [outcomes 5 & 7], respectively (2.87 (IQR 2.7, 3.4) versus 2.75 (IQR 2.5, 3.25), P = 0.52) | No difference in performance between the two groups. Simulated neonatal resuscitations induce a significant stress response in neonatal trainees; however, they were unable to identify any difference in stress measures between HFS and LFS.  Limitations  \*Small sample size  \* High number of female participants  \*Convenience sample  \*No power calculation performed  \*No baseline assessment of team training  \* Cannot directly attribute objective and subjective stress to simulation realism  \*Teams were only 2 people, both physicians |
| Katz 2020 | This study aimed to explore the utility of a voice-based VR ACLS team leader refresher as compared with HFS | RCT | USA  Unknown | 25 Anesthesia residents | Virtual reality | High fidelity simulator | Satisfaction  Technical skills | Satisfaction was similar between groups, but the HFS group felt they were provided better feedback 99 (IQR 89-100) vs 79 (IQR 71-88); P<.001.  Scores in technical domains, as measured in percentage correct without assistance, was significantly lower in the VR (exp) group than in the HFS (control) group (HFS: median 72.7, IQR 60.0-78.2 vs VR: median 47.0, IQR 40.0-58.0; P<.001; Mann-Whitney U Test). Scores were not dependent on the first modality encountered (VR first: median 40.5, IQR 35.5-42.75 vs HFS first: median 38.0, IQR 32.0-44.5; P=.810). | Utilization of a VR-based team leader refresher for ACLS skills is comparable with HFS in several areas, including learner satisfaction. The VR module was more cost-effective and was easier to proctor; however, HFS was better at delivering feedback to participants. Optimal education strategies likely contain elements of both modalities. Further studies are needed to examine the utility of VR-based environments at scale.  Limitations  \*Small sample size  \* Single VR experience utilized  \*Same participant type and institution  \*Tested voice skill/delegation only (no hands-on skill) |
| Kron 2017 | To assess advanced communication skills among second-year medical students exposed either to a computer simulation (MPathic-VR) featuring virtual patients, or to a multimedia computer- based learning module, and to understand each group’s experiences and learning preferences. | RCT | USA  Unknown | 421 Medical Students | Mpathic-VR (screen-based learning) | Computer based learning | Satisfaction  NTS  Teamwork | The mean ratings aggregated across the 12 survey items were significantly higher (e.g., more positive) among students with screen-based simulation (MPathic-VR) experience than for students with the computer-based learning experience, t (413) = 7.23, p < 0.0001, r2 = 0.1123.  Communication skills during post-exposure OSCE (openness/defensiveness, collaborative/competitive, nonverbal communication, presence/awareness of others): OSCE evaluators rated the communication skills of Mpathic-VR-trained students significantly higher (M = 0.806, SD = 0.201), than CBL students (M = 0.752, SD = 0.198). This analysis indicated a main effect for module, F (1, 414) = 6.09, p = 0.0140, n2 = 0.0145. In MPathic-VR, a lower score reflected better performance. For the intercultural communication scenario, scores decreased significantly from the first run-through (M = 11.67, SD = 6.26) to the second run-through (M = 5.89, SD = 5.12), F (1, 207) = 166.14, p <0.0001, h2 = 0.45.  For the interprofessional communication scenario, scores decreased significantly from the first run-through (M = 7.59, SD = 3.96) to the second run-through (M = 4.62, SD = 2.54), F (1, 207) = 104.64, p <0.0001, h2 = 0.36. Thus, students successfully learned how to improve their communication skills for both the intercultural and interprofessional scenarios. | MPathic-VR was effective in training advanced communication skills and in enabling knowledge transfer into a more realistic clinical situation.  Limitations  \*Different recruitment approaches between medical schools  \*Differences in intervals between training  \*Completing attitudinal survey before qualitative measurements  \*Small effect size in OSCE comparisons; different scenarios may have yielded different communication results |
| Lejus-Bourdeau 2021 | To compare the quality and educational impact of a full-scale simulation workshop with an HF infant simulator (SimBabyTM, Laerdal) or with a low-cost (LC) simulator composed of an inert infant mannequin with SimBabyTM software for medical education in pediatric difficult airway management. | RCT | France  2012-2013 | 129 anesthesiology, emergency medicine and intensive care physicians | High fidelity mannequin (SimBaby) | Low-cost inert mannequin | Satisfaction  NTS  Teamwork | No difference between high fidelity or low fidelity infant mannequin groups was noted in perceived education quality nor in simulation quality, with the exception that the HFS group perceived the mannequin to be more realistic and that they were more comfortable during the interaction.  ANTS (nontechnical skills)- self-evaluated by participants: The ANTS score was not different between either group. It was the same for all of its items except for ¨task management¨ considered easier in the LF (control) group with a slightly higher score (p= 0.0195).  ANTS with specific regard to teamwork no difference – low cost 14(12-15) vs high cost 14 (13-15), p=0.5257. | A low-cost simulator should be suggested as a less expensive alternative to an HF simulator for continuing medical education in pediatric difficult airway management.  Limitations  \*Facilitators required training in simulation and instrument completion  \*Tools have not been validated in these settings  \*Low response rates on delayed surveys |
| Liaw 2020 | To evaluate a team training program using virtual reality vs conventional live simulations on medical and nursing students’ communication skill performances and teamwork attitudes. | RCT | Singapore  2018 | 120 Nursing and Medical Students | Screen-based simulation | Immersive simulation | Satisfaction  NTS | There was a significant increase after team training in the Attitudes Toward IP Healthcare team (ATHCT) and IP Socialization and Valuing Scale (ISVS) posttest scores from the baselines scores for both groups. There was also a significant increase in the follow-up ISVS scores from the baseline scores for the virtual group (P=.047) but not for the simulation group (P=.14). No significant differences between the baseline and follow-up ATHCT scores were found for both groups. no significant differences in the trend between the virtual and simulation groups for both the ATHCT (P-interaction=.58, eta2 =0.005) and ISVS (P-interaction=.61, eta2 =0.004) scores. There were also no significant differences in the ATHCT (F2,118=0.507, P=.48, eta2 =0.004) and ISVS (F2,118=0.335, P=.56, eta2 =0.003) scores over the 3 time points between the virtual and simulation groups.  Team communication performance during sim -Unnamed scale developed by authors Team communication performance: no significant differences in the overall communication performance posttest scores (F2,58=1.46, P=.29, eta2 =0.33) between the virtual (mean 22.60, SD 5.31) and simulation groups (mean 23.97, SD 4.55). There were also no significant differences in the total checklist posttest scores (F2,58=3.654, P=.29, eta2 =0.28) and global posttest scores (F2,58=1.56, P=.29, eta2 =0.33) between the groups. | Study outcomes did not show an inferiority of team training using virtual reality when compared with live simulations, which supports the potential use of virtual reality to substitute conventional simulations for communication team training.  Limitations  \*Use of immediate post-test and self-reported attitudes questionnaire  \*Single location |
| Semler 2015 | To test the hypothesis that expert demonstration of teamwork principles and participation in high fidelity simulation would each result in objectively assessed teamwork behavior superior to traditional didactics | RCT | USA  2010 | 52 Internal Medical Residents | Immersive simulation  Expert demonstration | Didactic | Satisfaction  Technical skills  Teamwork | No difference in learner satisfaction was found between groups.  All learners completed an immersive simulation where the following were measured: Adherence to 2005 AHA ACLS algorithms (skills); Recognize, Respond, Reassess scoring system (skill re sequential components of providers' clinical management). No difference between the 3 groups. Teamwork behavior only weakly correlated with clinical performance (coefficient of determination [Rs2] = 0.267, P < 0.001).  All learners completed an immersive simulation where the following were measured: Teamwork Behavioral Rater score (teamwork behavior): Overall scores were significantly higher in the demonstration group compared with the didactic group (mean difference, 1.30; 95% [CI], 0.13–2.47; P = 0.045) but similar between demonstration and simulation (mean difference, 0.31; 95% CI, 21.23 to 1.84; P = 0.917). Simulation demonstrated a trend toward better scores than didactic training (mean difference, 0.99; 95% CI, 2.00 to 20.02; P = 0.068). Teamwork behavior only weakly correlated with clinical performance (coefficient of determination [Rs2] = 0.267, P < 0.001). | Teamwork training by expert demonstration resulted in similar teamwork behavior to participation in high fidelity simulation and was more effective than traditional didactics. Clinical performance was largely independent of teamwork behavior and did not differ between training modalities.  Limitations  \*Single center  \*Homogenous learners at single point in training  \*Different interventions may have yielded weaker or stronger results if structured differently  \*Does not examine combined approaches |
| Sorensen 2015 | To investigate the effect of in situ simulation (ISS) versus off-site simulation (OSS) on knowledge, patient safety attitude, stress, motivation, perceptions of simulation, team performance and organizational impact. | RCT | Denmark  2013 | 100 physicians, nurses and midwives working in the labor ward at a single institution. | Off-site simulation | In-situ simulation | Satisfaction  Knowledge  Indicators of stress  Teamwork | For almost all 20 questions in the evaluation questionnaire, the in-situ simulation and off-site simulation groups did not differ significantly. However, the two questions addressing the authenticity or fidelity of the simulations were scored significantly higher by the ISS participants compared with the OSS participants. Safety attitudes and perceived desire to change policy did not differ between groups.  There was no difference in mean post-intervention knowledge scores between in situ simulation vs off site simulation groups.  Measures of stress including salivary cortisol levels, Stress-Trait Anxiety Inventory, and cognitive appraisal showed no difference in mean change between groups doing simulation in situ vs off site.  Team performance (Team Emergency Assessment Measure): No significant differences were found in the team scoring of performance between the ISS (control) versus the OSS (exp) group. | No significant differences were found regarding knowledge, patient safety attitude, motivation or stress measurements when comparing ISS versus OSS. Although participant perception of the authenticity of ISS versus OSS differed significantly, there were no differences in other outcomes between the groups except that the ISS group generated more suggestions for organizational changes.  Limitations  \*Only immediate measurement of knowledge level and team performance  \*No clinical or patient safety data were measured  \*Previous simulation experience of 2/3rds of learners  \*No measurement of long-term retention or clinical skill level  \*Potential for Hawthorne effect |
| Stefanidis 2021 | To assess whether an integrated Advanced Modular Mannequin (AMM) provides improved participant experience compared with use of peripheral simulators alone during a standardized trauma team scenario. | RCT | USA  2020 | 42 first responders, anesthesiologists, and surgeons | Advanced Module Mannequin (AMM) | Standard mannequin | Satisfaction  Indicators of stress | Team experience ratings were higher for the computer with mannequin group (Advanced Module Mannequin) than with the mannequin group alone ("peripherals"). (Cohen’s d= .25, p= 0.016). Participant experience (GSR) varied by background with surgeons and first responders rating their experience significantly higher compared with anesthesiologists (p < 0.001). Focus groups revealed that participants preferred the integrated AMM condition based on its increased realism, physiologic responsiveness, and feedback provided on their interventions.  There were significant differences by condition (standardized B= 0.18; 95% CI, 0.04 to 0.31; p = 0.005, mixed-effects regression) with the integrated Advanced Module Mannequin [AMM] platform condition (which also involved multiple simulators) having significantly greater aggregate perceived workload than the condition that only integrated multiple simulators [peripherals condition] (Cohen’s d = .35, p = 0.014). Greater perceived workload was found for participants white and older than 35 in both the AMM (standardized b = 0.47; 95% CI, 0.29 to 0.65; p < 0.001; standardized b = 0.20; 95% CI, 0.01 to 0.38; p = 0.049) and peripherals only (standardized b = 0.26; 95% CI, 0.06 to 0.46; p = 0.003; b = 0.22; 95% CI, 0.02 to 0.41; p = 0.047) groups. In the AMM group, females found the workload conditions to be significantly higher than male participants (standardized b = 0.18; 95% CI, 0.04 to 0.31; p = 0.005) | Integration with the AMM platform provides benefits over individual peripheral simulators and has the potential to expand simulation-based learning opportunities and enhance learner experience, especially for surgeons.  Limitations  \*Anesthesiologists reported higher workload, which stemmed from frustration and unrealistic interactions with iPads, monitors, medication delivery, and other simulated technologies that were disparate from actual working conditions and flow  \*Single institution  \*Military setting may not translate to civilian setting  \*Use of novel (untested) assessment tools  \*Small sample size |
| Uslu 2020 | To determine the effectiveness of different simulation methods used for inter- professional training on gynecologic oncology palliative care knowledge, interdisciplinary education perceptions, and teamwork attitudes of health professional students and to compare these methods. | RCT | Türkiye  2016-2017 | 84 Interprofessional learners involved in palliative care training. | High fidelity simulation (HFS)  Hybrid simulation (HS)  Combination of HF and HS | Standard training | Satisfaction  Knowledge  Teamwork | Interdisciplinary Education Perception Scale (IEPS) results were as follows: No sig difference between pre-test mean scores. The High fidelity (HFS) + Hybrid simulation (HS) group's first post-test and second post-test mean scores were higher than its pre-test mean score (p = 0.001). The HFS + HS group’s first post-test (p = 0.056) and second post-test (p = 0.174) mean scores were the highest. There were not significant differences between pre and post test scores for other groups.  High-fidelity simulation (HFS) and hybrid simulation (HS) groups improved their palliative care knowledge. There was no SD between the students' PCKT pre-test mean scores (p > 0.05). The control group's first post-test mean score was higher than its pre-test mean score (p < 0.001). The first post-test and second post-test mean scores of the HFS (p < 0.001), HS (p < 0.001), and HFS + HS (p < 0.001) groups were higher than their pre-test mean scores. The HFS + HS group had the highest PCKT mean scores in the first post-test (p < 0.001) and second post-test (p < 0.001)  Teamwork Attitudes Questionnaire (TAQ): No SD between pre-test mean scores. No SD between the control and HFS groups’ pre-test, first post-test, and second post-test mean scores. The HS group's first post-test mean score was higher than its pre-test mean score (p = 0.007). The HFS + HS group's first post-test mean score was higher than its pre-test mean score (p = 0.012). | The introduction of high-fidelity simulation and hybrid simulation or hybrid simulation-based interprofessional training in undergraduate education can increase students' palliative care knowledge, inter- disciplinary education perception, and teamwork attitudes. Training programs that are used together with high- fidelity simulation and hybrid simulation applications in interdisciplinary training should be integrated into the undergraduate curricula of future cooperating health professions.  Limitations  \*Experimental groups had additional training/practice versus the control  \*Interprofessional education and palliative care training is new to Türkiye |
| Sparks 2017 | To test a simulation-based operating room team training strategy that challenges the communication abilities and teamwork competencies of surgeons while they are engaged in realistic operative maneuvers. | RCT | USA  Unknown | 15 surgical PGY2 and PGY3 surgical residents | ‘Synthetic anatomy’ mannequin (medium fidelity)  Deceased donor (high fidelity) | SimMan mannequin | Satisfaction  Confidence  Technical skills  NTS  Teamwork | Participant satisfaction post showed significant differences for both high and medium high vs low (Likert scale), where high > (but not SD) medium > (SD; 95%CI; p<.05) low fidelity.  There was no difference in self-efficacy between learners who interacted with a mannequin without surgical anatomy, a mannequin with Synthetic Anatomy for Surgical Tasks, and a deceased donor.  Technical skills (surgical management): change case 1 vs case 2: Surgical residents exposed to medium-fidelity simulation (synthetic anatomy for surgical tasks) team training scenarios demonstrated greater gains in teamwork skills compared with control groups (SimMan) (Trauma Management Skills video score: 95%CI, 0.61-2.90). Surgical residents exposed to medium-fidelity simulation (synthetic anatomy for surgical tasks) team training scenarios demonstrated equivalent gains in teamwork skills compared with high-fidelity simulations (deceased donor) (Trauma Management Skills video score: 95%CI, −1.70 to 0.49). TMS showed no SD between medium and high, and showed SD for both vs low/control. TMS showed SD for both medium and high for case 1 to case 2, and no SD for low.  NOTECHS: change case 1 vs case 2: Surgical residents exposed to medium-fidelity simulation (synthetic anatomy for surgical tasks) team training scenarios demonstrated greater gains in teamwork skills compared with control groups (SimMan) (Nontechnical Skills video score: 95%CI, 1.06-16.41). Surgical residents exposed to medium-fidelity simulation (synthetic anatomy for surgical tasks) team training scenarios demonstrated equivalent gains in teamwork skills compared with high-fidelity simulations (deceased donor) (Nontechnical Skills video score: 95%CI, −8.51 to 6.71). NOTECHS showed no SD between medium and high, and showed SD for both vs low/control. NOTECHS showed SD for both medium and high for case 1 to case 2, and no SD for low. Crisis Resource Management: change case 1 vs case 2: Surgical residents scored lower than expert surgeons; no SD between h/m/l groups.  Teamwork: [[the authors used the phrase teamwork with both NOTECHS and teamwork skills]] NOTECHS: change case 1 vs case 2: Surgical residents exposed to medium-fidelity simulation (synthetic anatomy for surgical tasks) team training scenarios demonstrated greater gains in teamwork skills compared with control groups (SimMan) (Nontechnical Skills video score: 95%CI, 1.06-16.41). Surgical residents exposed to medium-fidelity simulation (synthetic anatomy for surgical tasks) team training scenarios demonstrated equivalent gains in teamwork skills compared with high-fidelity simulations (deceased donor) (Nontechnical Skills video score: 95%CI, −8.51 to 6.71). NOTECHS showed no SD between medium and high, and showed SD for both vs low/control. NOTECHS showed SD for both medium and high for case 1 to case 2, and no SD for low. | Including a surgical task in operating room team training significantly enhanced the acquisition of teamwork skills among midlevel surgical residents. Incorporating relatively inexpensive, medium-fidelity synthetic anatomy in patient simulators was as effective as using high-fidelity anatomies from deceased donors for promoting teamwork skills in this learning group.  Limitations  \*Anonymous self-efficacy pre/post did not permit linkage to performance ratings  \*Small sample size  \*Surgeon played the role of senior resident which may have introduced bias or performance anxiety for trainees  \*Inter-rater reliability between video and in-person assessments was lower than expected  \*Single institution |
| Wang 2016 | To assess the performance of interprofessional teams of radiologists, technologists, and nurses trained with high-fidelity hands- on (HO) simulation and computer-based (CB) simulation training for contrast reaction management (CR) and teamwork skills (TS). | RCT | USA  Unknown | 40 nurses, radiology PGY2 residents and technologists | High fidelity hands-on team training | Computer based individual training | Satisfaction  Confidence  Knowledge  Technical skills  NTS  Teamwork | Hands-on team training (HO) training was more highly rated than computer-based individual (CB) as an effective educational tool (P = .01) and for effectiveness at teaching contrast reaction management (CR) and teamwork skills (TS) (P = .02).  Comfort levels in dealing with contrast reactions increased in both groups (high fidelity mannequin team training vs individual computer-based training) pre to post intervention.  Immediate post-test scores significantly improved from the pre-test in both the high-fidelity hands-on (79.6% versus 66.7%, P ¼ .002) and the computer-based groups (78.2% versus 67.6%, P ¼ .018), but the magnitude of improvement did not differ significantly between groups (p=.51).  Technical skills: contrast reaction (CR) management score: During the final scenario test, HO (exp) teams tended to receive higher grades than CB (control) teams on CR (95% versus 81%, P = .17) and made fewer errors in epinephrine administration (0/6 versus 2/5, P= .18).  Nontechnical skills: At baseline the CB (control) group reported being more comfortable using team communication skills (P = .025) and recognizing team communication skills (P = .002) than the HO (exp) group.  Teamwork: TeamSTEPPS (TS) score: HO (exp) and CB (control) teams scored similarly on TS (51% versus 52%, P = .66), but overall scores were lower for TS than for CR skills in both the HO (P = .03) and CB teams (P = .06). | High-fidelity simulation can be used to both train and test interprofessional teams of radiologists, technologists, and nurses for both CR and TS and is more highly rated as an effective educational tool by participants than similar CB training. However, a single session of either type of training may be inadequate for mastering TS.  Limitations  \*Despite both teams’ relative confidence in their teamwork and communication skills, their final scenario objective scores suggest that neither group mastered the skills with both groups only receiving 50% of total points possible  \*Computer based group statistically more comfortable at teamwork skills at baseline, which limited the possible improvement for the intervention  \*No preintervention simulation test performed so unknown if either group had significant improvement from baseline despite apparent low scores  \*Small sample size |
| Bultas 2014 | To compare the effectiveness of high-fidelity simulation (HFS) with traditional static mannequins as a teaching strategy for pediatric staff nurse education. | RCT | USA  Unknown | 33 Pediatric Nurses | High fidelity mannequin | Low fidelity mannequin | Satisfaction  Knowledge  Technical skills  Teamwork | Students in the HFS group indicated that they felt more knowledgeable and satisfied with their experience, per qualitative feedback.  No significant difference between high and low fidelity groups' knowledge acquisition was found during the immediate or delayed posttest.  PEARS Behavioral Measures Check-Off Tool (BMCT) expert rating of team performance: At 6 months, the HF (exp) mean score in the Respiratory scenario was 40.39 points compared with 26.96 points for the LF (control) group (p < 0.001). The HF group averaged a score of 30.87 points on the Circulatory scenario compared with an average score of 19.66 points for the LF group (p < 0.001).  Technical skills & Teamwork: PEARS Behavioral Measures Check-Off Tool (BMCT) expert rating of team performance: At 6 months, the HF (exp) mean score in the Respiratory scenario was 40.39 points compared with 26.96 points for the LF (control) group (p < 0.001). The HF group averaged a score of 30.87 points on the Circulatory scenario compared with an average score of 19.66 points for the LF group (p < 0.001). | Supports the use of HFS as an effective teaching modality when educating pediatric staff nurses in the identification and intervention of the deteriorating pediatric patient.  Limitations  \*Measurement tool not validated  \*Attrition of the control group  \*Small sample size  \*No blinding for participants or instructors  \*Possibility of prior exposure to cardiac and respiratory patients |
| **NON-RANDOMISED TRIALS** | | | | | | | | | |
| Weller 2014 | Primary hypothesis was that anesthesiologists would exhibit similar verbal communication patterns in a routine OR case and routine simulated case. Our secondary hypothesis was that anesthesiologists would exhibit different communication patterns in routine cases (in the OR or simulated environment) and simulated cases involving a crisis. | Cohort, observational, cross-over | New Zealand  Unknown | 40 anesthesiologists | High fidelity simulation | Real non-crisis Operating Room (OR) cases | Satisfaction  NTS | Participants considered communication in the simulated environment to be similar to that in the real environment, with team communication scored as being realistic when in simulation by 81% of responses.  Nontechnical skills: communication: The number of communication events in the CS (exp 2) was considerably higher than in either of the other two conditions - 16% (400) were in the OR (control) cases; 27% (683) in the RS; and 57% (1,419) in the CS. SD for communication 'to whom' anesthesiologist to circulating nurse OR vs RS (exp 1) (p=.010; CI 95%) and OR vs CS (p=.016; CI 95%). No SD ‘to whom’ for anesthesia technician, surgeon, scrub nurse. SD higher percentages of situational information for verbalizing patient status (p=.008), assessing patient status (p=.005), and proposing a plan of action (p=.002) occurred in the CS than in the OR. No SD for anticipates future events. No SD for CS vs RS and OR vs RS. Proposing a plan of action was SD between the RS and CS (p=.005). No SD for communication types (task assignment, requests information, responds to suggestions, statement of fact) among settings. | The similarity of teamwork-related communications elicited from anesthesiologists in simulated cases and the real setting lends support for the ecological validity of the simulation environment and its value in teamwork training. Different communication patterns and frequencies under the challenge of a crisis support the use of simulation to assess crisis management skills.  Limitations  \*Not a random sample  \*Novel coding framework  \*Future plans to explore validity of simulation for assessing crisis management not feasible |
| Coffey 2016 | To compare verbal interactions and the use of procedural touch by health care workers (HCWs) in scenarios with SPs and with mannequins | Cohort | United Kingdom  Unknown | 6 Interprofessional | Standardized patient | Mannequin | NTS  Teamwork | Nontechnical skills: Verbal and nonverbal (touch) communication: Counted the numbers of words and turns in the HCW-patient interaction and in the HCW-HCW interaction in each training session. Greater use of PAT and PT was observed in the scenarios involving an SP. Overall, there was greater use of T with the mannequins, although this pattern was more frequent with KSP than with Kman. The HCWs talked to the patient more in the cases with an SP than with a mannequin, although this did not reach statistical significance (p=.125). HCWs proposed a medical procedure to the patient with an explanation, were observed more frequently in the cases with an SP than with a mannequin.  Teamwork: Team interactions: There were more interactions among HCWs in the case of Jman than JSP, which differed from the observations of the cases with KSP and Kman. There were, however, similar patterns of team interaction within the individual teams, whether this was with an SP or with a mannequin. | The results suggest that training with an SP is particularly effective in replicating more authentic verbal and touch interactions with the patient.  Limitations  \*Lack of rigorous matching of the teams  \*Acuity of patients fluctuated in the scenarios |
| Nicolaides 2020 | To evaluate a simulation-based team performance course for medical students and compare its low- and high-fidelity components. | Cohort | Europe  2013 | 40 Medical Students | High fidelity simulated patient (hyper-realistic suit) | Low fidelity simulated patient | Satisfaction  Confidence  NTS  Indicators of stress  Teamwork | Students rated HF higher (4.88 out of 5, SD = 0.29) compared to LF (4.74 out of 5, SD = 0.67). A paired-samples t-test demonstrated a statistically significant difference (0.14 out of 5, SD = 0.37, p = .020) between evaluations of the two modules.  Overall mean confidence increased in both low and high-fidelity groups from pre to post intervention, with no difference in assessment between groups.  Nontechnical skills & Teamwork: TEAM tool (team emergency assessment measure): leadership, teamwork, task management, mean score (11 items) plus global rating: SD in mean score for LF (higher mean) vs HF (p=.023). Global score for LF was higher than HF but not SD (p=.18).  Median increases in stress levels were higher for students attending the high-fidelity module than for the low-fidelity module z = −2.08, p = 0.038.  TEAM tool (team emergency assessment measure): leadership, teamwork, task management, mean score (11 items) plus global rating: SD in mean score for LF (higher mean) vs HF (p=.023). Global score for LF was higher than HF but not SD (p=.18). | Both the low- and high-fidelity interventions demonstrated an improvement in team performance of the attending medical students. The high-fidelity intervention was more realistic, yet more stressful. Furthermore, it proved to be superior in harvesting leadership, teamwork, and task management skills. Both modules were evaluated highly by the students.  Limitations  \*Data collection was done at a single institution with a small sample size  \*Students were grouped by country of origin |
| Sanko 2012 | When learning objectives do not specifically dictate the use of one simulation modality over another, the authors sought to answer the question of which modality is preferred. They also assessed the impact of debriefing, and the frequency of participants asserting their leadership, as well as self-reported comfort and competence, and the ability to generate differential diagnoses when either a standardized patient (SP) or high-technology simulator (HTS) was used. | Cohort | USA  Unknown | 140 Medical Students | Simulated patient | Mannequin | Satisfaction  Confidence  Knowledge  Teamwork | 91% preferred the SP to the manikin and was not significantly altered by whether they encountered the SP or mannequin first (p=.279).  Confidence and competence increased amongst all learners after the second scenario regardless of whether the student first did a scenario with a mannequin or an SP (p< 0.001).  No significant difference in ability to list differential diagnoses for a scenario was found whether using an SP or a mannequin.  Nontechnical skills: Leadership: The likelihood of the participants to take on the leadership role was influenced by the debriefing intervention, but not by the modality. Nearly twice as many participants reported being a leader during their second scenario than in the first (p = 0.034). | Low and high technology have similar effectiveness for achieving learning objectives and for the demonstration of skills; however, students clearly preferred the SPs.  Limitations  \*Statistical differences may not be practical differences  \*Likert scale (5-point) may not be detailed enough to answer questions  \*Did not address finances of SP vs mannequin |
| Burns 2021 | To evaluate the impact of online, interactive TeamSTEPPS simulation versus an in-person simulation on medical students’ TeamSTEPPS knowledge and attitudes. | Cohort | USA  Unknown | 31 Medical Students | Online didactic + screen-based simulation | In person didactic + high fidelity mannequin | Satisfaction  Knowledge  Teamwork | Most participants, 92% (11; n=12), reported that they would use the online module again, and 94% (17; n=18) of the in-person participants who reported the same. When asked to quantify their agreement with the statement ‘The workshop enhanced my understanding of other professions' roles and responsibilities’, the average score for online participants was 3.8 (SD 0.8) and 4.4 (SD 0.5) for the in- person group  Pre/postintervention knowledge scores (TeamSTEPPS Benchmarks test) increased significantly from baseline (n=30, +2.0%, p=0.047), and these gains did not differ significantly based on whether participants received in- person versus online training (n=18, +1.5% CONT vs n=12, +2.9% INTV; p=0.49)  Nontechnical skills & Teamwork: As a retrospective pre/posttest, teamwork attitudes scores (TeamSTEPPS teamwork attitudes questionnaire) showed a statistically significant increase with training (n=30, +0.9, p<0.01) with no difference in the effect of training by group (n=18, +0.8 CONT vs n=12, +1.0 INTV; p=0.64) | Graduating medical students who received in-person and online teamwork training showed similar increases in TeamSTEPPS knowledge and attitudes. Online simulations may be used to teach and reinforce team communication skills when in-person, interprofessional simulations are not feasible.  Limitations  \*Small sample  \*Self-selection of students into groups  \*Ceiling effects in benchmarks and test scores  \*Online students were asked to “learn” teamwork individually |