Effect of simulation-based teaching on nursing skill performance: a systematic review and meta-analysis

Agezegn Asegid\textsuperscript{a,b,}\textsuperscript{*}, Nega Assefa\textsuperscript{b}

\textsuperscript{a}Department of Nursing, College of Medicine and Health Science, Wachemo University, Hosanna 667, Ethiopia
\textsuperscript{b}School of Nursing and Midwifery, College of Health and Medical Science, Haramaya University, Dire Dewa 138, Ethiopia

Received: 14 June 2020; Accepted: 17 July 2020; Published: 20 September 2021

Abstract: Objective: To summarize and produce aggregated evidence on the effect of simulation-based teaching on skill performance in the nursing profession. Simulation is an active learning strategy involving the use of various resources to assimilate the real situation. It enables learners to improve their skills and knowledge in a coordinated environment.

Methods: Systematic literature search of original research articles was carried out through Google Scholar, Medline, and Cochrane Cumulative Index to Nursing and Allied Health Literature (CINAHL) databases. Studies conducted on simulation-based teaching and skill performance among nursing students or clinical nursing staff from 2010 to 2019, and published in the English language, were included in this study. Methodological quality was assessed by Joanna Briggs Institute, and the risk of bias was also assessed by Cochrane risk of bias and the risk of bias assessment tool for non-randomized studies (ROBINS-I) checklists.

Results: Initially, 638 titles were obtained from 3 sources, and 24 original studies with 2209 study participants were taken for the final analysis. Of the total studies, 14 (58.3%) used single group prep post design, 7 (29.1%) used high fidelity simulator (HFS), and 7 (29.1%) used a virtual simulator (VS). Twenty (83.3%) studies reported improved skill performance following simulation-based teaching. Simulation-based teaching improves skill performance among types of groups (single or double), study regions, high fidelity (HF), low fidelity (LF), and standard patient (SP) users. But the effect over virtual and medium fidelity simulators was not statistically significant. Overall, simulation-based teaching improves the skill performance score among the experimental group (d = 1.01, 95% confidence interval [CI] [0.69–1.33], Z = 6.18, P < 0.01, 93.9%). Significant heterogeneity and publication bias were observed during the pooled analysis.

Conclusions: Simulation did improve skill performance among the intervention groups, but the conclusion is uncertain due to the significant heterogeneity. The large extent of difference among original research has necessitated the development of well-defined assessment methods for skills and standardized simulation set-up for proper assessment of their effects.

Keywords: checklist • clinical skill • education • experimental • nursing review • nursing staffs • quasi-experimental • simulation training • student nursing

© Shanxi Medical Periodical Press.

1. Introduction

Simulation is an active learning strategy involving the use of various resources to assimilate the real situation.\textsuperscript{1} Moreover, it allows students to practice skills, exercise clinical reasoning, and make patient care decisions in a safe environment.\textsuperscript{2} It is also ideal for teaching reflective skills and management of patients in a crisis situation.
2. Methods

2.1. Protocol and registration

To summarize and produce aggregated evidence on the effect of simulation-based teaching on nursing skill performance in the nursing profession, this review followed the guidelines proposed by Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA).

2.2. Eligibility criteria

Literature published in the English language, original, which deal with nursing students or nursing professionals, and which compare any type of simulation with no simulation or traditional lecture-based teaching, were included. Moreover, studies available in full text that measure the effect of simulation on skill performance, and published between 2009 and 2019 (10-year review), were also included. But, qualitative study, interprofessional study, non-nursing study, review study, study population patient, observational study, and combination training (simulation-based + other, and then simulation alone) were excluded from the review and analysis.

2.3. Participants

Participants were undergraduate nursing students and clinical nursing staff.

2.4. Intervention

Intervention was based on simulation-based teaching (using low fidelity [LF], HF, medium fidelity, standard patient (SP), and virtual based teaching).

2.5. Control

No treatment or other conventional training such as interactive lecture alone or in combination with conventional manikin-based teaching.

2.6. Outcomes

The primary outcome was skill performance score after intervention. The term score was used because an inconsistency was observed in separate reporting of acquisition and retention of skill performance. For this review, skill score was used as a general term representing a change in skill performance score following simulation-based teaching. The skill performance score was taken as it was reported by original researchers.

2.7. Information sources

Study data were obtained from the databases of Google Scholar, PubMed, Cochrane database (CINAHL), and other references.

Bland et al (2011) summarized features of simulation as a learning strategy, as it encompasses creating a hypothetical opportunity, authentic representation, active participation, integration, repetition, evaluation, and reflection. As a result, it promotes active learning, creative thinking, and high-level problem solving that can produce the capability of independent work among students. In contrast with this, the use of simulation also has disadvantages such as high cost, the need for staff development to manipulate the performance, limited time for training of faculty, and some chance of false transfer due to wrong adjustment of simulators. Again, higher psychological preparation of students is needed since most of the simulation activities cause students to be anxious and frustrated.

Some of the driving forces for current attention for simulation-based teaching are the patient bill of write, a greater need for high competency, and the changing trend of teaching approach from passive to experiential learning. Besides, a professional obligation to keep patient safety, difficulties to find clinical sites, and the greater need to provide high-quality clinical practice also influenced the current trends of teaching.

In nursing, there was a lack of high-stake research that can provide strong evidence on the effect of simulation with a well-organized procedure. This indicates the need to conduct more investigations and arrive at a consensus on the issue among nurse experts.

The individual studies reported both negative and positive effects of simulation-based teaching. For example, in medicine, the use of high fidelity (HF) simulation is criticized for causing overconfidence in students that was even hampering their real practice. On the other hand, nursing literature also reported no effect of simulation on knowledge, skill, and confidence. As a result, this analysis aimed to narrow this gap by producing pooled evidence about the effect of simulation-based teaching over skill performance in the nursing profession. Moreover, this study considers the students and clinical nursing staff as a comparison group to ascertain differences, if any, in skill performance.

Simulation has many advantages and effects for learners and as well as the health care industry as a whole. Studies reported that simulation helped the student to acquire knowledge, skill, and confidence in actual patient-based care.
2.8. Studies

Both non-randomized (quasi) and randomized original trail studies were included in the review and analysis.

2.9. Study selection

At first instance, literature were retrieved from original sources and merged using the software package EndNote X8 (reference management software) and an Excel sheet. Thereafter, the duplicate records were removed. Titles and abstracts were used for primary screening; then, the full text was used if needed. The two authors independently screened each study according to the inclusion criteria. Studies were included if they: (1) include undergraduate nursing students and/or clinical nursing staffs, (2) measure the effect of simulation-based teaching using various types of simulators, (3) use skill performance score as the primary outcome, (4) are randomized controlled trials (RCTs) or non-RCTs (quasi), and (5) produce sufficient data for calculation of sizes of effect. At the same time, the following criteria were used to exclude specific studies from the review process, including non-nursing, not assess simulation, interprofessional study, not original study, qualitative study, result that was not readily used as the report of median and different study populations.

2.10. Data collection process

The two review authors (AA and NA) independently extracted the data using an Excel sheet for a one-page summary. Accordingly, the information about the general overview of the article, the study design, country, population, sample size, intervention, the comparison, duration of the simulation, the outcome, and the methodological quality by JBI score checklist was filled over the pre-defined Excel sheet.

2.11. Risk of bias across studies

The risk of bias was assessed using the Cochrane Collaboration’s Risk of Bias Tool for RCTs. This tool has 6 areas to assess experimental study and the authors decide to use the tool without modifications. Each study was scored (1) for a high risk of bias, (2) for the unclear statements about specific areas of bias, and (3) for low risk of bias. The non-randomized trials were evaluated against the Risk of Bias Assessment tool for Non-randomized Studies (Robins-I). Robins-I have 5 domains to be scored for individual studies. They are (1) bias arising from the randomization process, (2) bias due to deviation from intended interventions, (3) bias due to missing outcome data, (4) bias in the measurement of the outcomes, and (5) bias in the selection of reported result. Each domain is expected to report scores of low, high, or concern.

The quality of the included studies was also done using JBI critical appraisal checklist. The tool was used to judge a study over 9 areas and researchers used 4 phrases with justification: Yes, No, Unclear, and Not applicable. Additionally, publication bias was tested by Trim and Fill methods to assess the effect of publication bias on effect size.

2.12. Summary measures

The composite score of skill performance reflects an overall aggregate score derived from various tools designed by the original researcher or adopted that were used to assess skill ability or performance before and after the experiment. The tools were varied in terms of their type, content, and number of points included in rubrics or checklists.

2.13. Synthesis of results

The analysis was performed by comprehensive meta-analysis version 2 (CMA) software. The quantitative description of pooled analysis was planned. The final discussion of pooled results is dictated by the level of heterogeneity obtained. Then subsequent subgroup analysis was done for the type of study groups, level of fidelity, study regions, types of participants, and types of outcome variables. The heterogeneity was assessed using the Cochran $\chi^2$ test (Q-test) with the alpha level of significance set at 0.10. The degree of heterogeneity was also estimated and interpreted using the $I^2$ statistic Cochrane Handbook for Systematic Reviews of Interventions recommendations with the alpha level of significance set at 0.10, which describes the percentage of total variation across studies that result from heterogeneity rather than chance. Finally, based on the final level of heterogeneity, pooled estimate was reported, discussed, and generalized to the group based on the significance level. The rest of the individual studies were included in a systematic review to avoid misleading readers.

The final size of effect was estimated and reported using a computed random standard deviation (SD) of mean difference ($d$) with a respective confidence interval (CI). This estimate is appropriate for effect size computed from a different study with different measurement context of outcome variables.
2.14. Risk of bias across studies

Assessment of quality of studies and risk of bias at study level was done by JBI and Cochrane checklist. Overall publication bias was tested by using Trim and Fill methods, which have a higher level of sensitivity to assess the effect of publication bias on effect size.\(^\text{18}\)

2.15. Patient and public involvement

This review had no contact with patients. All information was obtained from published studies and electronic databases.

3. Results

3.1. Study selection

Initially, 638 records were identified from 3 sources Cochrane, namely, (CINAHL), PubMed, and Google scholar. Then, 40 duplicated articles were removed using EndNote X8 citation manager\(^\text{19}\) and an Excel sheet. Then, 502 were removed due to focus on other issues (\(n = 78\)), non-nursing study (\(n = 96\)), out of date (\(n = 5\)), not assess simulation (\(n = 287\)), interprofessional study (\(n = 16\)), literature review (\(n = 15\)), and qualitative study (\(n = 5\)). From 96 studies, another 72 studies were removed because of results that were not ready for use (\(n = 9\)), not intended outcome (\(n = 24\)), populations are patients (\(n = 11\)), unclear interventions (\(n = 5\)), out of date (\(n = 7\)), and non-nursing study (\(n = 16\)). Twenty-four studies were used for the final analysis (Figure 1).

3.2. Study characteristics

The included studies varied in terms of their design, the population used, and duration of simulation, type of test used to evaluate outcome variable, type of interventions, learning theory used, and level of fidelity in the simulator.

Figure 1. Flow diagram showing the process of study identification and selection.
Totally 2209 study subjects participated in 24 original studies with a maximum of 367 to a minimum of 30 sample size. The proportion of studies that involved clinical nursing staff amounted to 13.4%, while the rest comprised undergraduate nursing students (86.6%). A large proportion of individual studies came from Turkey (33.3%) followed by the USA (29%), in which both constituted more than half of all studies. Moreover, more than three-fourth of the studies were quasi-experimental (n = 20; 83.3%), (29%) used HF, (29%) also used virtual simulators (VSs), and (58.3%) used both control and experimental group (double group). The total duration spent for simulation intervention ranged from a maximum of 24 h to a minimum of 20 min. The simulation duration was not clearly mentioned in 3 studies (24–26) (Table 1).

The control group was mostly taking the conventional or lecture method of teaching as a comparator or no intervention. The dominant scenario used by individual researchers was acute cases: mainly cardiopulmonary cases (41.6%). The second most common cases were drug dose calculation (8.3%), proper drug administration (8.3%), and securing peripheral intravenous line catheter and phlebotomy (8.3%) (Table 1).

To measure the effectiveness of the intervention, 12 (50%) used direct observation of skill performance using a checklist, 6 (25%) reported the use of OSCE, 4 studies (16.6%) used self-assessment of skill performance improvement, and 1 (4.2%) reported a rating of documents. In 3 studies the skill performance evaluation was assisted by VSs. Of this, virtual computer-guided performance was used in 1 (4.2%), 4 (16.7%) used self-assessment, and another one (4.2%) used direct actual patient-based performance evaluations (Table 1).

### 3.3. Types of studies

The majority (n = 20; 83.3%) of included studies were quasi-experimental. The rest (n = 4; 16.7%) were RCT (Table 1).

### 3.4. Type of scenario

Different type of scenarios were used for simulation activity in all studies. Almost half of the scenarios were having the nature of acute cases, such as CPR, resuscitation, arrhythmia, deteriorating patient, pre-post case, and shock. The remaining scenarios were non-acute or cold cases such as medication administration, phlebotomy, diabetes mellitus (DM), and communication skills.

### 3.5. Quality of individual studies

The risk of bias in included studies ranged from unclear to high due to issues with 6 areas of risk of bias assessments for RCTs. These are random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective reporting. From 7 studies, it could be ascertained that 5 of them scored moderate risk of bias while the rest were high risk of bias. Using Robins-I tool for non RCT, we discovered that 6 studies scored with no risk of bias, 7 with low risk of bias, 3 with moderate risk bias, and 1 with serious risk of bias. Moreover, from the total of 24 included studies, only 4 (6.7%) studies were categorized as high-quality research, 2 (8.3%) as low-quality research, and the remaining 18 (75%) studies ranked as medium-level quality studies.

In most of the studies, quality issues were related to lack of control group, unclear outcome measurements, and failure to clearly state what treatment was given for study groups.

### 3.6. Meta-analysis

#### 3.6.1. Result of individual studies

Even though individual studies reported additional outcomes as primary and/or secondary objective to their studies, this review considers and takes only the outcome related to skill performance. From a total of 24 studies, 20 reported positive effects of simulation-based teaching, while the rest reported a lack of evidence to support the positive effects of simulation-based teaching.

Simulation-based teaching improves skill performance among the experimental group with an overall random effect size of d = 1.01, 95% CI [0.69–1.33], Z = 6.18, P < 0.01. From this, it is understood that >79% of control group skill performance is below experimental group skill performance. But it is uncertain to conclude this finding because significant heterogeneity (I² = 93.9%) was observed during analysis.

The random effect size (d) for individual studies dispersed to small (d ≤ 2, n = 5, 20.8%), medium (d = 0.2–0.5, n = 4, 16.7%), and large (d ≥ 0.8 and above, n = 15, 62.5%). Moreover, 5 studies (33–35, 37) effect size showed statistically insignificant results during analysis (Figure 2; forest plot). This meta-analysis result is consistent with the original report by individual articles about the effect of simulation on skill performance.

Initially, 4 individual studies (33, 35, 36, 37) already reported that simulation has no statistically significant change
<table>
<thead>
<tr>
<th>Study</th>
<th>Interventions</th>
<th>Study type, duration, sample size</th>
<th>Scenario</th>
<th>Outcome measures</th>
<th>Result</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aqel &amp; Ahmed 2014, Jordan, RCT</td>
<td>Training of participant over simulated case with cardiac arrest scenario and debriefing discussion.</td>
<td>HFS, 25390</td>
<td>CPR</td>
<td>Direct observation using Checklist: mock codes were conducted over manikin over floor and evaluation using AHA checklist.</td>
<td>The results revealed the existence of a significant difference in the post-test CPR knowledge as well as the CPR skills in favor of participants in the intervention group.</td>
</tr>
<tr>
<td>2</td>
<td>Basak et al., 2016, Turkey, Quasi, Single pre-post</td>
<td>45 min paper-based drug dose calculation simulation and debriefing session for discussion.</td>
<td>LFS, 4582</td>
<td>Actual physician prescription</td>
<td>Rating: Drug dose calculation was evaluated from 100 points immediately after training and 1 month later.</td>
<td>The difference between the mean pre-test score and the mean post-test score was statistically significant ($t = 8.767$, df = 89, $P = 0.001$)</td>
</tr>
<tr>
<td>3</td>
<td>Basak et al., 2019, Turkey, RCT, equivalent control group</td>
<td>20 min simulation with 40 min debriefing and self-evaluation for 10 min generally 80 min discussion about teaching skill over SPs.</td>
<td>SP 8071</td>
<td>Inhaler drug administration</td>
<td>Direct Observation using Check list: Teaching skill measured by checklist consisted of 15 procedural steps developed and tested by principal investigators.</td>
<td>Total patient teaching skill score for control group was $26.73 \pm 5.63$ and $39.08 \pm 5.49$ for SP group which causes a statistically significant difference ($P \leq 0.01$)</td>
</tr>
<tr>
<td>4</td>
<td>Boyles et al., 2015, USA, Quasi Single pre-post</td>
<td>Interactive e simulation clinical scenario with video recording patient conditions, pop-up task, and respective response.</td>
<td>VS, 24367</td>
<td>Cardiac, shock, and respiratory</td>
<td>Virtual skill performance</td>
<td>A paired t-test showed a significant improvement in performance between the first and last scenarios ($t = -8.037$, df = 366, CI 2.05–1.24; $P = 0.00$).</td>
</tr>
<tr>
<td>5</td>
<td>Bowling et al., 2015, USA, Quasi, equivalent control group</td>
<td>50 min respiratory distress simulation case and participant required to react to simulated case.</td>
<td>MFS, 5073</td>
<td>Respiratory distress</td>
<td>OSCE with six station lasting 7 min and rater-based evaluations</td>
<td>There was a significant difference for both groups in knowledge and skill performance (measured with a mini OSCE), but not between the groups</td>
</tr>
<tr>
<td>6</td>
<td>Boyde M et al., 2018, Australia, Quasi Single pre-post</td>
<td>Innovative teaching of emergency management of patient using HF simulation with Jefferies simulation principles.</td>
<td>HFS, Not mentioned, 50</td>
<td>Emergency patient</td>
<td>Self-assessment: The self-efficacy in clinical performance scale was used to measure participant’s assessment and handover practice.</td>
<td>The mean change in handover skill from $7.88 \pm 1.76$ to $8.79 \pm 1.22$ was statistically significant with $t (41) = 3.41$, $P &lt; 0.01$</td>
</tr>
<tr>
<td>7</td>
<td>Chen et al., 2015, Canada, Quasi, equivalent control group</td>
<td>Auscultation skills training using low and HF training.</td>
<td>HFS, 4054</td>
<td>Pneumothorax and a systolic murmur: Auscultation skills</td>
<td>OSCE using Check list: Participants required to correctly identify 20 different sounds on simulators.</td>
<td>There was no evidence that the HF group performed better than the LFS group in clinical skills or in auscultation sounds recognition on HFS.</td>
</tr>
<tr>
<td>8</td>
<td>Durmaz et al., 2012, Turkey, RCT</td>
<td>Intervention: Participants receive 4 h computer-based education simulation about pre-operative and post-operative patient management.</td>
<td>VS, 4 h,82</td>
<td>Pre-post case</td>
<td>OSCE for pre and post-operative management and deep breathing and coughing exercise.</td>
<td>There was not a significant difference between the students’ post-education practical deep breathing and coughing exercise education skills ($P = 0.867$).</td>
</tr>
<tr>
<td>Study</td>
<td>Interventions</td>
<td>Study type, duration, sample size</td>
<td>Scenario</td>
<td>Outcome measures</td>
<td>Result</td>
<td>Effects</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>---------------------------------</td>
<td>---------</td>
<td>-----------------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>9</td>
<td>IV training over virtual IV simulator</td>
<td>Ismailoglu et al., 2018, Turkey, Quasi, equivalent control group</td>
<td>VS, Not clear, 62</td>
<td>Encoded case</td>
<td>Mean psychomotor skills score of the experimental group 45.18 (33.73 ± 4.22) was higher than that of the control group 20.44 (26.53 ± 4.45) with Z = 5.294, ( P = 0.000 ).</td>
<td>Improved</td>
</tr>
<tr>
<td>10</td>
<td>Abdominal examination skill was tested after teaching students using SP for 45 min.</td>
<td>Jaberi et al., 2019, Iran, RCT</td>
<td>SP, 45.87</td>
<td>Physical examination of abdomen</td>
<td>The mean score in intervention group changed from 5.35 ± 1.77 to 15.39 ± 3.2, while it was changed 4.98 ± 2.17 to 14.43 ± 3.93 in control group. There was a significant difference between the mean pre-test and post-test scores in each group (( P &lt; 0.05 )).</td>
<td>Improved</td>
</tr>
<tr>
<td>11</td>
<td>A 12 h theory and laboratory-based training using SP on selected fundamental of nursing skills.</td>
<td>Karabacak et al., 2019, Turkey, Quasi, Single pre-post</td>
<td>SP, 12 h, 65</td>
<td>Fundamental of nursing issues</td>
<td>No significant difference has been found between pre-scenario (7.05 ± 9.17) and post-scenario (5.89 ± 2.02) scores about self-assessment of safe patient transfer (( t = 1.01; \ P = 0.32 )).</td>
<td>No change</td>
</tr>
<tr>
<td>12</td>
<td>Virtual based 3 h training to improve/decrease IV reinsertion</td>
<td>Keleekai et al., 2016, USA, RCT, equivalent control group</td>
<td>VS, 3 h, 58</td>
<td>Peripheral IV securing</td>
<td>The intervention was effective and resulted in several statistically significant improvements in knowledge, confidence, and skills both within and between study groups over time.</td>
<td>Improved</td>
</tr>
<tr>
<td>13</td>
<td>Integrating simulation-based teaching over advanced acute care adult scenario on shock, resuscitations for 90 min.</td>
<td>Lee et al., 2019, Taiwan, China, Quasi, equivalent control group</td>
<td>HFS, 9052</td>
<td>Shock and resuscitations</td>
<td>No significant difference in clinical performance was observed among groups.</td>
<td>No change</td>
</tr>
<tr>
<td>14</td>
<td>The interactive web-based programmer 3 h training on patient identification, early recognition, vital sign monitoring, and management.</td>
<td>Liaw et al., 2015 Singapore, RCT, equivalent control group</td>
<td>VS, 3 h, 67</td>
<td>Deteriorating patients</td>
<td>There was a significant change in Assessing and managing clinical deterioration in experimental group pre-test 18.17 (3.55), post-test 25.83 (4.79), and Reporting clinical deterioration pre-test 10.09 (2.31) post-test 12.83 (2.41).</td>
<td>Improved</td>
</tr>
<tr>
<td>Study</td>
<td>Interventions</td>
<td>Study type, duration, sample size</td>
<td>Scenario</td>
<td>Outcome measures</td>
<td>Result</td>
<td>Effects</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>---------------------------------</td>
<td>----------</td>
<td>-----------------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>15</td>
<td>Lubbers et al., 2016, USA, Quasi, Single pre-post</td>
<td>1 h simulation, pre-post-simulation discussion.</td>
<td>HFS, 3 h and 3058</td>
<td>Not mentioned</td>
<td>Self-assessment of Knowledge, confidence and performance.</td>
<td>The Skill score, revealed significant increases from pre-test 2.25 to post-test 4.13, t = 21.21, P &lt; 0.001.</td>
</tr>
<tr>
<td>16</td>
<td>Meyer et al., 2011, USA, Quasi, equivalent control group</td>
<td>Replacing 2 weeks (25%) of clinical work or rotation with simulation-based teaching in skill lab.</td>
<td>VS, 24 h, 120</td>
<td>Various</td>
<td>Direct observation using rating scale Clinical faculty assessment of student performance in clinical work and compared with control group who spent 100% in clinical rotations.</td>
<td>Faculty rated students with patient simulation experience higher than those who had not yet attended simulation mean 1.74 (0.75), P = 0.02.</td>
</tr>
<tr>
<td>17</td>
<td>Morton et al., 2019, USA, Quasi, Single pre-post</td>
<td>Training using HFS portraying a patient with cardiac arrest.</td>
<td>HFS, Not mentioned, 37</td>
<td>CPR</td>
<td>Direct observation using Check list: Mock Code Evaluation Tool basically developed based on AHA (2015) guideline for basic life supports.</td>
<td>There is no statistically significant difference in performance obtained following simulation-based training.</td>
</tr>
<tr>
<td>18</td>
<td>Sarmasoge et al. 2016, Turkey, Quasi, equivalent control group</td>
<td>SP-based training of Arterial blood pressure and Subcutaneous injection, feedback, and discussion with SP.</td>
<td>SP, 4 h, 77</td>
<td>Hypertension and acute pain</td>
<td>Direct observation using Check list: Performance assessment using check list for arterial blood pressure measurements and subcutaneous injection by two raters.</td>
<td>The mean performance score for the measurement of arterial blood pressure was 76 ± 7.6 for the control group and 83 ± 3.1 for the experimental group (P &lt; 0.001). However, no significant difference was found between the groups’ performance scores on subcutaneous injection administration.</td>
</tr>
<tr>
<td>19</td>
<td>Stayt LC, et al., 2015, UK, RCT</td>
<td>2 h clinical skill teaching; systematic ABCDE assessment and management process on medium fidelity patient simulator (ALS Simulator, made by Laerdal Medical) using a clinical scenario of an acutely unwell patient who is exhibiting signs of clinical deterioration.</td>
<td>SP, 2 h, 98</td>
<td>Deteriorating patient</td>
<td>OSCE using check list. The OSCE comprised of a check list of 24 objective performance criteria that evaluated participants performance of assessing and managing a deteriorating patient using a patient simulator.</td>
<td>The results indicate that students who received simulation training performed a systematic ABCDE assessment and managed the deteriorating patient more effectively than those who received a didactic teaching approach.</td>
</tr>
<tr>
<td>20</td>
<td>Sumner et al., 2012, USA, Quasi, Single pre-post</td>
<td>Participants received the intervention by attending a 4-hour basic arrhythmia program on the second day of nursing orientation.</td>
<td>MFS, 4 h, 138</td>
<td>Arrhythmia cases</td>
<td>Self-assessment: post simulation self-report of caring and resource utilization in caring of patient with arrhythmias patients.</td>
<td>Following simulation there was transfer of knowledge to clinical practice.</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Interventions</th>
<th>Study type, duration, sample size</th>
<th>Scenario</th>
<th>Outcome measures</th>
<th>Result</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Toubasi et al., 2015, Jordan,†† Quasi, Single pre-post</td>
<td>Step by step simulation and debriefing of cardiac arrest scenario using AHA guidelines.</td>
<td>MFS, 8 h, 30</td>
<td>Cardiac arrest</td>
<td>Direct observation using Check list: Validated skill scenario testing tool which was developed by the AHA to assess performance according to the AHA 2010 guidelines.</td>
<td>There is a significant mean difference of 2.9 in overall skill performance and BLS score after simulation (t = 7.4, df = 29, P &lt; 0.01).</td>
</tr>
<tr>
<td>22</td>
<td>Unver et al., 2013, Turkey,‡‡ Quasi, Single pre-post</td>
<td>4 h training using SP</td>
<td>SR 4 h, 85</td>
<td>Medical administration</td>
<td>OSCE: OCEF were used.</td>
<td>There was a significant difference (30.26) in pre-test (24.02 ± 16.06) to post-test (54.28 ± 14.54) skill performance measurements (P &lt; 0.01; t = 14.35).</td>
</tr>
<tr>
<td>23</td>
<td>Vidal VL et al., 2013, Turkey,§§ Quasi, equivalent control group</td>
<td>Computer-based training with demonstration, return demonstration and verbal feedback regarding performance of phlebotomy.</td>
<td>VS, 3 h, 73</td>
<td>Phlebotomy</td>
<td>Direct observations using Check list: the skill checklist used by the mentors consisted of 21 items addressing the necessary steps for the completion of a phlebotomy procedure and 3 items related to overall performance.</td>
<td>There is significant among the group in mean skill performance score in pain factor (P = 0.006), hematoma formation (P = 0.000), and number of reinsertions (P = 0.000).</td>
</tr>
<tr>
<td>24</td>
<td>Woda et al., 2019, USA,†† Quasi, Single pre-post</td>
<td>A 20 min training using HFS and debriefing about care of patient with type I DM.</td>
<td>HFS, 20:233</td>
<td>Type one DM</td>
<td>Direct observation of using Check list: Performance evaluated using 10 item evaluation rubrics by research assistance on major areas of DM care.</td>
<td>Simulation did have a significant positive effect on performance change scores (P &lt; 0.001; r = 0.28). The mean pre-test score on performance items was 0.73 (SD = 0.14), and the mean post-test score on performance items was 0.76 (SD = 0.12).</td>
</tr>
</tbody>
</table>

Note: †: Minute; ABCDE, airway, breathing, circulation, disability, exposure; AHA, American Health Association; CI, confidence interval; CPR, cardiopulmonary resuscitation; df, degree of freedom; DM, diabetes mellitus; HF, high fidelity; HFS, high fidelity simulator; LFS, low fidelity simulator; MFS, medium fidelity simulator; OCEF, objectively constructed evaluation form; OSCE, objective structured clinical examination; RCT, randomized controlled clinical trial; RAPIDS, rescuing A patient in deteriorating situations; SD, standard deviation; SP, standard patient; t, t-distribution statistics; VS, virtual simulator.

Table 1. Characteristics of included studies.
over a participant’s skill performance. At the same time, the meta-analysis also confirmed this by reporting a statistically insignificant effect size for those studies. (Figure 3; forest plot).

### Effect of Simulation based teaching

#### 3.6.2. Subgroup analysis

Because of overall significant heterogeneity ($I^2 = 93.9\%$), subgroup analysis with moderator variables were done.

#### Sensativity Analysis with one study remove

![Forest plot showing sensitivity analysis by one study remove method.](image)
with types of study design, type of participants, study regions, and simulation fidelity. The heterogeneity level was maintained high despite variation in the effect size across the moderator variable analysis.

### 3.6.2.1. Effect of simulator type

Five types of the simulation were considered for this analysis. Except for medium fidelity simulator (MFS), all of the simulation types scored large effect size favoring the skill performance score among the experimental group. But only the low fidelity simulator (LFS) obtained a larger and statistically significant effect size with an acceptable level of heterogeneity $d = 0.89$ (CI [0.24, 2.29], $P = 0.02$, $I^2 0\%$). This group of studies involved study participants. We are confident that using LFS improved the skill performance of the experimental group (Table 2).

### 3.6.2.2. Types of group

The effect of group type used for the individual study was tested for all studies as subgroup analysis, which was tested as to whether individual studies used single pre-post or double group pre-post design. The single group pre-post users score large effect size $d = 1.02$ (CI [0.52, 1.50], $P < 0.01$). Again, the double group also score almost similar effect size $d = 1.00$ (CI [0.56, 1.44], $P < 0.01$). In both cases, significant heterogeneity was observed. So, it is understood there is no effect on size, whether we have used a single group or double group for the experiments (Table 2).

### 3.6.2.3. Type of study participants

Only 3 studies involved clinical nursing staff as study participants. The effect size for clinical nursing staff was $d = 1.08$ (CI [0.43, 1.74], $P < 0.01$, $I^2 85.8\%$). The almost similar effect size was observed for nursing students $d = 0.98$ (CI [0.61, 1.37], $P < 0.01$, $I^2 95\%$). Here also, we have no confidence to discuss the pooled analysis due to significant heterogeneity observed during analysis. But it is visible that the effect size was almost similar and statistically significant (Table 2).

<table>
<thead>
<tr>
<th>Comparison and Groups</th>
<th>Numbers of studies</th>
<th>Effect size (d)</th>
<th>I², %</th>
<th>Z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies Groups</td>
<td>24</td>
<td>1.01 (CI [0.62, 1.41], $P &lt; 0.01$)</td>
<td>93.9</td>
<td>5.13</td>
</tr>
<tr>
<td>Single group</td>
<td>10</td>
<td>1.02 (CI [0.52, 1.50], $P &lt; 0.01$)</td>
<td>95</td>
<td>4.46</td>
</tr>
<tr>
<td>Double groups</td>
<td>14</td>
<td>1.00 (CI [0.56, 1.44], $P &lt; 0.01$)</td>
<td>92.9</td>
<td>4.48</td>
</tr>
<tr>
<td>Simulator types</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF</td>
<td>7</td>
<td>1.23 (CI [0.55, 1.93], $P &lt; 0.01$)</td>
<td>94.8</td>
<td>3.5</td>
</tr>
<tr>
<td>Medium fidelity</td>
<td>3</td>
<td>0.89 (CI [−0.14, 1.93], $P = 0.09$)</td>
<td>86.5</td>
<td>1.69</td>
</tr>
<tr>
<td>LF</td>
<td>3</td>
<td>1.27 (CI [0.24, 2.29], $P = 0.02$)</td>
<td>0</td>
<td>2.4</td>
</tr>
<tr>
<td>SP</td>
<td>5</td>
<td>1.03 (CI [0.23, 1.84], $P = 0.01$)</td>
<td>96</td>
<td>2.5</td>
</tr>
<tr>
<td>VSs</td>
<td>6</td>
<td>0.69 (CI [−0.04, 1.4], $P = 0.06$)</td>
<td>95.4</td>
<td>1.85</td>
</tr>
<tr>
<td>Types of participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical staffs</td>
<td>3</td>
<td>1.08 (CI [0.43, 1.74], $P &lt; 0.01$)</td>
<td>85.8</td>
<td>3.25</td>
</tr>
<tr>
<td>Nursing students</td>
<td>8</td>
<td>0.98 (CI [0.61, 1.37], $P &lt; 0.01$)</td>
<td>95</td>
<td>5.11</td>
</tr>
<tr>
<td>Regions (country)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>America</td>
<td>8</td>
<td>1.22 (CI [0.62, 1.82], $P &lt; 0.01$)</td>
<td>94.6</td>
<td>4.02</td>
</tr>
<tr>
<td>Europe</td>
<td>10</td>
<td>0.76 (CI [0.24, 1.29], $P = 0.004$)</td>
<td>95.3</td>
<td>2.85</td>
</tr>
<tr>
<td>Middle East</td>
<td>6</td>
<td>1.17 (CI [0.48, 1.86], $P = 0.001$)</td>
<td>88.74</td>
<td>3.34</td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quasi</td>
<td>17</td>
<td>0.96 (CI [0.57, 1.34], $P &lt; 0.01$)</td>
<td>94.78</td>
<td>4.86</td>
</tr>
<tr>
<td>RCT</td>
<td>7</td>
<td>1.14 (CI [0.54, 1.75], $P &lt; 0.01$)</td>
<td>91.1</td>
<td>3.7</td>
</tr>
<tr>
<td>Types of scenarios</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute</td>
<td>12</td>
<td>1.07 (CI [0.73, 1.41], $P &lt; 0.01$)</td>
<td>88.1</td>
<td>6.18</td>
</tr>
<tr>
<td>Cold</td>
<td>12</td>
<td>0.92 (CI [0.35, 1.49], $P &lt; 0.02$)</td>
<td>95.16</td>
<td>3.16</td>
</tr>
</tbody>
</table>

Note: CI, confidence interval; HF, high fidelity; LF, low fidelity; RCT, randomized controlled clinical trial; SP, standard patient; VSs, virtual simulators.

Table 2. Summary of effect size for subgroup analysis.
3.6.2.4. Study design

There is no difference in whether RCT or quasi-experimental design was used to evaluate the effect of simulation on skill performance. The skill performance score was increased among study experimental group participants. The effect size for 7 RCTs was \( d = 1.14 \) (CI [0.54, 1.75], \( P < 0.01 \)) and for the rest of quasi-experimental was 0.96 (CI [0.57, 1.34], \( P < 0.01 \)). In both cases, considerable heterogeneity precludes us from drawing a conclusion and recommending the result (Table 2).

3.6.2.5. Types of scenario

Another comparison was done to ascertain whether nursing skill performance was different due to the use of categories of scenarios. The scenarios were categorized as acute and cold cases. The effect size for both groups of scenarios was similar and considerable heterogeneity was observed in both cases. Thus, we can conclude that in the current study, types of scenarios used for simulation have no effect on nursing skill performance (Table 2).

3.6.3. Sensitivity analysis

The pooled effect size was tested for a possible change by one study remove method. Accordingly, there is no large change over the overall effect size due to the removal of individual studies one by one.

The maximum change was observed (\( d = 1.11 \)) when Stayt et al.\(^\text{41}\) was removed from the analysis. Further, the minimum effect size (\( d = 0.97 \)) was also obtained when Jaberi and Momennasab\(^\text{34}\) was removed from the analysis. The overall variation was \( d = 0.13 \). Thus, it is understood that the removal of 1 study has no significant effect on overall effect size (Figure 2).

3.6.4. Risk of bias

The risk of publication bias was tested using 4 common methods. Except for Egger’s regression (intercept = 2.61, \( P = 0.08 \)), the Trim and Fill methods (\( d = 0.62, [0.28, 0.96] \)), classic Fail-safe N, and the Begg and Mazumdar (\( b = 0.35, P = 0.01 \)), all confirm the presence of publication bias under the random-effects model. The point estimate and 95% CI for the combined studies is 1.01035 (0.69, 1.33). Using Trim and Fill, the imputed point estimate is 0.62 (0.28, 0.95) (Figure 4).

4. Discussions

This review and meta-analysis were intended to present the result of the review, and produce a pooled estimate regarding the effect of simulation-based teaching on nursing skill performance in nursing. Most of the

![Funnel Plot of Standard Error by Std diff in means](image)

Figure 4. Funnel plot showing publication bias among included studies.
studies were from developed and middle-level countries, and original researches were varied in terms of the study context such as the types of the scenario used, the number of study participants, the duration of the simulation, and tools to measure outcomes. Moreover, the pooled estimate of included studies did prove the positive effect of simulation-based teaching in improving nursing skill performance. Since significant heterogeneity was observed during analysis, the reader needs to use the pooled analysis result with caution. The agreement among specific studies on the simulation was not complete. Some studies still reporting inconsequentially of simulation-based teaching got improving skill performance in nursing. This gives an assignment for researchers to answer why, and users to continuously assess their success after the implementation of the simulation.

The simulation-based teaching helps learners or users to assume the complexity of health service delivery and allow repeated exercise. Moreover, participation in simulation decrease mistakes in actual practice and increases flexibility during practice. In the current review, regardless of simulation types, the effect of simulation over skill performance showed a larger effect size that favors the users, which is consistent with a systematic review done by others. This indicates a need for further evidence and searching for potential factors significantly affecting the success and failure of this teaching strategy. Another factor may be the level of information contamination among controls and experimental groups. A significant number of specific studies were not strict on blinding participants and evaluators of performance.

This review and meta-analysis obtained significant heterogeneity in the overall and moderator analysis. Even though the sizes of effect were statistically significant, we lack the confidence to recommend this effect size due to large heterogeneity. Moreover, this might be due to a combination of studies with different scenarios, designs, and assessment tools. As a result, further work is expected from the nurse researchers to justify its effect confidently in a well-organized and standardized manner.

The larger proportion of studies was drawn from the developed and middle-level countries. Similar results were also reported consistently in various reviews and meta-analyses. This might be associated with a lack of financial support, simulation facility, and motivation on the part of the researchers to handle experimental studies that are accompanied by strict procedures.

We may think that high fidelity simulators (HFS) are better than LFS, but the current review shows the opposite. The estimated effect size for LF was higher for LF with an acceptable range of heterogeneity. Even in medicine, the students prefer LF, focused, and shorter duration of the simulation. Again, Massoth et al., reported in 2019 that LFS helped to improve skill performance as compared to HFS, and the HF was criticized for letting the students have overconfidence. Another RCT reported HF that had no effect on students’ retention of neonatal resuscitations.

The students’ preference as well as larger effect size in LFS-based teaching may be associated with the extent of time spent in simulation and mental adjustment of students for the simulation environment. It tends to happen that that students spent more time in LF. Moreover, the level of anxiety at the time of teaching in LF may also favor learning. Another justification may be the distracting nature of HFS from basic concept learning by increasing extraneous cognitive load; this was also given as the reason for impaired learning in HF simulation room.

In contrast with the current study, many reviews of original studies showed a higher advantage of HFS than LFS in neonatal resuscitation, identification, and management of deteriorating patient, and performance of basic life support. As a controversial finding, having different types of fidelity levels has not shown a significant difference in student skill performance in all types of simulation. This result indicated not to depend on the level of fidelity and has rather resulted in the revelation that use of the mixed method may be more advisable. Again, it helps us to conclude that focused training, student handling, and duration of simulation matters more than types of fidelity used. Thus, the upcoming research needs to identify and address the factors that determine success in using simulator other than changing fidelity.

The use of standardized patients is preferred for the noninvasive procedure and skills, such as physical examination, history taking, communication exercise, and improvement of confidence for clinical skill management. This review also identified the use of standardized patients as a simulator improves the skill performance of participants with large effect sizes. Similar results were reported from different reviews. Oh et al. (2015), show that the use of standardized patients improves communication skills with large effect size.

5. Conclusions

Assisting teaching with simulation did improve nursing skill performance. Again, the use of simulation-based teaching showed a positive effect both for student and
Effect of simulation-based teaching on skill performance

clinical nursing staff training. The level of fidelity showed little difference and even LFS produced a greater effect size than others. Along with investing in equipment and teaching aid, equal attention should be given to faculty development to improve the style of teaching, student handling, and facilitation of teaching sessions. Since most studies were done in simulated environments, their application and significance for actual patient care need to be proved with further research.

Strength and limitations

Analysis of single outcome of simulation-based teaching aid is understood to cause focused result and implication. Moreover, focusing on the most important aspect of nursing education (skill) also helps to inform the most important aspect of nursing.

The confidence in generalizability and overall recommendation is limited by significant heterogeneity in the pooled analysis. Variety and difference in the type of scenario and outcome measuring tool were the major challenges of these combined studies.

The scope of the literature search was narrow due to the subscription challenge, which might reduce the depth of the literature search. Bias may also be introduced during searching, screening, and selecting literature, which directly affect the pool of literature for the final analysis. The number and quality of included and excluded literature were dependent on the critical appraisal ability of researchers. Again, this review was not specific and it considers every study that assessed a skill performance while they were using a different scenario, and research context that ends up with significant heterogeneity. The true effects of simulation-based teaching may be obscured due to the inclusion of freely available literature.

Author contributions

All authors developed the protocol, interpreted the results, and approved the final version. AA and NA completed the search, screened articles for inclusion, and synthesized the findings. AA and NA extracted data and drafted the manuscript.

Ethical approval

Ethical issues are not involved in this paper.

Conflicts of interest

All contributing authors declare no conflicts of interest.

References


