

# Simulation-based assessments in health professional education: a systematic review

Tayne Ryall<sup>1</sup>  
Belinda K Judd<sup>2,3</sup>  
Christopher J Gordon<sup>3</sup>

<sup>1</sup>Physiotherapy Department, Canberra Hospital, ACT Health, Canberra, ACT,  
<sup>2</sup>Faculty of Health Sciences, <sup>3</sup>Sydney Nursing School, The University of Sydney, Sydney, NSW, Australia

**Introduction:** The use of simulation in health professional education has increased rapidly over the past 2 decades. While simulation has predominantly been used to train health professionals and students for a variety of clinically related situations, there is an increasing trend to use simulation as an assessment tool, especially for the development of technical-based skills required during clinical practice. However, there is a lack of evidence about the effectiveness of using simulation for the assessment of competency. Therefore, the aim of this systematic review was to examine simulation as an assessment tool of technical skills across health professional education.

**Methods:** A systematic review of Cumulative Index to Nursing and Allied Health Literature (CINAHL), Education Resources Information Center (ERIC), Medical Literature Analysis and Retrieval System Online (Medline), and Web of Science databases was used to identify research studies published in English between 2000 and 2015 reporting on measures of validity, reliability, or feasibility of simulation as an assessment tool. The McManis Critical Review for quantitative studies was used to determine methodological value on all full-text reviewed articles. Simulation techniques using human patient simulators, standardized patients, task trainers, and virtual reality were included.

**Results:** A total of 1,064 articles were identified using search criteria, and 67 full-text articles were screened for eligibility. Twenty-one articles were included in the final review. The findings indicated that simulation was more robust when used as an assessment in combination with other assessment tools and when more than one simulation scenario was used. Limitations of the research papers included small participant numbers, poor methodological quality, and predominance of studies from medicine, which preclude any definite conclusions.

**Conclusion:** Simulation has now been embedded across a range of health professional education and it appears that simulation-based assessments can be used effectively. However, the effectiveness as a stand-alone assessment tool requires further research.

**Keywords:** health care, technical skills, competency, students

## Introduction

Assessment, in the most expansive definition, is used to identify appropriate standards and criteria and ascertain quality through judgment.<sup>1</sup> There are a multitude of assessment modes, adopted for various reasons, such as measuring performance or skill acquisition, and these can be used at different stages of the learner's educational trajectory. There has been much debate, however, about the effectiveness of various forms of assessment, such as multiple-choice question examinations, and this has influenced educators' desire to develop assessments that are more realistic

Correspondence: Christopher J Gordon  
M02 – Sydney Nursing School, The  
University of Sydney, Sydney, New South  
Wales 2006, Australia  
Tel +61 2 9351 0586  
Fax +61 2 9351 0615  
Email christopher.gordon@sydney.edu.au

and performance based.<sup>2,3</sup> The types of assessment used in pre- and postregistration health professional education have been widely reviewed.<sup>4-9</sup> A compounding challenge with assessment for health professionals and health students is determining competency of practice. This is a complex but necessary component of education and training. In more recent decades, performance-based assessment practices have gained strong momentum<sup>4</sup> as educators have sought to examine authentic learner performance with the knowledge that these types of assessments are a driving influence on learning and teaching practices. Out of this need for authentic assessment came the adoption of simulation-based assessment.

Simulation, as a technique for both training and assessment, has been used in the aeronautical industry and military fields since the early 1900s, with the first flight simulator being developed in 1929.<sup>10</sup> The complexity and sophistication of simulation improved progressively from the 1950s, driven primarily by the integration of computer-based systems. The translation of simulation into health education has resulted in an almost exponential growth in the use of simulation as an educational tool. Simulation aims to replicate real patients, anatomical regions, or clinical tasks or to mirror real-life situations in clinical settings.<sup>11</sup> The increasing implementation of simulation-based learning and assessment within health education has been driven by training opportunities to practice difficult or infrequent clinical events, limited clinical placement opportunities, increasing competition on clinical educators' time, new diagnostic techniques and treatment, and greater emphasis being placed on patient safety.<sup>11-15</sup> Accordingly, health educators have adopted simulation as a viable educational method to teach and practice a diverse range of clinical and nonclinical skills. Simulation modalities such as standardized patients (SPs), anatomical models, part-task trainers, computerized high-fidelity human patient simulators, and virtual reality are in use within health education.<sup>10,11,16</sup> In particular, these techniques have been used in preregistration health professional training, as simulation allows learners to practice prior to clinical placement and patient contact, maximizing learning opportunities and patient safety.<sup>6,17,18</sup> Simulation provides a safe environment to practice clinical skills in a staged progression of increasing difficulty, appropriate for the level of the learner. Practicing skills on real patients can be difficult, costly, time consuming, and potentially dangerous and unethical.<sup>11,12,14,15</sup> As such, health professional educators have increasingly adopted simulation-based assessment as a viable means of evaluating student and health professional populations. In addition, simulation-based

assessments are a means of creating an authentic assessment, replicating aspects of actual clinical practice.

While there has been widespread acceptance of simulation as an educational training tool, with evidence supporting its use in health education, the effectiveness of simulation-based assessments in evaluating competence and performance remains unclear. With an increasing use of simulation in health education worldwide, it is salient to review the literature related to simulation-based assessments. Therefore, the aim of this systematic literature review was to evaluate the evidence related to the use of simulation as an assessment tool for technical skills within health education.

## Methods

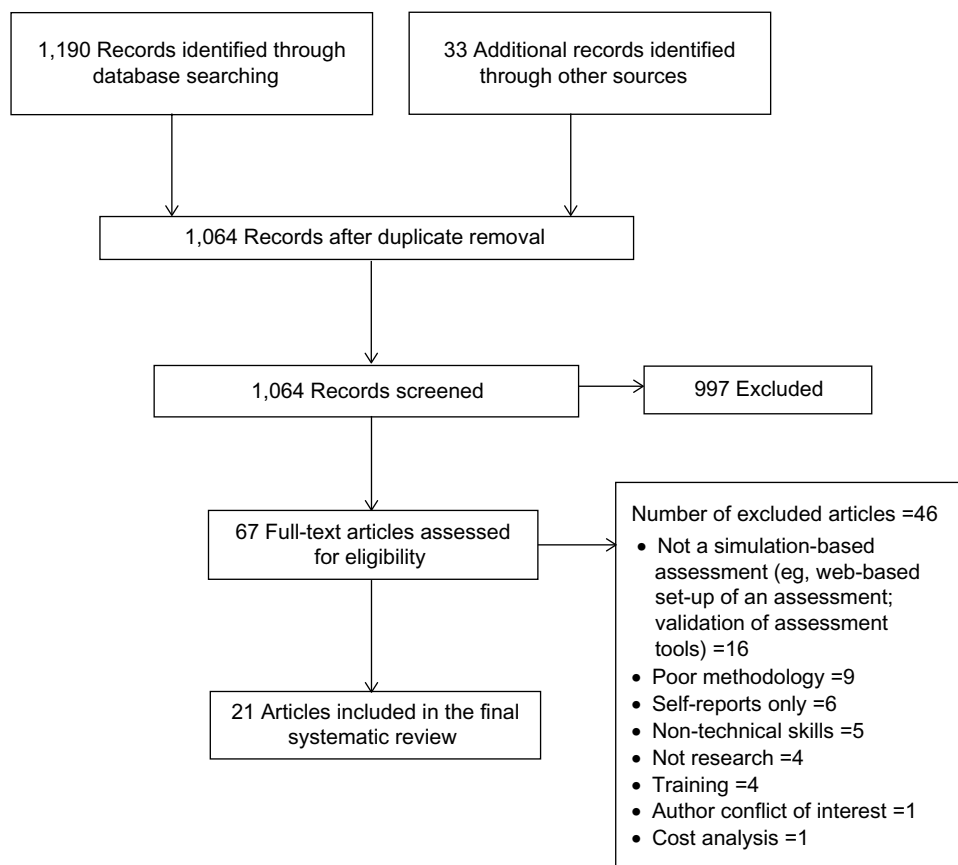
This systematic review was undertaken using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.<sup>19</sup> The review involved searching health and education databases, followed by structured inclusion and exclusion criteria with consensus across reviewers. Two raters (TR, BJ) independently screened all abstracts for eligibility. There was high agreement on the initial screen, and both raters showed excellent interrater reliability (Cohen's kappa =0.91). Any disagreements with article eligibility were reconciled via consensus or referred to a third reviewer (CJG).

## Search databases and terms

Literature was searched in the following key databases: Cumulative Index to Nursing and Allied Health Literature (CINAHL), Education Resources Information Center (ERIC), Medical Literature Analysis and Retrieval System Online (Medline), and Web of Science. The following search terms were included: allied health, medical education, nursing education, assessment, and simulation. This initial search located 1,190 articles, with another 33 located through reference list searches and gray literature (n=1,223). Following removal of duplicates, 1,064 abstracts were screened for eligibility. We reviewed 67 full-text articles for eligibility, with 21 articles chosen for the final systematic review (Figure 1). An adapted critical appraisal tool<sup>20</sup> (Table 1) was used to determine the methodological rigor of the articles.

## Inclusion/exclusion criteria

Our inclusion criteria required all articles to be in English, and the databases were searched for articles published between the years 2000 and 2015. Articles needed to be research based and to have examined simulation as an assessment tool for health professionals or health professional students. Articles incorporating simulation-based assessments that explored



**Figure 1** Flow diagram of search.

technical and nontechnical skills were included. However, those studies that focused solely on nontechnical skills only, such as communication, interpersonal skills, and team work were excluded. The focus of this systematic review was on technical skills and we, therefore, only included studies that examined technical and nontechnical skills in combination. Technical skills were defined as those requiring the participant to complete a physical assessment (or part thereof) or to perform a treatment technique/s that required a hands-on component. The included articles all focused on simulation as an assessment tool and ideally were compared to other established forms of assessment.

Due to the large number of studies and reviews that have previously investigated objective structured clinical examinations (OSCEs),<sup>21-40</sup> all papers that investigated OSCEs were excluded. This was beyond the scope of this systematic review. Research articles focusing on the use of simulation as a training modality only were excluded. This included studies on simulation training program validity conducted by incorporating a simulation-based assessment at the end of the training. We excluded these articles as they did not address the effectiveness of the simulation as an assessment

but rather as a training tool. Studies that researched a specific simulation-based assessment grading tool were also excluded as these focused on tool validation and not on the assessment process. All search outcomes were assessed by two investigators (TR, BJ), and each abstract was read by both investigators for quality control.

Articles were also assessed for eligibility by their outcome measures. Many articles evaluated simulation as an assessment technique, but their primary outcome measure was a survey of participants' attitudes on the simulation experience. In such studies, nearly all participants found it to be a positive experience, with only minor suggestions for improvement.<sup>41-50</sup> Such studies were excluded as they did not focus on our primary aim of objectively determining the reliability, validity, or feasibility of simulation as an assessment technique.

## Critical appraisal

The methodological quality of all included full-text articles was assessed using a modified critical appraisal tool.<sup>20</sup> The McMasters Critical Review Form for Quantitative Studies has been used repeatedly in systematic reviews of health care<sup>51-54</sup>

**Table 1** Adapted critical appraisal tool

|  | Yes<br>(=1) | No<br>(=0) | Not<br>addressed<br>(=0) | Not<br>applicable<br>(=0) |
|--|-------------|------------|--------------------------|---------------------------|
| 1. Was the study purpose clearly stated?   |             |            |                          |                           |
| 2. Was relevant background literature reviewed?                                    |             |            |                          |                           |
| 3. Was the sample described in detail?   |             |            |                          |                           |
| 4. Was sample size justified?  |             |            |                          |                           |
| 5. Were the outcome measures reliable?   |             |            |                          |                           |
| 6. Were the outcome measures valid?  |             |            |                          |                           |
| 7. Was intervention described in detail?   |             |            |                          |                           |
| 8. Was contamination avoided?  |             |            |                          |                           |
| 9. Was cointervention avoided?   |             |            |                          |                           |
| 10. Were results reported in terms of statistical significance?                    |             |            |                          |                           |
| 11. If multiple outcomes, was that taken into account in the statistical analysis? |             |            |                          |                           |
| 12. Were the analysis methods appropriate?   |             |            |                          |                           |
| 13. Was educational importance reported?   |             |            |                          |                           |
| 14. Were dropouts reported?  |             |            |                          |                           |
| 15. Were the conclusions appropriate given study methods and results?              |             |            |                          |                           |

**Note:** Data from Law et al.<sup>20</sup>

and it demonstrates strong interrater reliability. We modified this tool using 15 items, which were scored dichotomously (Yes =1/No =0, and “Not addressed” and “Not applicable” were also scored zero). As such, a maximum score of 15 was permissible, and all 67 full-text articles were appraised and scored. Both reviewers (TR, BJ) independently scored the articles, and the final inclusion of full-text articles was discussed with the third reviewer (CJG) to find consensus, with the critical appraisal tool score used as a measure of methodological rigor. Forty-six articles were excluded for the reasons shown in Figure 1.

## Findings

Of the 21 articles included, the majority were from the field of medicine (n=16), with the remaining being from the

disciplines of paramedics (n=2), nursing (n=1), osteopathy (n=1), and physical therapy (n=1) (Table 2). Studies were undertaken in Australia, Denmark, New Zealand, Switzerland, and USA. There were no randomized controlled trials, and the majority of studies were of an observational study design. As such, blinding of participants and assessors to the simulation intervention was not undertaken in any of the studies. In addition, many studies used convenience samples that were not powered, and none of the studies calculated the number of participants required to achieve statistical significance, with one study commenting that their study was not adequately powered to detect differences between their two academic sites.<sup>55</sup> Many of the studies (n=13) were pilots or had small numbers (<50) of participants (range: n=18 to n=45). A small number of studies were conducted across different health centers<sup>55–57</sup>; however, the majority of these were conducted in a single health setting or university, making it difficult to generalize the findings. The included articles had scores on the critical appraisal tool ranging from 8 to 14/15. Some of the main reasons for low scores were a lack of description of the participants and a lack of either statistical analysis or description of the analysis.

## Themes

Eight (40%) of the studies used high-fidelity patient simulators to assess medical students, professionals, or applicants for a postgraduate nursing degree; with six (27%) of the studies using an SP combined with a clinical examination as the form of assessment. All of these studies varied in the format that they were completed, eg, the number of stations, time allowed, and the number of assessors used. The SP studies were conducted in the disciplines of medicine, physical therapy, and osteopathy and typically involved students rather than health professionals. Of the remaining studies, three studies (14%) used a virtual reality simulator to assess novice (medical students and professionals) to experienced professionals; two studies (9%) used manikins with varying levels of fidelity, from low-fidelity manikins to high-fidelity human patient simulators, to assess paramedics and intensive care unit (ICU) medical trainees; one study (5%) used a medium-fidelity patient simulator to assess paramedics; and one study (5%) included a part-task trainer to assess medical residents. Two studies compared SPs or low- to high-fidelity patient simulator assessments to other form(s) of assessment, such as paper-based examinations, oral examinations, or current university grade point averages. As such, the major themes that emerged are related to the type of simulation modality chosen for assessment (Supplementary material presents the definitions).

**Table 2** Summary of included articles

| Author (year)                           | Sample size/<br>discipline   | Primary outcome  | Simulation type                                 | Assessment type   | Main findings   |
|---|--|--|---|---|---|
| Asprey et al (2007) <sup>67</sup>       | 101 Medical students,<br>71 physician assistant<br>students  | Comparison of third<br>year medical student<br>scores with physician<br>assistant scores on SP<br>assessments - clinical skill<br>checklist  | SP  | Checklists –<br>completed by the SP   | SP-based assessments<br>were not able to<br>distinguish differences<br>between medical and<br>physician assistant<br>performance and are<br>therefore measuring<br>clinical experience as<br>the two groups had<br>comparable clinical<br>experience. |
| Bick et al (2013) <sup>55</sup>         | 26 Anesthetists  | Comparison of expert<br>vs novice anesthetist<br>performance on a<br>simulated transesophageal<br>echocardiography<br>examination – time and<br>accuracy checklist                                       | Virtual reality                                 | Checklists –<br>completed via video<br>by blinded (voice<br>masking technology)<br>experts  | Simulated<br>transesophageal<br>echocardiography<br>examination is able to<br>discriminate between<br>expert and novice<br>performers.  |
| Boulet et al (2003) <sup>59</sup>       | 24 Medical students, 13<br>medical residents   | Reliability and validity of<br>simulation assessment<br>for final year medical<br>students and residents   | High-fidelity<br>patient simulator              | Checklists – scored<br>via video by four<br>raters (two unaware<br>of the training levels<br>of participants)   | Simulation can validly<br>and reliably assess acute<br>care skills, but multiple<br>encounters are required<br>to predict performance.  |
| Burns et al (2013) <sup>62</sup>        | 26 Medical doctors<br>(16 interns, five PGY-2,<br>two PGY-3, two chief<br>residents, and three<br>pediatric hematology/<br>oncology fellows) | Performance scores<br>of interns (<1-year<br>experience in pediatrics)<br>vs the rest of the<br>residents and fellows<br>(>1-year experience)  | High-fidelity<br>pediatric patient<br>simulator | Checklists and<br>global rating scale –<br>scored via video by<br>two raters  | The manikin-<br>based assessment<br>provided reliable<br>and valid measures<br>of participants'<br>performance.   |
| Edelstein et al<br>(2000) <sup>68</sup> | 147 Medical students   | Comparison of<br>three measures of<br>medical student<br>performances - SP<br>examinations, computer-<br>based case simulations,<br>and traditional<br>performance indicators<br>(GPAs)                  | SP and computer-<br>based simulations           | Unclear how they<br>were assessed on<br>the SP examinations<br>– rated by both the<br>SP as well as the<br>author   | Differences were<br>found between the<br>SP assessments and<br>computer simulations<br>and traditional<br>assessments.<br>A multipronged<br>assessment approach is<br>supported.  |
| Fehr et al (2011) <sup>61</sup>         | 27 Anesthetic residents<br>and eight anesthetic<br>fellows   | Reliability and validity of<br>simulation assessments<br>for anesthesia residents<br>and pediatric fellows   | High-fidelity<br>patient simulator              | Checklists – rated<br>by lead investigator<br>and two experts   | The multiple-scenario<br>assessment could<br>reliably assess pediatric<br>residents and determine<br>the skill level of<br>participants. Further<br>validation required,<br>including comparing of<br>clinical performance.                           |
| Gimpel et al<br>(2003) <sup>71</sup>    | 121 Osteopathy<br>students   | Reliability and validity of<br>simulation assessment -<br>the Comprehensive<br>Osteopathic Medical<br>Licensing Examination–<br>USA performance-based<br>clinical skills examination<br>(COMPLEX-USA-PE) | SP  | Checklists – rated<br>by SPs on all<br>stations and experts<br>on the stations that<br>the participants<br>may have chosen<br>to perform a<br>manipulation<br>technique as part of<br>their treatment | Acceptable reliability<br>and validity of the<br>assessment. Use of the<br>Complex-USA-PE is<br>supported for future<br>use to assess the<br>readiness of osteopathic<br>medical students for<br>clinical practice.                                   |

(Continued)

**Table 2** (Continued)

| Author (year)                          | Sample size/<br>discipline  | Primary outcome  | Simulation type                                     | Assessment type   | Main findings  |
|--|---|--|---|---|--|
| Grantcharov et al (2005) <sup>75</sup> | Ten medical students, ten medical residents, and eight medical doctors  | Validity of virtual reality simulator for assessment of gastrointestinal endoscopy skills. Comparison of time and accuracy scores in three groups (experienced, residents, and medical students)                     | Virtual reality                                     | Parameters were calculated and recorded by the computer   | Virtual reality simulator was able to distinguish between the three levels of experience and therefore possessed acceptable construct validity.  |
| Hawkins et al (2004) <sup>70</sup>     | 54 Medical doctors  | Evaluation of computer-based case simulations and SP methods for patient management skills via checklists  | Computer-based case simulations and SP              | Computer-based scoring and checklists and a patient perception questionnaire for the SP station – rated by the SP                           | Computer-based case simulations and SP examinations were unable to distinguish between different experience levels; it appears that they may be useful in a multipronged assessment approach.                |
| Iyer et al (2013) <sup>78</sup>        | 16 Medical residents  | Validity of objective structured assessment of technical skills for neonatal lumbar puncture via competency-based scoring tool   | Task trainer  | Checklist and global rating scale – rated by six raters via video   | Reasonable evidence of validity of this assessment tool. Adoption of this tool in clinical environments is recommended for providing formative and summative feedback to improve resident skills.            |
| Lammers et al (2009) <sup>76</sup>     | 212 Paramedics  | Identifying performance of paramedics in simulated pediatric emergencies using prevalidated scoring checklists   | Low-, medium-, and high-fidelity patient simulators | Checklist – rated by one evaluator at the time and an author via video later, who amended scores if required and made the final decision    | Scores from manikin-based simulations objectively identified multiple performance deficiencies in paramedics.  |
| Lipner et al (2010) <sup>56</sup>      | 115 Medical cardiology fellows and cardiologists  | Evaluation of simulation-based assessment of technical and cognitive skills of physicians during coronary interventions by determining whether assessment differentiates performances of novice, skilled, and expert | Virtual reality                                     | Computer-based scoring system based on the consensus of a committee rating that included rating of the potential risk of taking that action | This assessment approach was able to identify poor-performance physicians who are unlikely to be providing appropriate patient care.   |
| McBride et al (2011) <sup>58</sup>     | 29 Medical residents (13 interns, nine second year residents, six third year residents, and one chief resident) | Multiple-scenario assessment (n=20) of residents' acute pediatric management skills using action checklists and global rating scales   | High-fidelity pediatric patient simulator           | Checklists and global rating scale – rated by two raters: one at the time, and one via video later  | Residents' scores in the assessments were found to be reliable and valid measures of their ability as long as multiple scenarios were used. Scores were able to discriminate between different skill levels. |

(Continued)



**Table 2** (Continued)

| Author (year)                              | Sample size/<br>discipline  | Primary outcome  | Simulation type                               | Assessment type  | Main findings   |
|--|---|--|---|--|---|
| Murray et al (2007) <sup>60</sup>          | 64 Medical residents and 35 anesthesiologists                         | Evaluation of anesthesia resident performances in a simulation-based intraoperative environment using an item score checklist  | High-fidelity patient simulator               | Checklists and time taken to perform the key actions – rated by two blinded raters via video; where there were large discrepancies, a third rater scored the performance   | The simulation-based assessment was found to be a reliable and valid measure of performance and was able to distinguish more experienced from early training participants.  |
| Nagoshi et al (2004) <sup>69</sup>         | 39 Medical students, 49 medical residents, and eleven medical fellows | Evaluation of a simulation-based standardized assessment of geriatric management skills by examining reliability and validity of the tool and scored using checklists        | SP  | Checklists – rated by the SP   | The assessment was found to be reliable but did not discriminate between levels of training of participants.  |
| Nunnink et al (2010) <sup>57</sup>         | 45 Medical doctors (ICU trainees)                                     | Validity of simulation-based assessments of intensive care trainees procedural skills by comparing with written exams and oral viva formats                                  | Low-high fidelity patient simulators and SP's | Templates for written responses and checklists for all other – one rater (also rated the written questions on the same topic)  | There was a lack of correlation between exam formats suggesting a multi-modal approach to assessment is favorable; and that simulation may be more useful in assessment of procedural skills.   |
| Panzarella and Manyon (2008) <sup>66</sup> | 34 Physical therapy students  | Evaluation of an Integrated Standardized Patient Examination for Physical therapy students   | SP  | Checklists for the subjective and objective assessment, four-point rubric for the integration question – rated by four raters: two in the room (the SP and an expert rater), one via video (the participant) and a “criterion” rater (one of the authors blinded to the others’ score) | Student scores need comparison to other forms of assessment of performance. Interrater reliability was found to vary between the raters - training for raters needs investigation. Good content validity reported by participants and expert raters. Needs further testing. |
| Penprase et al (2012) <sup>65</sup>        | 70 Registered nurses  | Evaluation of simulation-based assessment of applicants to Nurse Anesthesia programs by comparing scores to candidate interview scores                                       | High-fidelity patient simulator               | Checklists – rated by one of two trained raters concealed behind a one-way mirror  | Simulation-based assessment may provide a useful adjunct to the admission process.  |
| von Wyl et al (2009) <sup>77</sup>         | 30 Paramedics   | Determine the interrater reliability of assessing technical and nontechnical skills of paramedics during simulated emergency scenarios using previously validated checklists | Medium-fidelity patient simulator             | Checklists – rated by two raters (an experienced emergency physician and a psychologist) via video   | The assessment of technical and nontechnical skills using simulation-based assessment was shown to be feasible and reliable. There was a significant positive correlation between both skill types.   |

(Continued)

**Table 2** (Continued)

| Author (year)                      | Sample size/<br>discipline                     | Primary outcome   | Simulation type                 | Assessment type  | Main findings   |
|------------------------------------|--|---|---------------------------------|--|---|
| Waldrop et al (2009) <sup>63</sup> | 12 Medical interns and<br>44 medical residents | Evaluate the reliability and validity of simulation-based assessment in the management of intraoperative equipment-related errors in anesthesia | High-fidelity patient simulator | Checklists – rated initially by two raters for the first four participants; as scores were 100% in agreement, only one rater was used for the remaining participants | The assessment was found to be an effective, reliable, and valid method to determine individual performance. The assessment was also able to discriminate between the experience levels of participants.  |
| Weller et al (2005) <sup>64</sup>  | 21 Medical doctors                             | Evaluation of the psychometric properties of a simulation-based assessment for anesthesiologists  | High-fidelity patient simulator | Global rating scale - rated by four blinded raters via video   | The results show that 12–15 cases were required for acceptable reliability in this assessment modality. At lower level of performance, trainees overrated performance compared to those of high-performance levels, suggesting self-assessment may not be reliable. |

**Abbreviations:** GPA, grade point average, ICU, intensive care unit; PGY, postgraduate year; SP, standardized patient.

## High-fidelity simulation

Eight studies<sup>58–65</sup> conducted simulation-based assessment using high-fidelity human patient simulators. The simulators used were METI Emergency Care Simulator® (Medical Education Technologies Inc., Sarasota, FL, USA)<sup>58</sup>; METI HPS® (Sarasota, FL, USA)<sup>63,64</sup>; METI BabySim® (Sarasota)<sup>58</sup>; METI PediaSIM HPS® (Sarasota)<sup>61,62</sup>; SimMan 3.3 (Laerdal Medical, Wappingers Falls, NY, USA)<sup>65</sup>; SimNewB® (Laerdal Medical)<sup>58,61,62</sup>; and a life-size simulator developed by MEDSIM-EAGLE® (Med-Sim USA, Inc., Fort Lauderdale, FL, USA).<sup>59,60</sup>

All of the studies were conducted with medical students or practicing doctors, except for one focusing on postgraduate nursing applicants.<sup>65</sup> Generally, the reliability and validity of assessment using high-fidelity human patient simulators was found to be good. All of the medical-related studies used multiple scenarios (eg, trauma, myocardial infarction, and respiratory failure) using high-fidelity human patient simulators to assess the candidates, with the postgraduate nursing applicants only being assessed on one anesthetic scenario within a group of three.

As all of the assessments were targeting the clinical performance of students and doctors on high-risk skills, it was not surprising that high-fidelity patient simulators were a popular assessment modality. Unfortunately, this type of

assessment attracted generalizability coefficients less than what are acceptable for a high-stakes examination such as a summative performance assessment (G coefficients <0.8). All high-fidelity human patient simulator assessments were found to be suitable for low-stakes examinations (eg, a formative assessment of performance).

The evidence from these studies showed that increasing scenario numbers, rather than increasing the number of raters, increased assessment reliability.<sup>58–64</sup> Some researchers suggested that 10–12 scenarios, with three to four assessors, would be required to reach an acceptable level of reliability of 0.8,<sup>64</sup> while others observed that ten scenarios with two raters did not reach these levels (0.57).<sup>61</sup> When multiple raters (two to four) were used, interrater reliabilities of 0.59–0.97 (the majority being >0.8) were achieved.<sup>58–62</sup> While it was unclear how some of the raters reviewed the scenarios,<sup>61,63</sup> the majority rated the performance via a video recording of the scenarios,<sup>59,60,62,64</sup> with one study having a rater present at the time of the scenario as well as one scoring the performance via a video recording,<sup>58</sup> whereas one study used a one-way mirror to rate participants at the time.<sup>65</sup> When assessing pediatric interns, residents, and hematology/oncology fellows on sickle cell disease scenarios, checklists had superior interrater reliability than global rating scales.<sup>62</sup>



Construct validity was high in studies that used high-fidelity human patient simulators for assessing participants with varying degrees of experience (medical students through to specialists) as they were able to differentiate between the different levels of experience.<sup>58–63</sup> The pilot study investigating the correlation between high-fidelity human patient simulator assessment and face-to-face interviews for applicants applying for a postgraduate anesthetic nursing course found that there was a significant positive relationship ( $r=0.42$ ) between the two and that high-fidelity human patient simulator assessment was a suitable adjunct to the admissions process.<sup>65</sup>

## Standardized patients

There were six studies that used SPs within a clinical examination. These simulation-based assessments varied significantly in total duration, the number of stations, the amount of time per station, the number of SPs used, the types of stations used, and skills assessed, but all had the common feature of using SPs. These studies investigated SP encounters, but with fewer SP encounters than traditional OSCEs, and allowed participants longer time with each SP and expected more than just one technical skill to be performed at each station, eg, a full physical therapy assessment and treatment<sup>66</sup> under an assessment format. Four were from the medical profession,<sup>67–70</sup> with the others being from physical therapy<sup>66</sup> and osteopathy.<sup>71</sup> Two of the studies combined the SP assessments with computer-based assessments to assess medical students<sup>68</sup> as well as emergency medicine, general surgery, and internal medicine doctors.<sup>70</sup> Unfortunately, only three of these articles listed the presenting problem of the SP.<sup>66,69,71</sup>

Within SP assessments, participants' performance was assessed by trained assessors,<sup>66,71</sup> clinical experts,<sup>66</sup> self-assessment,<sup>66</sup> and the SPs themselves.<sup>66,67,69–71</sup> In all studies, assessors were trained to score the encounters, but only one study commented on the assessor's reliability. This study found that for physical therapy students, SP ratings did not significantly correlate with the ratings of other raters.<sup>66</sup> In contrast, strong agreement between experts and the criterion rater were evident. This suggests that experts and criterion raters are better placed than SPs to rate performances during high-stakes examinations.<sup>66</sup> When checklists were used by SPs, they negatively correlated with experience, possibly as more experienced doctors may solve problems and make decisions using fewer items of information, therefore checklists may lead to less valid scores.<sup>70</sup> Results varied as to whether SP examinations were able to determine clinical experience.<sup>67,69</sup> Nonetheless, they were found to

be reliable in assessing osteopathic students' readiness to treat patients.<sup>71</sup> The correlation between SP examinations and computer-based assessments was varied, with minimal correlation ( $r=0.24$  uncorrected and  $r=0.40$  corrected)<sup>68</sup> to low-to-moderate correlation ( $r=0.34$ – $0.48$ ).<sup>70</sup> However, SP examinations showed low correlation with curriculum results within physical therapy ( $<0.3$ ).<sup>66</sup> Overall, it was concluded that SP-based assessments should not be used in isolation to assess clinical competence.

## Virtual reality

Virtual reality is increasingly being adopted as a simulation tool. In the health professions, virtual reality simulation uses computers and human patient simulators to create a realistic and immersive learning and assessment environment.<sup>72–74</sup> Three studies used virtual reality<sup>55,56,75</sup> in simulation-based assessments comparing novice (medical students or residents), skilled (residents), and expert medical clinicians. Three different virtual reality systems were applied and all were shown to be able to differentiate between participants' skill levels. The systems used were the SimSuite system (Medical Simulation Corporation, Denver, CO, USA), which includes an interactive endovascular simulator<sup>56</sup>; GI Mentor II computer system (Simbionix Ltd, Cleveland, OH, USA)<sup>75</sup>; and a Heartworks TEE Simulator (Inventive Medical Ltd, London, UK), which includes a manikin and haptic-simulated probe.<sup>55</sup> All three systems were found to have construct validity as they were able to distinguish between the technical ability among the groups and therefore they are useful in determining those that require further training prior to clinical practice on real patients.

## Mixed-fidelity patient simulators

Two studies that used low-, medium-, and high-fidelity human patient simulators during an assessment of paramedics<sup>76</sup> and ICU medical trainees' resuscitation skills<sup>57</sup> were included in this review. All three levels of patient simulator fidelity were found to have high interrater reliability in these populations.<sup>76</sup> Intensive care trainees were assessed on medium- and high-fidelity human patient simulators as well as by written and oral viva examinations.<sup>57</sup> The written examination was shown not to correlate with either medium- or high-fidelity human patient simulation-based assessments, indicating that written and simulation assessments differed in their ability to evaluate knowledge and practical skills. Specific skill deficiencies were able to be determined when low- to high-fidelity simulators were used, therefore allowing subsequent training to be targeted to individuals' needs.<sup>76</sup>

## Medium-fidelity simulation

One study<sup>77</sup> investigated a medium-fidelity simulator (human patient simulator designed to allow limited invasive procedures with lower fidelity needs) and a volunteer with moulage. The paramedics were assessed in pairs on two simulated scenarios (acute coronary syndrome and a severe traumatic brain injury). Both their technical and nontechnical skills were assessed via separate checklists; the two assessors rated the performance from a video-recording and were allowed to rewind as necessary. Interrater reliability (between an emergency physician and psychologist) showed good correlations, especially for technical skills such as assessment of primary airway, breathing, circulation, and defibrillation. A positive and significant relationship was found between technical and nontechnical skills. Accordingly, one rater was found to be sufficient to adequately assess technical skills, but two raters were required to demonstrate equivalent reliability for nontechnical skills.

## Task trainer

One pilot study<sup>78</sup> investigated the performance of pediatric residents in lumbar puncture using a neonatal lumbar puncture task trainer. This simulation-based assessment used a video-delayed format, in which six raters reviewed the video and assessed the pediatric residents' performance based on the seven criteria of preparation, positioning, analgesia administration, needle insertion technique, cerebrospinal fluid (CSF) fluid return/collection, diagnostic purpose/laboratory management of CSF, and creating and maintaining a sterile field. There was good interrater reliability and validity in regard to the response process (potential bias if the raters recognized the residents; voices were not altered, but faces were not shown) and relationship to external variables (eg, previous experience in neonatal or pediatric ICUs).

## Discussion

We undertook a systematic review to examine simulation as an assessment tool across health professional education. Although this review demonstrated that simulation-based assessments of technical-based skills can be used reliably and are valid, the research was constrained by the findings that simulation-based assessments were commonly used in isolation, not in combination with other assessment forms or with more than one simulation scenario. This review also demonstrated that assessments using high-fidelity simulators and SPs have been more widely adopted. High-fidelity simulation was more widely adopted in medicine and commonly used in the emergency and anesthetic specialties in which

high-risk skill assessments are used more frequently. The evidence suggests that participants can be assessed reliably with high-fidelity human patient simulators combined with multiple station assessment tasks with well-constructed scenarios. Overall, the results are promising for the future use and development of simulation-based assessment in the health education field.

Due to the multiplicity of simulation-based assessments, it was difficult to compare data between studies and definitive statements on which form of assessment type would be best for health disciplines and for students and practicing health professionals. In regard to health students, standardizing assessments created a fairer and more consistent approach, leading to greater equity and reliability. Simulation appears to achieve this in competency-based assessments as well as being a useful tool for predicting future performances. This area of research needs exploration as it may have the potential to determine future performances of students and their competency, especially in relation to whether students are ready for clinical environments and exposure to real patients. Simulation-based assessments may also assist newly graduated health professionals who could be deemed competent by using reliable and valid authentic assessments prior to commencing practice in a new area. In addition, simulation-based assessment is a promising approach for determining the skill level and capability for safe practice, as it appears to be able to distinguish between different levels of performance among novice and expert groups as well as being able to identify poor performers, allowing for safe practice.

The methodological rigor was an issue, with many of the studies having scores on the critical appraisal tool ranging from 8 to 14/15. Many of the studies had modest participant numbers, a common limitation noted in several studies,<sup>58,62,64,70,78</sup> which may limit the generalizability of the results. Sample size was not justified in many instances, and there was little mention of participant dropouts. In contrast, three studies had substantive participant numbers (>120 participants)<sup>68,71,76</sup> and provided robust analyses, which increased external validity.

A noticeable gap in this literature is that only three of the articles reviewed compared their simulation-based assessment to another assessment form or simulation type. These comparative studies provided a higher degree of critique of the assessment type and permitted observation of differences, which may be of assistance for health educators. We believe that comparative studies should be conducted in future research to provide evidence of assessment superiority and enhanced informed assessment choices.

A continuation of this theme is that studies examining the reliability and validity of simulation-based assessments need stronger research approaches, such as blinding assessors and participants, providing precise details of the intervention, and – where possible – to avoid contamination. While we appreciate that educational research is often challenging, robust study design should be tantamount.

Overall, further research is required to determine which form of simulation-based assessment is best suited in specific health professional learner situations. While it is suggested that simulation-based assessments should not be used in isolation to make an overall assessment of an individual's clinical and theoretical skills, simulation-based assessments are being widely used and sometimes for this discrete purpose. Development of simulation-based assessment needs to continue as it will provide clarity and consistency for the assessors and participants, in addition to furthering the use of simulation in health education. As simulation is increasingly being used to replace a proportion of health student's clinical practice time,<sup>79,80</sup> it is expected that simulation-based assessment will become an integral component of health professional curricula and, therefore, it needs to be evidence based and valid. This will provide stronger conclusions for the use of simulation-based assessment in health professional education.

## Limitations

There were several limitations of this systematic review. Studies included were limited to the English language, and there may well be other studies conducted and published in non-English-speaking publications. Due to the varying nature of the studies, we were unable to complete any form of pooled data analysis. We did not include studies that investigated the cost-effectiveness and cost analysis of simulation-based assessments. Studies of this type may highlight other areas of practicalities not highlighted in this systematic review. As already mentioned, studies investigating OSCEs were also excluded due to the extensive previous research conducted in this area. The inclusion of OSCE-based studies may have helped to strengthen the argument for SP use, as this tends to be the most common form of simulation used within the OSCE literature.

## Conclusion

The use of simulation within health education is expanding; in particular, its use in the training of health professionals and students. The evidence from this review suggests that the use of SPs would be a practical approach for many clinical situations, with the use of part-task trainers or patient simulators to aid in areas in which the actors are unable to “act”

or in cases wherein invasive procedures are undertaken. In assessments in which clinical skills need to be evaluated in high-pressure situations, the evidence of simulation-based assessments is that the use of patient simulators in high-fidelity environments may be more suitable than using task trainers. High-fidelity simulation assessments could also be used to incorporate and assess multidisciplinary team assessments. Overall, there is a clear need for further methodologically robust research into simulation-based assessments within health professional education.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Boud D. Sustainable assessment: rethinking assessment for the learning society. *Stud Contin Educ*. 2000;22(2):151–167.
2. Wiggins G. A true test: toward more authentic and equitable assessment. *Phi Delta Kappan*. 1989;70(9):703–713.
3. Falchikov N, Goldfinch J. Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Rev Educ Res*. 2000;70(3):287–322.
4. Swanson DB, Norman GR, Linn RL. Performance-based assessment: lessons from the health professions. *Educ Res*. 1995;24(5):5–11.
5. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med*. 1991;66(12):762–769.
6. Watson R, Stimpson A, Topping A, Porock D. Clinical competence assessment in nursing: a systematic review of the literature. *J Adv Nurs*. 2002;39(5):421–431.
7. Redfern S, Norman I, Calman L, et al. Assessing competence to practise in nursing: a review of the literature. *Res Pap Educ*. 2002;17(1):51–77.
8. Greiner AC, Knebel E. *Health Professions Education: A Bridge to Quality*. Washington (DC): National Academies Press; 2003.
9. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387–396.
10. Rosen KR. The history of medical simulation. *J Crit Care*. 2008; 23(2):157–166.
11. Issenberg SB, Scalese RJ. Simulation in health care education. *Perspect Biol Med*. 2008;51(1):31–46.
12. Gaba DM. The future vision of simulation in health care. *Qual Saf Health Care*. 2004;13(suppl 1):i2–i10.
13. Weller JM, Nestel D, Marshall SD, Brooks PM, Conn JJ. Simulation in clinical teaching and learning. *MJA*. 2012;196(6):1–5.
14. Ziv A, Small SD, Wolpe PR. Patient safety and simulation-based medical education. *Med Teach*. 2000;22(5):489–495.
15. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Acad Med*. 2003;78(8):783–788.
16. Ker J, Mole L, Bradley P. Early introduction to interprofessional learning: a simulated ward environment. *Med Educ*. 2003;37(3):248–255.
17. Alinier G, Hunt WB, Gordon R. Determining the value of simulation in nurse education: study design and initial results. *Nurse Educ Pract*. 2004;4(3):200–207.
18. Seropian MA, Brown K, Gavilanes JS, Driggers B. Simulation: not just a manikin. *J Nurs Educ*. 2004;43(4):164–169.
19. Moher DP, Liberati AMDD, Tetzlaff JB, Altman DG; PRISMA Group. The PG. preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264–269.
20. Law M, Stewart D, Pollock N, Letts L, Bosch J, Westmorland M. *Critical Review Form – Quantitative Studies*. McMaster University Occupational Therapy Evidence-Based Practice Research Group; Hamilton, Ontario: 1998.

21. Hodges B. Validity and the OSCE. *Med Teach*. 2003;25(3):250–254.
22. Wallace J, Rao R, Haslam R. Simulated patients and objective structured clinical examinations: review of their use in medical education. *Adv Psychiatr Treat*. 2002;8(5):342–348.
23. Pell G, Fuller R, Homer M, Roberts T; International Association for Medical Education. International Association for Medical E. How to measure the quality of the OSCE: a review of metrics – AMEE guide no. 49. *Med Teach*. 2010;32(10):802–811.
24. Carraccio C, Englander R. The objective structured clinical examination: a step in the direction of competency-based evaluation. *Arch Pediatr Adolesc Med*. 2000;154(7):736–741.
25. Bartfay WJ, Rombough R, Howse E, Leblanc R. Evaluation. The OSCE approach in nursing education. *Can Nurse*. 2004;100(3):18–23.
26. Barman A. Critiques on the objective structured clinical examination. *Ann Acad Med Singapore*. 2005;34(8):478–482.
27. Rushforth HE. Objective structured clinical examination (OSCE): review of literature and implications for nursing education. *Nurse Educ Today*. 2007;27(5):481–490.
28. Khattab AD, Rawlings B. Use of a modified OSCE to assess nurse practitioner students. *Br J Nurs*. 2008;17(12):754–759.
29. Casey PM, Goepfert AR, Espey EL, et al; Association of Professors of Gynecology and Obstetrics Undergraduate Medical Education Committee. To the point: reviews in medical education – the objective structured clinical examination. *Am J Obstet Gynecol*. 2009;200(1):25–34.
30. Patricio M, Juliao M, Fareleira F, Young M, Norman G, Vaz Carneiro A. A comprehensive checklist for reporting the use of OSCEs. *Med Teach*. 2009;31(2):112–124.
31. Mitchell ML, Henderson A, Groves M, Dalton M, Nulty D. The objective structured clinical examination (OSCE): optimising its value in the undergraduate nursing curriculum. *Nurse Educ Today*. 2009;29(4):398–404.
32. Walsh M, Bailey PH, Koren I. Objective structured clinical evaluation of clinical competence: an integrative review. *J Adv Nurs*. 2009;65(8):1584–1595.
33. Hodges BD, Hollenberg E, McNaughton N, Hanson MD, Regehr G. The psychiatry OSCE: a 20-year retrospective. *Acad Psychiatry*. 2014;38(1):26–34.
34. Hastie MJ, Spellman JL, Pagano PP, Hastie J, Egan BJ. Designing and implementing the objective structured clinical examination in anesthesiology. *Anesthesiology*. 2014;120(1):196–203.
35. Phillips D, Zuckerman JD, Strauss EJ, Egol KA. Objective structured clinical examinations: a guide to development and implementation in orthopaedic residency. *J Am Acad Orthop Surg*. 2013;21(10):592–600.
36. Patricio MF, Juliao M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Med Teach*. 2013;35(6):503–514.
37. Lillis S, Stuart M, Sidonie FS, Stuart N. New Zealand registration examination (NZREX Clinical): 6 years of experience as an objective structured clinical examination (OSCE). *N Z Med J*. 2012;125(1361):74–80.
38. Smith V, Muldoon K, Biesty L. The objective structured clinical examination (OSCE) as a strategy for assessing clinical competence in midwifery education in Ireland: a critical review. *Nurse Educ Pract*. 2012;12(5):242–247.
39. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Med Educ*. 2011;45(12):1181–1189.
40. Turner JL, Dankoski ME. Objective structured clinical exams: a critical review. *Fam Med*. 2008;40(8):574–578.
41. Costello E, Plack M, Maring J. Validating a standardized patient assessment tool using published professional standards. *J Phys Ther Educ*. 2011;25(3):30–45.
42. Ruesseler M, Weinlich M, Byhahn C, et al. Increased authenticity in practical assessment using emergency case OSCE stations. *Adv Health Sci Educ Theory Pract*. 2010;15(1):81–95.
43. Ebbert DW, Connors H. Standardized patient experiences: evaluation of clinical performance and nurse practitioner student satisfaction. *Nurs Educ Perspect*. 2004;25(1):12–15.
44. Alinier G. Nursing students' and lecturers' perspectives of objective structured clinical examination incorporating simulation. *Nurse Educ Today*. 2003;23(6):419–426.
45. Landry M, Oberleitner MG, Landry H, Borazjani JG. Education and practice collaboration: using simulation and virtual reality technology to assess continuing nurse competency in the long-term acute care setting. *J Nurses Staff Dev*. 2006;22(4):163–171.
46. Sharpnack PA, Madigan EA. Using low-fidelity simulation with sophomore nursing students in a baccalaureate nursing program. *Nurs Educ Perspect*. 2012;33(4):264–268.
47. Forsberg E, Georg C, Ziegert K, Fors U. Virtual patients for assessment of clinical reasoning in nursing – a pilot study. *Nurse Educ Today*. 2011;31(8):757–762.
48. Paul F. An exploration of student nurses' thoughts and experiences of using a video-recording to assess their performance of cardiopulmonary resuscitation (CPR) during a mock objective structured clinical examination (OSCE). *Nurse Educ Pract*. 2010;10(5):285–290.
49. Brimble M. Skills assessment using video analysis in a simulated environment: an evaluation. *Paediatr Nurs*. 2008;20(7):26–31.
50. Richardson L, Resick L, Leonardo M, Pearsall C. Undergraduate students as standardized patients to assess advanced practice nursing student competencies. *Nurse Educ*. 2009;34(1):12–16.
51. Lekkas P, Larsen T, Kumar S, et al. No model of clinical education for physiotherapy students is superior to another: a systematic review. *Aust J Physiother*. 2007;53(1):19–28.
52. Egan M, Hobson S, Fearing VG. Dementia and occupation: a review of the literature. *Can J Occup Ther*. 2006;73(3):132–140.
53. Schabrun S, Chipchase L. Healthcare equipment as a source of nosocomial infection: a systematic review. *J Hosp Infect*. 2006;63(3):239–245.
54. Gordon J, Sheppard LA, Anaf S. The patient experience in the emergency department: a systematic synthesis of qualitative research. *Int Emerg Nurs*. 2010;18(2):80–88.
55. Bick JS, DeMaria S Jr, Kennedy JD, et al. Comparison of expert and novice performance of a simulated transesophageal echocardiography examination. *Simul Healthc*. 2013;8(5):329–334.
56. Lipner RS, Messenger JC, Kangilaski R, et al. A technical and cognitive skills evaluation of performance in interventional cardiology procedures using medical simulation. *Simul Healthc*. 2010;5(2):65–74.
57. Nunnink L, Venkatesh B, Krishnan A, Vidhani K, Udy A. A prospective comparison between written examination and either simulation-based or oral viva examination of intensive care trainees' procedural skills. *Anaesth Intensive Care*. 2010;38(5):876–882.
58. McBride ME, Waldrop WB, Fehr JJ, Boulet JR, Murray DJ. Simulation in pediatrics: the reliability and validity of a multiscenario assessment. *Pediatrics*. 2011;128(2):335–343.
59. Boulet JR, Murray D, Kras J, Woodhouse J, McAllister J, Ziv A. Reliability and validity of a simulation-based acute care skills assessment for medical students and residents. *Anesthesiology*. 2003;99(6):1270–1280.
60. Murray DJ, Boulet JR, Avidan M, et al. Performance of residents and anesthesiologists in a simulation-based skill assessment. *Anesthesiology*. 2007;107(5):705–713.
61. Fehr JJ, Boulet JR, Waldrop WB, Snider R, Brockel M, Murray DJ. Simulation-based assessment of pediatric anesthesia skills. *Anesthesiology*. 2011;115(6):1308–1315.
62. Burns TL, DeBaun MR, Boulet JR, Murray GM, Murray DJ, Fehr JJ. Acute care of pediatric patients with sickle cell disease: a simulation performance assessment. *Pediatr Blood Cancer*. 2013;60(9):1492–1498.
63. Waldrop WB, Murray DJ, Boulet JR, Kras JF. Management of anesthesia equipment failure: a simulation-based resident skill assessment. *Anesth Analg*. 2009;109(2):426–433.



64. Weller JM, Robinson BJ, Jolly B, et al. Psychometric characteristics of simulation-based assessment in anaesthesia and accuracy of self-assessed scores. *Anaesthesia*. 2005;60(3):245–250.
65. Penprase B, Mileto L, Bittinger A, et al. The use of high-fidelity simulation in the admissions process: one nurse anesthesia program's experience. *AANA J*. 2012;80(1):43–48.
66. Panzarella KJ, Manyon AT. Using the integrated standardized patient examination to assess clinical competence in physical therapist students. *J Phys Ther Educ*. 2008;22(3):24–32.
67. Asprey DP, Hegmann TE, Bergus GR. Comparison of medical student and physician assistant student performance on standardized-patient assessments. *J Physician Assist Educ*. 2007;18(4):16–19.
68. Edelstein RA, Reid HM, Usatine R, Wilkes MS. A comparative study of measures to evaluate medical students' performances. *Acad Med*. 2000;75(8):825–833.
69. Nagoshi M, Williams S, Kasuya R, Sakai D, Masaki K, Blanchette PL. Using standardized patients to assess the geriatrics medicine skills of medical students, internal medicine residents, and geriatrics medicine fellows. *Acad Med*. 2004;79(7):698–702.
70. Hawkins R, Gaglione MM, LaDuca T, et al. Assessment of patient management skills and clinical skills of practising doctors using computer-based case simulations and standardised patients. *Med Educ*. 2004;38(9):958–968.
71. Gimpel JR, Boulet JR, Errichetti AM. Evaluating the clinical skills of osteopathic medical students. *J Am Osteopath Assoc*. 2003;103(6):267.
72. Aggarwal R, Grantcharov TP, Eriksen JR, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. *Ann Surg*. 2006;244(2):310–314.
73. Banerjee PP, Luciano CJ, Rizzi S. Virtual reality simulations. *Anesthesiol Clin*. 2007;25(2):337–348.
74. Haubner M, Krapichler C, Losch A, Englmeier KH, van Eimeren W. Virtual reality in medicine-computer graphics and interaction techniques. *IEEE IEEE Trans Inf Technol Biomed*. 1997;1(1):61–72.
75. Grantcharov TP, Carstensen L, Schulze S. Objective assessment of gastrointestinal endoscopy skills using a virtual reality simulator. *JSLs*. 2005;9(2):130–133.
76. Lammers RL, Byrwa MJ, Fales WD, Hale RA. Simulation-based assessment of paramedic pediatric resuscitation skills. *Prehosp Emerg Care*. 2009;13(3):345–356.
77. von Wyl T, Zuercher M, Amsler F, Walter B, Ummenhofer W. Technical and non-technical skills can be reliably assessed during paramedic simulation training. *Acta Anaesthesiol Scand*. 2009;53(1):121–127.
78. Iyer MS, Santen SA, Nypaver M, et al; Accreditation Council for Graduate Medical Education Committee, Emergency Medicine and Pediatric Residency Review Committee. Assessing the validity evidence of an objective structured assessment tool of technical skills for neonatal lumbar punctures. *Acad Emerg Med*. 2013;20(3):321–324.
79. Blackstock FC, Watson KM, Morris NR, et al. Simulation can contribute a part of cardiorespiratory physiotherapy clinical education two randomized trials. *Simul Healthc*. 2013;8(1):32–42.
80. Watson K, Wright A, Morris N, et al. Can simulation replace part of clinical time? Two parallel randomised controlled trials. *Med Educ*. 2012;46(7):657–667.

## Supplementary materials

### Definitions

#### High-fidelity patient simulators

These are designed to allow a large range of noninvasive and invasive procedures to be performed and offer realistic sensory and physiological responses, with outputs such as heart rate and oxygen saturation usually displayed on a monitor. They can be run by a computer technician or preprogrammed to react to the participant's actions.

#### Objective structure clinical examinations

These involve participants progressing through multiple stations at predetermined time intervals. They may have active or simulation-based stations that assess practical skills or

passive stations such as written or video analysis, commonly used to assess theoretical knowledge.

#### Standardized patients

These are people trained to portray a patient in a consistent manner and present the case history of a real patient using predetermined subjective and objective responses.

#### Task trainers

These are models that are designed to look like a part of the human anatomy and allow individuals to perform discrete invasive procedures, for example a pelvis for internal pelvic examinations, or an arm to practice cannulation.

### Journal of Multidisciplinary Healthcare

#### Publish your work in this journal

The Journal of Multidisciplinary Healthcare is an international, peer-reviewed open-access journal that aims to represent and publish research in healthcare areas delivered by practitioners of different disciplines. This includes studies and reviews conducted by multidisciplinary teams as well as research which evaluates the results or conduct of such teams or health-

care processes in general. The journal covers a wide range of areas and welcomes submissions from practitioners at all levels, from all over the world. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/journal-of-multidisciplinary-healthcare-journal>

Dovepress