

COPYRIGHT NOTICE



FedUni ResearchOnline

<https://researchonline.federation.edu.au>

This is the peer-reviewed version of the following article:

Cant, R. P., et al. (2018). "Do Simulation Studies Measure up? A Simulation Study Quality Review." *Clinical Simulation in Nursing* 21: 23-39.

Which has been published in final form at:

<https://doi.org/10.1016/j.ecns.2018.06.002>

Copyright © 2018 Elsevier Inc. All rights reserved.

Do simulation studies measure up? A simulation study quality review

AUTHORS

Robyn P Cant ¹, Tracy Levett-Jones ², Ainsley James ³

¹ Research Fellow, Nursing and Midwifery, Monash University, Clayton VIC, Australia

² Professor of Nursing Education, University of Technology Sydney, Ultimo NSW, Australia.

³ Lecturer, Federation University Australia, Churchill VIC, 3842, Australia.

Key Points

- We assessed the quality of simulation studies in nursing using two valid tools
- Medical Education Research Study Quality Instrument and Simulation Research Evaluation Rubric were found valid and feasible
- Most studies were high in quality however some elements of reporting can be improved.

ABSTRACT

BACKGROUND

Simulation-based education has become a ubiquitous teaching approach in nursing.

However, ensuring the quality of simulation research is critical.

METHODS

We reviewed the methodological quality of 26 quantitative studies published in Clinical Simulation in Nursing, 2017.

RESULTS

Medical Education Research Study Quality Instrument and Simulation Research Evaluation Rubric showed that nearly all studies were moderate to high quality (rated $\geq 50\%$).

Correlation coefficients showed inter-rater agreement was high overall ($\geq .94$).

CONCLUSIONS

This was a valid approach for examining simulation study quality. While most included studies were of high quality, some elements of study reporting can be improved upon.

INTRODUCTION

Simulation-based education (SBE) has become a ubiquitous component of nursing education and a strategy recognized for its potential to improve learners' competence and confidence (Motola, Devine, Chung, Sullivan & Issenberg, 2013). SBE offers a wide range of realistic and 'life-like' simulated clinical experiences where learners can engage with computerized 'patient' manikins, standardized patients (actor), or other simulation modalities, in preparation for practice (Cant & Cooper, 2017; Cantrell, Franklin, Leighton, & Carlson, 2017). A key element of SBE is the provision of formative and/or summative feedback which enables reflection on performance to enhance learning (INACSL, 2016; Levett-Jones & Lapkin, 2014).

SBE has been utilized in both pre-licensure and post-graduate nurse training (Cantrell et al., 2017; Rutherford-Hemming & Alfes, 2017). Internationally, several countries have acknowledged the value of simulation for nursing education by approving the use of SBE as partial replacement of clinical practice hours. For example, in the US, the National Council for State Boards of Nursing Study determined that SBE could be substituted for clinical practice hours (NCSBN, 2010; Hayden, Smiley, Alexander, Kardong-Edgren, & Jeffries, 2014). A national survey of 432 academics in US nursing schools reported that three-quarters of the schools were substituting SBE for clinical hours (Breymer et al., 2015). In the UK, the Nursing and Midwifery Council (2010) ruled that up to 300 hours (13%) of the required 2300 hours of clinical practice can be replaced with simulation-based practice (p. 9).

One of the challenges to the use of SBE and its use as a replacement for or else supplementation of clinical placement hours, is the variable quality of simulation studies. The strength of evidence for the contribution of SBE to learning is under scrutiny. In nursing, reviews of literature have regularly cited a lack of studies with comparative research designs as a limiting factor in reporting overall effectiveness (Cant & Cooper, 2017; Doolen et al., 2017; Rutherford-Hemming & Alfes, 2017). In a review of simulation studies in nursing, Cantrell et al. (2017) found the evidence relied too heavily on self-reported measures. A review of 72 studies of SBE in pre-licensure nurse education, identified that although many studies presented valid empirical evidence for knowledge gain, larger parallel studies were

required to enable cross-sectional comparisons (Cant & Cooper, 2017). A meta-analysis of randomized controlled trials of simulation investigating nurses' skills and knowledge reported a small effect in favour of simulation, but the quality of evidence was regarded as low and the overall impact was thus uncertain (Hegland, Aarlie, Strømme, & Jamtvedt, 2017). Doolan et al. (2017) in reviewing simulation use in undergraduate nursing, suggested a need for more methodologically sound research. Teaching and learning research in simulation-based studies is thus difficult to plan for, and to conduct.

The reviews cited above include 172 primary simulation studies and 104 reviews of simulation literature. None of these were able to reach unreserved conclusions about the overall benefit of SBE because of a lack of similarity and various limitations in the design and implementation of reviewed studies. Ideally, we would base education practices on unequivocal approval of the empirical evidence provided in research. All this suggests an imperative to examine the methodological quality of current simulation studies in order to inform the development of future research.

The aim of this study is, therefore, to examine the methodological quality of quantitative research publications that evaluate the use of SBE in nursing. We will critique recent nursing simulation research using valid assessment instruments.

METHODS

This study presents a 'systematic search and review' of current literature, based on Grant and Booth's (2009, p. 95) operational definitions. It will produce a best evidence synthesis of the state of simulation designs/methodologies and study outcomes and make recommendations for high quality research.

Sample

A purposeful sample of relevant publications was extracted from a key simulation journal in publications between January and December, 2017. This period was chosen as it reflects the most current state of the art of simulation in nursing. The journal *Clinical Simulation in Nursing* (CSN) is the official journal of the *International Nursing Association for Clinical Simulation and Learning* (INACSL) and thus is highly regarded for nursing simulation publications worldwide. Forty-seven primary simulation studies published in 2017 were identified and screened for inclusion in the review (Figure 1).

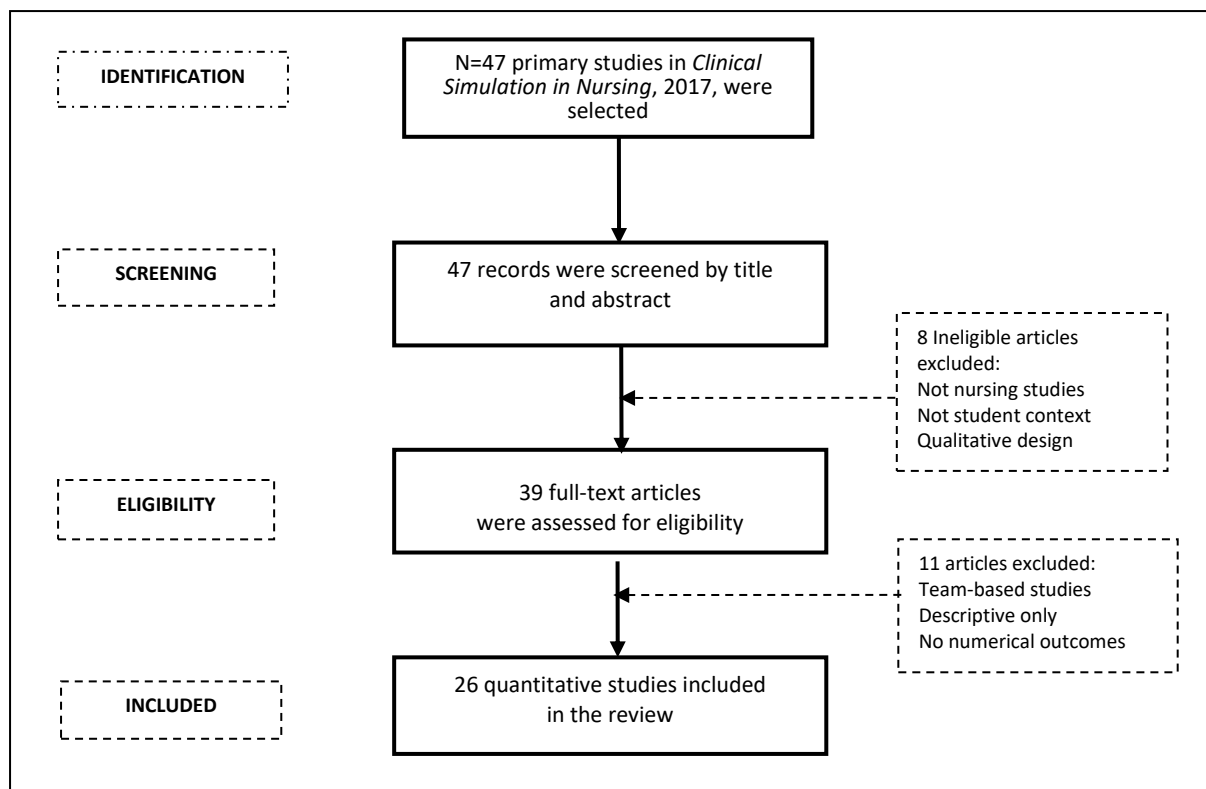


Figure 1 Flow chart of study selection

Inclusion and exclusion criteria

All primary quantitative simulation research studies published in CSN in 2017 that reported a simulation intervention and study outcomes were selected. Studies of nurses or nursing students using any level of simulation fidelity were included. Interprofessional (eg, team training) studies were excluded, as well as descriptive or methods studies such as studies of course design, pre-briefing, or debriefing as they did not meet the assessment instrument criteria. The publications were downloaded into an electronic library database and examined by title and abstract. Eight ineligible studies were removed at this stage, leaving 37 potential studies to be included.

Synthesis

The study characteristics were tabulated to identify study details, including country of origin, topic, research design, sample, simulation intervention, evaluation measures, and main outcome. Once this detail was examined, 11 studies were found ineligible and were removed, leaving 26 studies to be included in the analysis. Ineligible studies were, for example, not a quantitative design, were not nursing data, were descriptive, or else they

reported instrument development. All three authors reviewed the publication list to ensure inclusion criteria were met.

The papers were assessed for methodological quality using two quality assessment instruments (see below) and independent assessments that were undertaken by two authors. The first author (RC) independently completed a set of study ratings using both appraisal instruments, and the two other authors rated the studies each using a different appraisal instrument. Thus, two sets of quality scores from two measures were provided for each study, enabling further scrutiny.

‘Quality of research’ measures

The properties of the two assessment instruments are described below.

1. Medical Education Research Study Quality Instrument (MERSQI)

The MERSQI was designed in 2007 to measure the methodological quality of experimental, quasi-experimental, and observational studies (Reed, Beckman, & Wright, 2009), and since then has frequently been used to evaluate the quality of research in medical education (Braga, et al., 2015; DeCoste-Lopez, Madhok, & Harman, 2015; Hsieh et al., 2016; Ilgen, Sherbino, & Cook, 2013; McKinney, & Wood, 2013). The MERSQI includes 10 items clustered into the six domains of: study design, sampling, type of data, validity evidence for evaluation instrument scores, rating of data analysis, and strength of outcomes. Scores between 1 and 3 points per domain are given according to operational definitions, with possible total scores between 5 and 18 points.

Reed et al. (2007) in developing the MERSQI reported strong Intraclass Correlation Coefficient ranges for interrater and intra-rater reliability (0.72 to 0.98 and 0.78 to 0.99, respectively) in an early study of research publications versus funding. Cook and Reed (2015) reported further evidence of the scale’s validity in a comparison with the Newcastle–Ottawa Scale-Education, reporting high interrater and intra-rater reliability and favourable correlations with global quality ratings from two independent experts. They concluded MERSQI is a reliable tool for appraising methodological quality of medical education research, also suggesting that interpretation and use of scores should focus on item-specific codes rather than overall scores.

The MERSQI has rarely been used in reviews of nursing research other than when used in multidisciplinary training reviews (team training) that included nurses. A search of the

literature did not identify any study that explored the quality of clinical simulation studies in nursing with the MERSQI and thus, the current study may help to fill this gap.

2. The Simulation Research Evaluation Rubric (SRR)

Specifically focused on evaluation of simulation research studies, the SRR was developed by Fey, Gloe and Mariani (2015) and first published in CSN. It includes a rating scheme for the reporting of 16 items including: study rationale, sample and setting, description of simulation, implementation, study instruments, debriefing/feedback, results and discussion. There is also an option to rate qualitative and mixed methods designs by substituting variables. Each element is rated on a scale of 0-4 points (0= unsatisfactory, 4= excellent), earning a total score of 56 points for quantitative or qualitative studies and 64 points for a mixed methods study. The scale was found reliable during the development phase with a content validity index of 0.96 and inter-rater reliability of 0.92, although no further study or testing of psychometric properties was identified.

Analysis

The sets of quality scores were collated and consistency between assessments was explored using IBM-SPSS vs 25 (IBM Corp, 2016). A Kappa Measure of Agreement (k) between raters was computed for each study. Kappa, whilst measuring absolute agreement, also accounts for differences that occur by chance; a value of $k = .5$ shows moderate agreement; $k = .7$ represents good agreement, $k \geq .8$ represents very good agreement (Pallant, 2013). Correlations between assessments using the elements rated by two raters were computed using Spearman's Correlation (rho: r). Relationship strength was indicated by $r = .30- .49$ (medium) and $r = .50-1.0$ (large). The overall consistency and reproducibility of all the ratings derived from each instrument was tested using the Intra-class Correlation Co-efficient (ICC), two-way model. The association between the mean total proportional scores in each instrument was examined using Wilcoxon Signed Ranks Test. $P < 0.05$ was regarded as significant.

RESULTS

Twenty-six quantitative primary studies of simulation education in nursing published in CSN in 2017 were reviewed (see Table 1 *at end of this file*). The majority of studies ($n=20$) described SBE use in pre-licensure nurse education; five studies described professional

nurse training and one study reported both student and trained nurse outcomes. Eighteen studies were conducted in USA, four in Canada, and one each in Singapore, South Korea, China, and Australia.

The level of research design varied from experimental controlled trials (n=2), to single-group or two-group pre-post evaluation studies (n=13), to post-test only single group designs (n=11).

The participant samples for the 20 nursing student studies ranged from 23 to 207 (mean n=78), while the five studies of professional nurses sampled between 11 and 749 nurses (M= 482). The single study that sampled both nurses and students sampled 1,229 qualified nurses and n= 1,742 students.

Fidelity

The level of simulation fidelity across the 26 studies ranged from high fidelity to low fidelity. High fidelity simulations included fully computerised manikins (n=7), virtual reality/simulation (n=2) and standardised patients (n=7). Low fidelity simulations included part task trainers/non-computerised manikins (n=2), scripted scenarios (n=6) and simulation stations (n=2).

Study outcomes

Study outcomes varied, based on very different simulation learning objectives. Twenty of 26 studies reported positive benefits of SBE and six studies found no benefit or else equivalency in terms of improvement. All the studies were founded on the instruments selected for evaluating learners' performance. As presented in Table 1, the majority of instruments used were related to program evaluation outcomes based on a variety of topics around learners' perceptions, rather than objective evaluation measures (such as knowledge, or observed behaviours). Subjective evaluations included self-ratings of clinical judgement (Lasater Clinical Judgement Rubric; Clinical Competency Tool); confidence and self-efficacy (State-trait Anxiety Inventory; Self-confidence Scale Revisited; Self-efficacy Survey; Nursing Satisfaction and Self-confidence with Learning Scale). The most numerous measures were those related to simulation course satisfaction (Simulation Effectiveness Scale; Patient Simulation Satisfaction Tool; Learning Climate Evaluation Tool). Several studies used both subjective and objective measures to complement their evaluation and

Table 2 – Methodological quality of simulation studies in nursing according to MERSQI and SRR

Study	MERSQI				SRR			
	Mean of Total Scores/18 Mean (%)	Inter-rater Correlation <i>r</i> (p-value)	Interrater Agreement (Kappa) <i>k</i> (p-value)	Assessed Study Quality	Mean of Total scores /56 Mean (%)	Inter-rater Correlation <i>r</i> (p-value)	Interrater agreement (Kappa) (<i>k</i> / p-value)	Assessed Study Quality
Andrea & Kotowski, 2017	8.50 (47.2)	1.00 (<0.001)	1.00 (<0.001)	Low	42.5 (75.9)	.698 (0.006)	.485 (0.003)	High
Beaird et al, 2017	16.25 (90.3)	.602 (0.066)	.643 (<0.001)	High	52.5 (93.7)	.055 (0.852)	.084 (0.837)	High
Boling et al, 2017	6.50 (36.0)	.747 (0.013)	.833 (<0.001)	Low	43.0 (76.8)	.692 (0.006)	.354 (0.041)	High
Booth et al, 2017	11.50 (63.9)	.481 (0.160)	.692 (<0.001)	Moderate	44.5 (79.5)	.795 (0.001)	.240 (0.168)	High
Cason et al, 2017	11.25 (62.5)	.881 (0.001)	.753 (<0.001)	Moderate	40.5 (72.3)	.841 (<0.001)	.661 (<0.001)	Moderate
Cooper et al, 2017	14.00 (77.8)	1.00 (<0.001)	1.00 (<0.001)	High	48/52 (92.3)	.587 (0.027)	.404 (0.010)	High
Gu et al, 2017	11.50 (63.9)	.839 (0.002)	.750 (<0.001)	Moderate	49.0 (87.5)	.841 (<0.001)	.661 (<0.001)	High
Haley et al, 2017	12.50 (69.4)	.872 (0.001)	.744 (<0.001)	Moderate	45.5 (81.2)	.905 (<0.001)	.873 (<0.001)	High
Ignacio et al, 2017	13.00 (72.2)	.850 (0.002)	.643 (<0.001)	Moderate	49.0 (87.5)	.705 (0.005)	.481 (0.012)	High
Kaplan et al, 2017	12.25 (68.0)	.764 (0.010)	.630 (<0.001)	Moderate	54.5 (97.3)	1.00 (1.00)	.481 (<0.001)	High
Kim et al, 2017	13.0 (72.2)	.950 (<0.001)	.878 (<0.001)	Moderate	50.5 (90.2)	.329 (0.251)	.468 (0.025)	High
Kubin & , 2017	13.0 (72.2)	.773 (0.009)	.747 (<0.001)	Moderate	52.0 (92.5)	.090 (0.125)	.125 (0.542)	High
Labant & Palmer, 2017	11.0 (61.0)	.900 (<0.001)	.759 (<0.001)	Moderate	41.5 (74.1)	.447 (0.109)	.037 (0.847)	Moderate
Lee et al, 2017	10.25 (56.9)	1.00 (<0.001)	1.00 (<0.001)	Moderate	48.0 (85.7)	.047 (0.873)	.069 (0.626)	High
Letcher et al, 2017	14.25 (79.2)	.618 (<0.001)	.595 (<0.001)	High	48.0 (85.7)	.166 (0.610)	.125 (0.263)	High
Luckter-Flude et al, 2017	11.0 (61.0)	.468 (0.173)	.500 (<0.001)	Moderate	47.5 (84.8)	.616 (0.190)	.097 (0.565)	High
Maharaj, 2017	11.0 (61.1)	.930 (<0.001)	.873 (<0.001)	Moderate	49.0 (87.5)	.282 (0.379)	.375 (0.350)	High
Mariani et al, 2017	12.0 (66.6)	.868 (0.001)	.744 (<0.001)	Moderate	50.0 (89.3)	.340 (0.235)	.300 (0.089)	High
McWilliams et al, 2017	10.5 (58.3)	.911 (<0.001)	.756 (<0.001)	Moderate	43.5 (77.7)	.412 (0.143)	.513 (0.010)	High
Murphy & Janisse, 2017	9.0 (50.0)	.848 (0.002)	.868 (<0.001)	Moderate	45.5 (81.2)	.664 (0.010)	.548 (0.012)	High
Ross & Carney, 2017	9.75 (54.2)	.686 (0.029)	.571 (<0.001)	Moderate	43.5 (77.7)	.169 (0.564)	.119 (0.159)	High
Sapyta & Eiger, 2017	11.5 (63.9)	.797 (0.006)	.518 (<0.001)	Moderate	43.5 (77.7)	.114 (0.698)	.217 (0.083)	High
Skinner, 2017	10.5 (58.3)	.906 (<0.001)	.878 (<0.001)	Moderate	39.0 (69.6)	0.00 (1.00)	.113 (0.447)	Moderate
Turkelson & Keiser, 2017	12.25 (68.1)	.774 (0.009)	.643 (<0.001)	Moderate	47.5 (84.8)	.553 (0.040)	.170 (0.200)	High
Verkuyt et al, 2017	13.25 (73.6)	.797 (0.006)	.634 (<0.001)	Moderate	53.0 (94.6)	.576 (0.031)	.576 (0.031)	High
Walters et al, 2017	11.5 (63.9)	.846 (0.002)	.872 (<0.001)	Moderate	45.0 (80.3)	.188 (0.519)	.045 (0.341)	High

NOTE: Quality outcomes are based on the quartiles within each rating scale: Low = ≤ 50% of score, Moderate = 51%-75%; High = 76%-100%. Each outcome is derived from the mean of ratings by two researchers, each of whom used one instrument to assess each study in the dataset.

others used purposely developed subjective or objective measures where validity or reliability mostly went unreported.

Higher level objective measures used in approximately 10 studies included measurements of learners' knowledge (multiple choice questionnaires; tests, or examination results), electronically collected data (haptic tool measurements in two cases) or direct observation by a trained observer using a valid checklist. Immediate post-intervention measurements were reported in nearly all studies and longer term impact was lacking. The 26 studies reported use of a wide range of measures as well as a wide range of topics, an indication of the curriculum areas where SBE is being applied.

Methodological quality evaluation

Table 2 presents the mean values for the quality evaluation for each of 26 studies according to MERSQI and SSR.

1. Quality according to MERSQI

The majority of studies ($n = 21$) were found to reflect a moderate methodological quality (50% to 75%; range 9.0 to 13.5 points of 18). Three received high ratings on quality ($>76\%$; >13.6 points of 18); two were rated low in quality, $\leq 50\%$ (≤ 9 points). Most scores attributed within the protocol were limited by one point or more because two outcome points were awarded to actions with 'real patients' or 'actual effects on real patients' that were not applicable.

Inter-rater agreement computed by the Kappa test showed good agreement in assessments in each study ($\geq .7$ in 16 studies and very good agreement ($\geq .8$) for eight studies). The reliability of Kappa scores was confirmed in further testing. Correlational analyses showed statistically significant correlations between raters for MERSQI scores in 24 of 26 studies. All but two studies had a large correlation ($r = \geq .5$) with the average quite high (mean $r = .81$) and twofold scores in three studies were perfectly correlated ($r = 1.0$). An Intra-class Correlation Co-efficient of .988 on average measures confirmed high overall consistency in the 52 ratings ($F = 81.520$, $df 9, 459$ [CI .974-.996], $p < 0.001$). Owing to variations in study scores within MERSQI domains (sub-sections) that were either generally well achieved, or else poorly achieved, we highlight some features below.

Sampling domain: Sampling from a larger number of institutions (1, 2 or >3) gained extra points, although the majority of studies sampled only one training institution. The sampling response rate was poorly recorded in 13 studies; 11 studies reported a high sampling response rate of >75% and earned higher points.

Data analysis domain: Data analysis was well reported with 21 studies providing evidence of analysis that went beyond descriptive analysis to include inferential tests such as t-tests or ANOVA, or a non-parametric equivalent.

Validity evidence domain: The validity evidence for instruments that were utilized was lacking. This included a lack of description of relevant content, evidence and explanation; internal structure evidence (such as reliability and validity, inter-rater, prior test-retest, or factor analysis) and relationships to other variables (such as concurrent or predictor correlations with other variables- regression, etc). The scores in 12 studies were lacking regarding content evidence, and validity evidence was lacking in 13 studies. Only three studies reported relationships with other variables. Reporting in this domain that potentially could have earned up to three points, often earned low scores.

Outcomes domain: Most studies reported objective outcomes at the level of knowledge tests and skills improvement, earning 1.5 points. Three studies used self-reported measures alone and earned lower points.

2. Quality according to SRR

The SRR is focused on evaluating the processes used in developing and delivering a SBE program to learners. Nearly all studies (n= 23 of 26) were assessed as high in quality ($\geq 75\%$ of a possible 56 points; range 44-56 points). Three were rated moderate quality (50% to 75%; 28-43 points); none were rated low in quality (Table 2).

In approximately half the studies (12/26) Kappa measures of agreement between raters were significantly correlated. Of these, six studies showed moderate agreement ($\geq .5$) with others lower: the range was $k=.354$ to $k=.868$. Significant inter-rater correlations (r) occurred in only 12 of 26 studies, with five being strongly correlated ($r \geq .7$). The differences suggest a likelihood of variance increases when a rating comprises 16 variables, each rated on a 4-point Likert scale and based on decisions about compliance with the assessment rubric. Despite differences, high overall consistency in the 52 assessments was confirmed by

a significant Intra-class Correlation Co-efficient (ICC) of .941 on average measures ($F=16.967$, df 13, 663 [CI .887-.977], $p<0.001$) although the individual study correlations were low (ICC: .235).

The majority of studies earned ratings of 3 (very good) or 4 (excellent) for most criteria, however a lack of description of the whole breadth of a program, with a criterion missed or poorly described, meant that points could be lower. Highest scores were awarded for experimental studies that used valid instruments and reported a power calculation. Description of development of the simulation and description of debriefing/feedback methods were two areas where detail was often lacking. However, depending on the research design not all variables were applicable to each study and there is a need to consider adding a 'not applicable' response option in place of the 'unsatisfactory' option (score = 0). Furthermore, it was difficult to discriminate between ratings 2 (good), 3 (very good) and 4 (excellent) in some of the descriptive information owing to varying research foci.

3 Comparison between quality ratings using MERSQI and SRR

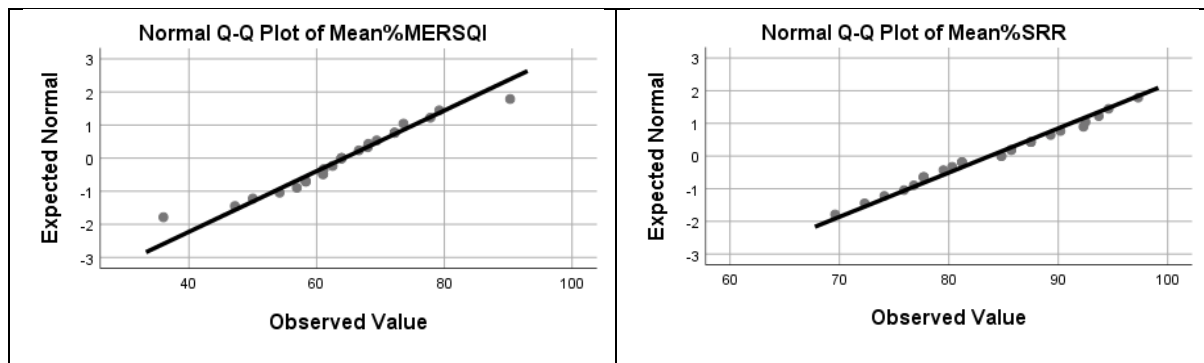
Although there was a large correlation between the proportional (%) mean total MERSQI scores and the mean total SRR scores (.716, $p<0.001$), a significant difference was identified between the two instruments' in ranking the overall quality of studies ($p<0.001$). As seen in Table 3, the SRR ranked the proportional total scores almost 20% higher on average than the MERSQI (83.7% versus 64.3% respectively) with approximately half the variance (range: 27.7%, versus 54.3%).

Table 3 Comparison between quality ratings using proportional MERSQI and SRR scores

	MERSQI Total score (%/SD)	SSR Total score (%/SD)	Z (p-value)
Percent Mean Total Score	64.3 (10.9) (CI: 59.9%, 68.7%)	83.7 (7.38) (CI: 80.8%, 86.7%)	-5.354 (<0.001)
Percent Median Score	63.9%	84.8%	
Percent Total Score (Range)	36.0% - 90.3%	69.6% - 97.3%	
Interquartile range	13.9%	11.8%	
<u>Quality Outcome</u>			
High (n/%)	2 (7.7)	23 (88.5)	
Moderate (n/%)	21 (80.8)	3 (11.5)	
Low (n/%)	3 (11.5)	0 (0.0)	

The quartile cut points for MERSQI were 58.3% (25th), 63.9% (50th), and 72.0% (75th), while those for SRR were higher: 77.7%, 84.8%, 89.5%. Thus, the magnitude of assessed study quality differed between instruments. However, Figure 2 shows that both instruments produced valid plots with linearity between the observed and expected values, with few outliers.

Figure 2 Plot of proportional instrument scores showing trends in observed versus expected values (n=26)



DISCUSSION

This review found that the methodological quality of included studies, according to the Medical Education Research Quality Instrument (MERSQI), ranged from moderate to high (M= 64%). Although the Simulation Research Evaluation Rubric (SRR) ratings also ranged from moderate to high, the overall scores were approximately 20% higher (M= 84%).

The MERSQI is a generalist instrument for rating the quality of medical education studies across a broad range of designs and topics, while the SSR is a simulation-specific instrument which tests best practice in simulation design and delivery; hence the instruments rank somewhat different variables. Two studies ranked as low quality by MERSQI were ranked as high quality by SRR and this juxtapositioning suggests the need to review whether the SRR rating system adequately discriminates between research methods. For example, as SRR is focused on development and implementation of a simulation program but not on the objectivity of measures used, studies may be awarded higher points on study instruments. Nevertheless, the reliability of both the MERSQI and the SRR was confirmed by strong intra-rater and inter-rater correlations ($\geq .94$) for the 26 studies under review. We therefore suggest that both instruments are feasible and suited for use in quality evaluations of

simulation research. Further guidance on how to report simulation studies is provided in a 24-item checklist suggested as best reporting practice: extensions to the CONSORT and STROBE statements (Cheng, et al., 2016)

Variability in the assessed levels of quality in the included studies was often related to study design, with stronger research designs resulting in higher quality scores. It became apparent from the methodological assessment that the best level of evidence is derived from experimental studies that sample a random selection of participants and compare outcomes with a second group who do not receive the same intervention. Only three of the included studies presented this level of evidence (Gu, Zou, & Chen, 2017; McWilliams, Malecha, Langford, & Clutter, 2017; Verkuyl, Romaniuk, Atack, & Mastrilli, 2017; the majority were quasi-experimental, single group or two-group pretest and post-test designs. It is reasonable for these designs to be awarded lower rating scores because it is likely that observations taken before and after an SBE intervention are influenced by extraneous factors that threaten internal validity. It was noted there were less stringent rating requirements in some variables in the SRR- eg., regarding instruments used, there was no scoring distinction between objective or subjective measurements. Other designs that were used such as post-test only designs with no comparison group can, at best, describe features of an educational intervention but do not measure the effect on learning.

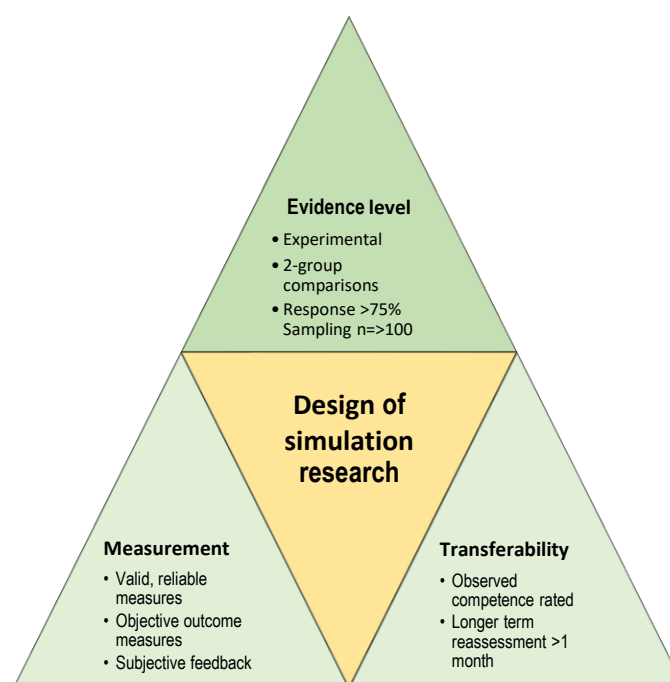


Figure 3 Schema for inclusions in high quality simulation-based education research designs

In Figure 3 we present a schema showing indicators of high quality research in SBE designs, drawn from the current findings.

A major discussion point in the design of SBE programs is the choice of measures and assessment instruments that provide evidence of program impact. This review found that the majority of studies gained a maximum score for the quality of data collection instruments, as 18/26 studies chose objective measures to report the impact of SBE (eg, knowledge tests, examinations, observations of performance).

Eight studies chose self-reported measures. Even though self-reported measures may be validated instruments (such as a Self-Efficacy Survey), self-reported evidence is less reliable because it is prone to bias (Polit & Beck, 2012). Response bias can be introduced by, for example, a social desirability effect as participants attempt to present themselves in the best possible light. There is also evidence that students rate their performance more highly than faculty observers (Kubin & Wilson, 2017). Self-report, when used as a proxy measure for performance, can be weak – for example, ‘confidence’ as a proxy for clinical competence has been found not to correlate with faculty-rated clinical competence (Liaw, Scherpbier, Rethans, & Klainin-Yobas, 2012). Self-reported measurements are best regarded *as an indicator* and studies using these approaches as intervention outcomes are likely to receive lower scores on quality.

Incorporation of program evaluation outcome measures such as participant satisfaction were also frequently used in the reviewed studies (eg, Simulation Effectiveness Scale; Patient Simulation Satisfaction Tool). Although learner satisfaction is often considered to be the ‘low hanging fruit’ in evaluation (Adamson, Kardong-Edgren, & Willhaus, 2013), research suggests that satisfaction influences engagement in learning and level of self-confidence in the ability to provide patient care (Khalaila, 2014; Lubbers, & Rossman, 2017). Furthermore, feedback from participants is also desirable as a quality improvement measure, used to identify areas of the SBE program that may need enhancement (Lioce et al., 2015). For these reasons we recommend that a combination of objective measures of program impact (effect) and subjective program evaluation outcome measures (participant feedback) are appropriate, and both should be used in the evaluation of SBE.

Limitations

This study extracted a purposive sample of quantitative studies regarded as current examples of simulation research, but we acknowledge that the included studies may not represent SBE reported in journals other than CSN or in fields other than nursing. While valid instruments were chosen to rate the methodological quality of studies, there may still have been some subjectivity in the authors' assessments, which may have affected rating and ranking of data. Nevertheless, this review provides a snapshot of current reporting of SBE research and gives some guidance for future improvement in the quality of simulation research designs.

CONCLUSION

Research into simulation based education (SBE) in nursing has escalated exponentially in the last decade along with the rapid increase in the use of simulation more generally. The 26 primary nursing studies included in this review provide an overview of the broad range of programs where SBE is currently being utilized. Although the included studies were of moderate to high methodological quality (M=64%, M=87%), those rated the highest using the MERSQI and SRR were experimental studies. While we acknowledge the importance of using a combination of objective measures of program impact (effect) and the more subjective program evaluation outcome measures (feedback), we recommend that future research aiming to determine the true impact of SBE should consider experimental designs to limit threats to external and internal validity.

References

- Adamson, K. A., Kardong-Edgren, S., & Willhaus, J. (2013). An updated review of published simulation evaluation instruments. *Clinical Simulation in Nursing*, 9(9), e393-e400. <http://dx.doi.org/310.1016/j.ecns.2012.1009.1004>.
- Andrea, J., & Kotowski, P. (2017). Using standardized patients in an undergraduate nursing health assessment class. *Clinical Simulation in Nursing*, 13(7), 309-313. doi: <https://doi.org/10.1016/j.ecns.2017.05.003>
- Beaird, G., Nye, C., & Thacker, L. R. (2017). The use of video recording and standardized patient feedback to improve communication performance in undergraduate nursing students. *Clinical Simulation in Nursing*, 13(4), 176-185. doi: <https://doi.org/10.1016/j.ecns.2016.12.005>
- Boling, B., Hardin-Pierce, M., Jensen, L., & Hassan, Z.-U. (2017). Implementing simulation training for new cardiothoracic intensive care unit nurses. *Clinical Simulation in Nursing*, 13(1), 33-38.e12. doi: <https://doi.org/10.1016/j.ecns.2016.09.001>
- Booth, R. G., Sinclair, B., Strudwick, G., Brennan, L., Tong, J., Relouw, H., . . . & Vlastic, W. (2017). Identifying error types made by nursing students using eMAR technology. *Clinical Simulation in Nursing*, 13(10), 492-500. doi: <https://doi.org/10.1016/j.ecns.2017.05.016>
- Braga, M., Tyler, M. D., Rhoads, J. M., Cacchio, M. P., Auerbach, M., Nishisaki, A., & Larson, R. J. (2015). Effect of just-in-time simulation training on provider performance and patient outcomes for clinical procedures: a systematic review. *BMJ Simulation and Technology Enhanced Learning*, 1, 94-102.
- Breymier, T. L., Rutherford-Hemming, T., Horsley, T. L., Atz, T., Smith, L. G., Badowski, D., & Connor, K. (2015). Substitution of clinical experience with simulation in prelicensure nursing programs: a national survey in the United States. *Clinical Simulation in Nursing*, 11(11), 472-478. doi: [10.1016/j.ecns.2015.09.004](https://doi.org/10.1016/j.ecns.2015.09.004)
- Cant, R. P., & Cooper, S. J. (2017). The value of simulation-based learning in pre-licensure nurse education: A state-of-the-art review and meta-analysis. *Nurse Education in Practice*, 27, 45-62. doi: [10.1016/j.nepr.2017.1008.1012](https://doi.org/10.1016/j.nepr.2017.1008.1012)
- Cantrell, M. A., Franklin, A., Leighton, K., & Carlson, A. (2017). The evidence in simulation-based learning experiences in nursing education and practice: an umbrella review. *Clinical Simulation in Nursing*, 13(12), 634-667. doi: [10.1016/j.ecns.2017.08.004](https://doi.org/10.1016/j.ecns.2017.08.004)
- Cason, M., Atz, T., & Horton, L. F. (2017). New nursing graduates' self-efficacy ratings and urinary catheterization skills in a high-fidelity simulation scenario. *Clinical Simulation in Nursing*, 13(2), 71-77. doi: <https://doi.org/10.1016/j.ecns.2016.12.006>
- Cheng, A., Kessler, D., Mackinnon, R., Chang, T. P., Nadkarni, V. M., Hunt, E. A., ... & International Network for Simulation-based Pediatric Innovation, Research & Education (INSPIRE) Reporting Guidelines Investigators. (2016). Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Advances in Simulation*, 1, 1:25. DOI [10.1186/s41077-41016-40025-y](https://doi.org/10.1186/s41077-41016-40025-y).
- Cook, D. A., & Reed, D. A. (2015). Appraising the quality of medical education research methods: The Medical Education Research Study Quality Instrument and the Newcastle–Ottawa Scale–Education. *Academic Medicine*, 90(8), 1067-1076. doi: [10.1097/acm.0000000000000786](https://doi.org/10.1097/acm.0000000000000786)
- DeCoste-Lopez, J., Madhok, J., & Harman, S. (2015). Curricular innovations for medical students in palliative and end-of-life care: a systematic review and assessment of study quality. *Journal of Palliative Medicine*, 18, 338-349. <https://doi.org/310.1089/jpm.2014.0270>.
- Doolen, J., Mariani, B., Atz, T., Horsley, T. L., Rourke, J. O., McAfee, K., & Cross, C. L. (2017). High-fidelity simulation in undergraduate nursing education: a review of simulation reviews. *Clinical Simulation in Nursing*, 12(7), 290-302. doi: [10.1016/j.ecns.2016.01.009](https://doi.org/10.1016/j.ecns.2016.01.009)
- Fey, M. K., Gloe, D., & Mariani, B. (2015). Assessing the quality of simulation-based research articles: a rating rubric. *Clinical Simulation in Nursing*, 11(12), 496-504. doi: [10.1016/j.ecns.2015.10.005](https://doi.org/10.1016/j.ecns.2015.10.005)
- Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91-108. doi: [10.1111/j.1471-1842.2009.00848.x](https://doi.org/10.1111/j.1471-1842.2009.00848.x)
- Gu, Y., Zou, Z., & Chen, X. (2017). The effects of vSIM for Nursing™ as a teaching strategy on fundamentals of nursing education in undergraduates. *Clinical Simulation in Nursing*, 13(4), 194-197. doi: <https://doi.org/10.1016/j.ecns.2017.01.005>
- Haley, B., Heo, S., Wright, P., Barone, C., Rettigantid, M. R., & Anders, M. (2017). Effects of using an advancing care excellence for seniors simulation scenario on nursing student empathy: a randomized controlled trial. *Clinical Simulation in Nursing*, 13(10), 511-519. doi: <https://doi.org/10.1016/j.ecns.2017.06.003>

- Hayden, J., Smiley, R. A., Alexander, M., Kardong-Edgren, S., & Jeffries, P. R. (2014). The NCSBN National Simulation Study: A longitudinal, randomized, controlled study replacing clinical hours with simulation in prelicensure nursing education. *Journal of Nursing Regulation*, 5, C1-S64.
- Hegland, P. A., Aarlie, H., Strømme, H., & Jamtvedt, G. (2017). Simulation-based training for nurses: Systematic review and meta-analysis. *Nurse Education Today*, 54, 6-20. doi: <https://doi.org/10.1016/j.nedt.2017.04.004>
- Hsieh, M.-J., Bhanji, F., Chiang, W.-C., Yang, C.-W., Chien, K.-L., & Ma, M. H.-M. (2016). Comparing the effect of self-instruction with that of traditional instruction in basic life support courses—A systematic review. *Resuscitation*, 108(Suppl C), 8-19. doi: <https://doi.org/10.1016/j.resuscitation.2016.08.021>
- IBM Corp. (2016). IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.
- Ignacio, J., Scherpbier, A., Dolmans, D., Rethans, J.-J., & Liaw, S. Y. (2017). Mental rehearsal strategy for stress management and performance in simulations. *Clinical Simulation in Nursing*, 13(7), 295-302. doi: <https://doi.org/10.1016/j.ecns.2017.04.005>
- Ilgen, J. S., Sherbino, J., & Cook, D. A. (2013). Technology-enhanced simulation in emergency medicine: a systematic review and meta-analysis. *Academic Emergency Medicine*, 20(2), 117-127. doi: 10.1111/acem.12076
- INACSL. (2016). INACSL Standards of Best Practice: SimulationSM Debriefing. *Clinical Simulation in Nursing*, 12, S21-S25.
- Kaplan, B., Murihead, L., & Zhang, W. (2017). Leveraging partnerships: nursing student veteran-centered simulation in situ. *Clinical Simulation in Nursing*, 13(6), 258-263. doi: <https://doi.org/10.1016/j.ecns.2017.02.001>
- Khalaila, R. (2014). Simulation in nursing education: An evaluation of students' outcomes at their first clinical practice combined with simulations. *Nurse Education Today*, 34, 252-258.
- Kim, Y.-J., Noh, G.-O., & Im, Y.-S. (2017). Effect of step-based prebriefing activities on flow and clinical competency of nursing students in simulation-based education. *Clinical Simulation in Nursing*, 13(11), 544-551. doi: <https://doi.org/10.1016/j.ecns.2017.06.005>
- Kubin, L., & Wilson, C. E. (2017). Effects of community volunteer children on student pediatric assessment behaviors. *Clinical Simulation in Nursing*, 13(7), 303-308. doi: <https://doi.org/10.1016/j.ecns.2017.04.011>
- Labant, A. L., & Palmer, E. A. (2017). Teaching neonatal abstinence syndrome using simulation. *Clinical Simulation in Nursing*, 13(6), 254-257. doi: <https://doi.org/10.1016/j.ecns.2017.01.009>
- Lee, A. H., Kelley, C., Alfes, C. M., Bennington, L. K., & Dolansky, M. A. (2017). High-fidelity patient simulation to evaluate student nurse patient safety competency. *Clinical Simulation in Nursing*, 13(12), 628-633. doi: <https://doi.org/10.1016/j.ecns.2017.08.006>
- Letcher, D. C., Roth, S. J., & Varenhorst, L. J. (2017). Simulation-based learning: improving knowledge and clinical judgment within the NICU. *Clinical Simulation in Nursing*, 13(6), 284-290. doi: <https://doi.org/10.1016/j.ecns.2017.03.001>
- Levett-Jones, T., & Lapkin, S. (2014). A systematic review of the effectiveness of simulation debriefing in health professional education. *Nurse Education Today*, 34(6), e58-e63 DOI: <https://doi.org/10.1016/j.nedt.2013.1009.1020>
- Liaw, S. Y., Scherpbier, A., Rethans, J.-J., & Klainin-Yobas, P. (2012). Assessment for simulation learning outcomes: A comparison of knowledge and self-reported confidence with observed clinical performance. *Nurse Education Today*, 32(6), e35-e39. doi: 10.1016/j.nedt.2011.10.006
- Lioce, L., Meakim, C. H., Fey, M. K., Chmil, J. V., Mariani, B., & Alinier, G. (2015). Standards of best practice: Simulation standard IX: Simulation design. *Clinical Simulation in Nursing*, 11(6), 309-315.
- Lubbers, J. & Rossman, C. (2017). Satisfaction and self-confidence with nursing clinical simulation: Novice learners, medium-fidelity, and community settings. *Nurse Education Today*, 48, 140-144.
- Luctkar-Flude, M., Wilson-Keates, B., Tyerman, J., Larocque, M., & Brown, C. A. (2017). Comparing instructor-led versus student-led simulation facilitation methods for novice nursing students. *Clinical Simulation in Nursing*, 13(6), 264-269. doi: <https://doi.org/10.1016/j.ecns.2017.03.002>
- Maharaj, T. (2017). Live-model simulation: improving nursing students' attitudes and knowledge of Alzheimer's disease. *Clinical Simulation in Nursing*, 13(9), 446-451. doi: <https://doi.org/10.1016/j.ecns.2017.05.002>
- Mariani, B., Ross, J. G., Paparella, S., & Allen, L. R. (2017). Medication safety simulation to assess student knowledge and competence. *Clinical Simulation in Nursing*, 13(5), 210-216. doi: <https://doi.org/10.1016/j.ecns.2017.01.003>
- McKinney, J., Cook, D.A., Wood, D., & Hatala, R. (2013). Simulation-based training for cardiac auscultation skills: systematic review and meta-analysis. *Journal of General Internal Medicine*, 28, 283 DOI <https://doi.org/10.1007/s11606-11012-12198-y>.

- McWilliams, L. A., Malecha, A., Langford, R., & Clutter, P. (2017). Comparisons of cooperative-based versus independent learning while using a Haptic Intravenous Simulator. *Clinical Simulation in Nursing*, 13(4), 154-160. doi: <https://doi.org/10.1016/j.ecns.2016.12.008>
- Motola, I., Devine, L., Chung, H., Sullivan, J., & Issenberg, B. (2013). Simulation in healthcare education: A best evidence practical guide. AMEE Guide No. 82. *Medical Teacher*, 35, e1511–e1530.
- Murphy, L. J., & Janisse, L. (2017). Optimizing transition to practice through orientation: a quality improvement initiative. *Clinical Simulation in Nursing*, 13(11), 583-590. doi: <https://doi.org/10.1016/j.ecns.2017.07.007>
- National Council of State Boards of Nursing. (2010). Simulation use in prelicensure nursing programs. Chicago, IL.
- Nursing and Midwifery Council. (2010). Standards for Preregistration Nursing Education. NMC, London.
- Pallant, J. (2013). *SPSS Survival Manual; a step by step guide to data analysis using SPSS* (5th ed.). Sydney: Allen and Unwin.
- Polit, D., & Beck, C. (2012). *Nursing Research- Generating and Assessing Evidence for Nursing Practice* (9th ed.). Philadelphia: Wolters Kluwer.
- Reed, D. A., Cook, D.A., Beckman, T.J., Levine, R.B., Kern, D.E., & Wright, S.M. (2007). Association between funding and quality of published medical education research. *JAMA*, 298, 1002-1009.
- Reed, D. A., Beckman, T. J., & Wright, S. M. (2009). An assessment of the methodologic quality of medical education research studies published in The American Journal of Surgery. *The American Journal of Surgery*, 198(3), 442-444. doi: <https://doi.org/10.1016/j.amjsurg.2009.01.024>
- Ross, J. G., & Carney, H. (2017). The effect of formative capstone simulation scenarios on novice nursing students' anxiety and self-confidence related to initial clinical practicum. *Clinical Simulation in Nursing*, 13(3), 116-120. doi: <https://doi.org/10.1016/j.ecns.2016.11.001>
- Rutherford-Hemming, T., & Alfes, C. M. (2017). The use of hospital-based simulation in nursing education—a systematic review. *Clinical Simulation in Nursing*, 13(2), 78-89. doi: <https://doi.org/10.1016/j.ecns.2016.12.007>
- Sapya, Y. Y., & Eiger, C. (2017). Improving pediatric nurses' knowledge, accuracy, and confidence through code documentation simulation. *Clinical Simulation in Nursing*, 13(6), 278-283. doi: <https://doi.org/10.1016/j.ecns.2017.02.003>
- Skinner, H. M. (2017). Simulation: preparing nursing students to work with community-dwelling older adults. *Clinical Simulation in Nursing*, 13(10), 520-523. doi: <https://doi.org/10.1016/j.ecns.2017.04.012>
- Turkelson, C., & Keiser, M. (2017). Using checklists and repetitive simulation to improve patient safety: a pilot project with the Impella® Left Ventricular Assist Device. *Clinical Simulation in Nursing*, 13(2), 53-63. doi: <https://doi.org/10.1016/j.ecns.2016.10.009>
- Verkuyl, M., Romaniuk, D., Atack, L., & Mastrilli, P. (2017). Virtual Gaming Simulation for nursing education: an experiment. *Clinical Simulation in Nursing*, 13(5), 238-244. doi: <https://doi.org/10.1016/j.ecns.2017.02.004>
- Walters, B., Potetz, J., & Fedesco, H. N. (2017). Simulations in the classroom: an innovative active learning experience. *Clinical Simulation in Nursing*, 13(12), 609-615. doi: <https://doi.org/10.1016/j.ecns.2017.07.009>

Citation: Cant, R.P., Levett-Jones, T., James, A. Do simulation studies measure up? A simulation study quality review. *Clinical Simulation in Nursing*, 2018; 21: 23-39. At: <https://doi.org/10.1016/j.ecns.2018.06.002>.

Table 1 – Characteristics and findings of quantitative simulation studies in nursing

Study/origin	Study design and mode of simulation	Topic	Study sample	Measures	Findings/Implications
Andrea & Kotowski, 2017 Chicago, USA.	Quantitative design. Three simulation scenarios with standardized patients.	Health assessment: Obtaining a health history from patients.	n=80 first semester baccalaureate nursing students in groups of 5-6 students in each (sample collected over 3 semesters).	Confidence, communication skills, clinical judgment (Lasater Clinical Judgment Rubric (LCJR)) at pretest, during placement and one week after simulation.	Participant's level of confidence increased.
Beaird, Nye, & Thacker, 2017 Virginia, USA.	Prospective, repeated measures survey design. Study participants were randomized into video-recording and non-video-recording groups.	Use of video recording and standardized patient feedback to improve communication performance.	n=94 pre-registration nursing students taking a women's health course.	Adapted Macy Communication Scale: SPs rated students' performance, and students used the same tool to rate their own performance.	Intentional communication training using SP feedback improved communication skills performance.
Boling, et al, 2017 Kentucky, USA.	Observational study: development and pilot of training based on Jeffrey's Nursing Education Simulation Framework, using high fidelity manikins.	Cardiothoracic simulation.	n=11 qualified intensive care nurses working in two groups of 5-6, over 4 hours.	13-item valid tool: Simulation Evaluation Tool completed at end of session.	Participants valued the simulation for learning and felt it was more beneficial than listening to a lecture on the topic. Creating a custom program is effective in the training of new nurses.
Booth et al., 2017 Canada	Observational study: Simulated scenario of medication administration to a stable medical-surgical patient.	Identifying medication error types made by nursing students using eMAR (electronic medication administration record) technology.	n=25 nursing students	Purposely developed observation guide and recorded interview (debrief)	Students generated a variety of errors during the medication administration process, likely due to a range of individual, contextual, and eMAR factors. Although students completed online learning activities prior, the fidelity of these learning activities was possibly insufficient to prepare students to use the eMAR system.
Cason et al 2017 Charleston, USA.	Descriptive correlational design.	Self-efficacy ratings in urinary catheterization skills	n=47 new nursing graduates.	Self-reported Clinical Competency Questionnaire (CCQ)	No significant association between self-efficacy and competency in the skill of urinary catheterization.

	High fidelity manikins.			(46 items on professional behavior, general performance, core nursing skills, advanced nursing); observational assessment used Creighton Competency Evaluation Instrument (C-CEI).	
Cooper et al, 2017 Australia	Quasi-experimental pre-test, post-test design. Virtual interactive simulation (three e-simulations with standardized patients).	Impact of an e-simulation program on management of patient deterioration.	n=2,971: n=1,229 qualified nurses and n= 1,742 final year nursing students (international sample).	Knowledge (11-item questionnaire) and performance ratings (rated on-line), confidence survey, post session evaluation.	E-simulation may enhance students' preparation for practice and improve qualified nurses' management of deteriorating patients.
Gu et al 2017 China	Experimental, randomized controlled, posttest design. 10 virtual interactive cases added to curriculum.	Effects of vSIM for Nursing™ as a teaching strategy.	n= 28 undergraduates were assigned and n=13 completed the intervention, n=14 in the control group.	Post-test knowledge	The experimental group had significantly higher knowledge scores than those in the control group. vSIM for Nursing might be an effective supplementary teaching strategy to improve students' knowledge of fundamentals of nursing.
Haley et al 2017 Arkansas USA	Quasi-experimental post-test design. An Advancing Care Excellence for Seniors simulation scenario.	Nursing student empathy and self-awareness.	n=50 nursing students.	Self-Consciousness Scale Revised (22-item Likert-type scale to measure self-reported self-awareness); Kiersma Chen Empathy Scale (15-item Likert-type scale to measure self-reported empathy); Patient Practitioner Orientation Scale (18-item Likert-type scale used to measure self-reported PCC).	Active listening, empathy, and PCC scores in the intervention group compared with the control group and baseline significantly differed at post intervention and four-week follow-up (ps < .001). Self-awareness scores between groups were significantly different only at four-week follow-up (p < .001).
Ignacio et al 2017 Singapore	Mixed methods design. Use of a mnemonic strategy with manikin simulation to enhance patient deterioration management.	Mental rehearsal strategy for stress management and performance using simulation.	n=35 final year nursing students; n= 21 completed interviews.	Strait-Trait Anxiety Inventory, heart rate, systolic blood pressure, and skin temperature were used as stress measures. Post-test interviews.	Both MR and the mnemonic strategies had similar effects on performance and stress during patient deterioration management.
Kaplan et al 017	Quasi-experimental pre-post-test design.	Congestive heart failure - Nursing student veteran-	n=23 nursing students, both undergraduate and postgraduate entry to BSN.	Knowledge: pretest/posttest consisted of 14 items to measure level of knowledge acquisition;	Knowledge tests improved by 16%. The realism added by in situ simulation is

Atlanta USA	Veteran-centered care simulation scenarios over 3 weeks, 30 mins scripted scenario with manikin each week followed by group debrief.	centered simulation in situ.		the Medical Education Technologies, Inc. Simulation Effectiveness Tool was used to measure student confidence and learning.	valuable and may be suitable for expanded use.
Kim et al 2017 South Korea	Quasi-experimental, non-equivalent control group, non-synchronized design. Three different pre-briefing steps.	Step-based prebriefing activities on flow and clinical competency of nursing students	n= 207 junior/senior nursing students (control group n=62; experimental group 1 = 67, and group 2 = 76.	10-item Flow Short Scale; Korean version of Clinical Competency Self-reported Tool completed by nursing students and an evaluation tool by an instructor.	The second experimental group showed the highest amount of flow, satisfaction, and self-confidence. In simulation-based education; several prebriefing activities should be developed and integrated.
Kubin & Wilson 2017 Texas, USA	Pre-post-test two-group design. One group of participants practiced pediatric assessment on high-fidelity simulators and a second group practiced assessment on community volunteer children.	Pediatric assessment behaviors	N= 99 undergraduate nursing students in a pediatric nursing course.	Students were administered the Pediatric Student Comfort and Worry Assessment Tool at the beginning of semester and following intervention. Students self-evaluated and faculty-evaluated completing a pediatric assessment using the Effective Noticing and Responding domains of the Lasater Clinical Judgment Rubric.	Overall, students had similar worry and comfort scores regardless of group; additionally, assessment performances were similar.
Labant & Palmer, 2017 Indiana USA	Quasi-experimental, non-equivalent control group design. Experimental group received a lecture/video and hands-on training using a high-fidelity simulation manikin. Control group had traditional didactic and video instruction.	Comparing whether simulation training is more effective than traditional didactic and video instruction in teaching nursing students' assessment skills for scoring Neonatal Abstinence Syndrome (no illicit drugs).	n=26 nursing students.	NAS scoring system (student scores were compared with the nurse expert rater NAS scores)	No significant differences in NAS scores were found between the experimental group and the control group.
Lee et al 2017 Ohio USA	Observational study. High-fidelity patient simulation to determine nursing safety competence	Basic patient safety competency	n= 52 undergraduate BSN nursing students.	Students were observed and evaluated on their ability to identify or perform six basic safe patient care competencies (such as hand hygiene, patient ID, appropriate communication).	Only 19% of students were competent in all six patient safety competencies.

Letcher, 2017) South Dakota USA	Quasi-experimental pre–post-test design. Three sessions of simulated scenarios and structured debriefing.	Simulation-based learning: improving knowledge and clinical judgment within the NICU.	n=130 nurses	Lasater Clinical Judgment Rubric and patient indicators.	Differences in knowledge scores were seen: p = .0167 (Year 1) and p = .0021 (Year 2), together with clinical outcome trends of less ventilator days, increased utilization of alternative oxygen delivery methods, and stable intraventricular hemorrhage rate.
Luckter et al 2017 Canada	Mixed-methods crossover study. Instructor-led simulation with in-scenario feedback and postscenario debriefing and student-led simulation with postscenario debriefing only.	Comparing Instructor-led versus student-led health assessment knowledge and satisfaction with simulation facilitation methods.	n=114 nursing students completed the training and n=90 novice nursing students completed the evaluation.	11-item Patient Simulation Satisfaction subscale of the Health Assessment Educational Modality Evaluation survey; Facilitation Style Preference Survey, multiple choice quiz, and qualitative feedback on likes/dislikes about facilitation styles.	Novice learners preferred instructor-led to student-led simulation (p<.001); there was no association between simulation facilitation methods and knowledge scores.
Maharaj 2017 Texas USA	Quasi-experimental pre–post test two- group design. Standardized patient simulation (SPs).	Nursing students' attitudes and knowledge of Alzheimer's disease	n=65 bacclaureate senior nursing students	Alzheimer's Disease Knowledge Scale (ADKS) measured Alzheimer's knowledge and Dementia Attitude Scale (DAS) measured student attitudes; both surveys were repeated at the end of session.	Lectures and simulation experience had a positive effect on Alzheimer's Disease knowledge and attitude.
Mariani et al, 2017 Philadelphia USA	Experimental two-group pretest-posttest design. Medication Safety Simulation (1.25 hours).	Knowledge, competency, and perceptions of medication safety	n = 86 junior-level medical-surgical baccalaureate nursing students.	Medication Safety Knowledge Assessment (MSKA) and the Healthcare Professionals Patient Safety Assessment (HPPSA) completed at beginning of semester and post program; competency was observed using Medication Safety Critical Element Checklist (MSCEC).	Statistically significant differences in student knowledge (MSKA) and competency (MSCEC) were seen for students who participated in the medication safety enhanced simulations.
McWilliams et al, 2017 Texas USA	Post-test-only experimental design (using four random groups). IV simulator (virtual haptic simulator) for comparisons of cooperative-based versus independent learning.	Performance scores and number of IV cannulation attempts.	n=180 nursing students were randomized into four group assignments.	Initial performance score earned by each learner on the haptic IV simulator, and the number of attempts to earn a passing performance score were recorded, downloaded from the program and analysed.	Cooperative team members performed better with fewer attempts than independent learners when using an IV simulator; learning from observing and helping one another.

Murphy & Janisse, 2017 Canada	Quasi-experimental two-group evaluation. Three days of multiple simulation stations in a nurse orientation program.	Competencies and confidence.	n= 749 new nurse hires participated in an orientation program in 2013-14, and n=521 during 2015-16.	Quantitative data sources included human resources and summative assessment of de-identified data; comparative analysis between year groups.	Recent new hires demonstrated improved entry-level competencies and confidence through experiential learning using simulation education. When re-tested there was an improved competency pass rate.
Ross & Carney 2017 New Jersey USA	Quasi-experimental pre-post design. 4-hr simulation workshop - Formative Capstone Simulation Scenarios (three simulation scenarios prior to first clinical practicum.	Anxiety and self-confidence related to initial clinical practicum.	N = 95 novice nursing (sophomore) baccalaureate nursing students.	Spielberger State-Trait Anxiety Inventory (STAI) State Form Y used to assess current state anxiety related to initial clinical practicum; Nursing Anxiety and Self-Confidence with Clinical Decision-Making Scale (NASC-CDM) rated anxiety and self-confidence related to clinical decision making.	State anxiety and anxiety related to clinical decision making were significantly lower following formative capstone simulation scenarios ($p < .001$), and self-confidence related to clinical decision making was significantly higher following formative capstone simulation scenarios ($p < .001$).
Sapya & Eiger 2017 Chicago USA	One-group pre- and post-test quasi-experimental design. Two code simulations (using baby manikin and code documentation).	Nurses' knowledge, accuracy, and confidence.	n=48 pediatric acute care nurses from three units.	Purposely developed knowledge test and a code documentation checklist (8 items).	A statistically significant increase in knowledge, documentation accuracy and confidence was seen after the training program ($p < .001$).
Skinner 2017 Missouri USA	Quasi-experimental, mixed-method, one-group study design. A simulation of community-dwelling older adults.	Impact of simulation on students' preparedness to work with older people.	n= 23 senior nursing students (over two semesters)	Facts on Aging Quiz pre- and post-test score, and student interviews at end of course.	Scores were not statistically improved, however students indicated greater willingness to work with older people.
Turkelson & Keiser 2017 Minnesota USA	Quasi-experimental pre-post repeated-measure design. Impella® Left Ventricular Assist Device	Effectiveness of crisis checklists and repetitive simulations on patient/manikin outcomes and adherence to best practice guidelines.	n = 26 cardiac intensive care unit nurses.	Manikin outcomes (best practice checklists) and patient outcomes (records review); adapted National League for Nursing Satisfaction and Self-Confidence with Learning Tool.	Participants had improved adherence to critical processes of care and reduced errors in management of patients with an Impella® left ventricular assist device in simulated as well as actual patient events.
Verkuyt et al 2017 Canada	Experimental two-group pre-post study.	Students' pediatric knowledge, self-efficacy, and satisfaction.	n= 47 second year nursing students.	Pediatric Nursing Care Knowledge Test, a Pediatric Skills Self-Efficacy (SE) Survey, and Simulation Satisfaction Survey.	Both groups made modest knowledge gains; significant gains were seen in self-efficacy scores with the gaming group making greater gains. Satisfaction survey scores were high. Virtual gaming simulation

	Interactive virtual gaming simulation and a hands on manikin-based simulation.				combined with hands-on simulation is a valuable teaching resource.
Walters 2017 Indiana USA	Quasi-experimental post-test study. Standardized patients (or students acting as SP) and low fidelity manikin).	Students' knowledge and perceptions	n = 199 third-year baccalaureate nursing students enrolled in an adult health alterations course	Six-item Learning Climate Questionnaire; 21 items from the Basic Psychological Needs at Work Scale; adapted Situational Motivation Scale; student examination records.	Students enrolled in semesters that included simulations performed better on quizzes and the first examination and perceived the learning environment to be more student centred, felt more autonomous, competent and connected to the class, and were more motivated.