

Criteria to evaluate graduate nurse proficiencies in obtaining a health history and perform physical assessment in simulation-based education: A narrative review

Luis Alexandre Costa^{a,*}, Eloise Jane Monger^b

^a Department of Social Sciences and Nursing, Solent University, Southampton, United Kingdom

^b School of Health Sciences, University of Southampton, Southampton, United Kingdom

ARTICLE INFO

Keywords:

Clinical history
Competency
Evaluation
Graduate nurse
Physical assessment
Proficiencies
Simulation

ABSTRACT

Background: Simulation is a technique being used increasingly in healthcare education which offers opportunities to evaluate nursing proficiencies. The use of valid and reliable instruments is recognised as the foundation for a robust assessment, however competency-based health assessment courses for graduate nurses can consequently become reductionist in measuring proficiencies.

Objective: The specific review question was: In simulation-based education, what are the criteria that evaluate graduate nursing student's competence in obtaining a health history and performance of patient assessment?
Methods: Eleven studies were included in the review. Papers were critically appraised with The Joanna Briggs Institute quasi-experimental studies checklist. Bloom's taxonomy was used to structure this narrative review.

Results: Seven papers evaluated cognition through questionnaires and two papers used a Likert-scale to determine self-perceived knowledge. Six papers evaluated psychomotor skills with a behavioural checklist. Diversity of application was factored into the studies when testing affective skills. Three papers used Likert-scales to evaluate preparedness, six papers used Likert-scales to evaluate self-confidence and one used a Likert-scale to evaluate autonomy. Three papers used a checklist to evaluate professionalism. Four papers used faculty member/standardised patient feedback.

Conclusion: Reductionist evaluation instruments create a barrier when evaluating competency. The limited validity and reliability of assessment instruments in simulation, as well as the lack of standardisation of affective skills assessment, presents a challenge in simulation research. Affective skills encompass attitudes, behaviours and communication abilities, which pose a significant challenge for standardised assessments due to their subjective nature.

This review of the simulation literature highlights a lack of robustness in the evaluation of the affective domain. This paper proposes that simulation assessment instruments should include the standardisation of affective domain proficiencies such as: adaptation to patients' cognitive function, ability to interpret and synthesise relevant information, ability to demonstrate clinical judgement, readiness to act, recognition of professional limitations and faculty/standardised-simulated patient feedback. The incorporation of the affective domain in standardised assessment instruments is important to ensure comprehensive assessment of simulation particularly in the development of health history and physical assessment proficiencies. Attention to all of the domains in Blooms taxonomy during simulation assessment has the potential to better prepare professionals for the patient care setting.

1. Introduction

Post-registration health assessment courses are focused on establishing firm foundations in the art of conducting a thorough health exam (Mahoney, 2002). These courses allow registered nurses to advance

their skill set by incorporating educational activities that articulate clinical judgment with critical thinking (Rushforth, 2008).

Simulation is an established technique in healthcare education (McGaghie et al., 2014). Donaldson (2009) proposed the use of simulation as one of the top priorities in educating healthcare professionals in

* Correspondence to: Solent University, E Park Terrace, Southampton SO14 0YN, UK.

<https://doi.org/10.1016/j.nepr.2024.103984>

Received 5 February 2024; Received in revised form 9 April 2024; Accepted 20 April 2024

Available online 26 April 2024

1471-5953/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

the UK. Simulation-based education (SBE) creates opportunities for experiential learning and evaluation in a risk-free environment where learners can integrate theory and practice without fear of harming patients (Decker et al., 2008). Simulation integrates the development of clinical and decision-making skills, stimulating critical thinking through different modalities (Jeffries, 2005), offering opportunities to perform competency-based learning and allowing the evaluation of proficiencies (Livesay and Lawrence, 2018).

Nurse education aspires to develop competent and autonomous practitioners (Nursing & Midwifery Council, 2018). Competence is the functional adequacy and capacity to integrate knowledge and skills with attitudes and values into clinical practice context (Meretoja et al., 2004). Professional values, communication, clinical decision-making, leadership and team working are the set of core proficiencies required by the Nursing & Midwifery Council (2018). The standardisation of proficiencies allows the delivery of evidence-based care with professionalism and integrity; providing protection for members of the public and opportunities to promote health and prevent illness.

Background-based education (CBE) is an approach to teaching and learning that focuses on the mastery of proficiencies which are typically aligned with industry standards or learning outcomes and often involves flexible learning pathways, frequent assessments to assess progress and provide feedback (Collier-Sewell et al., 2023). CBE often overengineers task-orientated teaching models and reductionist assessments; where the assessment approach breaks down complex processes into simpler components for the purpose of analysis or measurement of skill needing students focus on content over process (Collier-Sewell et al., 2023). There is debate as to whether a competency-based and task-orientated model leads to accountable practice and to what degree, reductionist evaluations provide objective indicators of competence (Collier-Sewell et al., 2023). The evaluation of nursing competence has developed arbitrarily over years and differs all over the globe due to differences in practice, legislation and policies (Xu et al., 2022).

Assessment of competency in simulation is a challenging aspect of nursing education. The Association for Simulated Practice in Healthcare (ASPiH) is a UK organisation that focuses on the development and application of Simulation Based Education (SBE). ASPiH aims to mitigate challenges encountered in simulation by promoting what is considered good practice. With respects to evaluation, ASPiH guidelines ensure process adherence to specific rules. According to ASPiH (Association for Simulated Practice in Health Care, 2023), assessment in SBE can be both formative and summative. Formative assessments (FA) provide ongoing feedback to improve performance. Guidance on formative assessment states that competencies should be guided by curricular information, guidelines and strategies targeting the learner's experience. FA in SBE aim to induce self-reflection and self-evaluation, which can be emphasised during debriefing (Cant and Cooper, 2011).

SBE is increasingly used to measure competence and summatively evaluate academic achievement (Arrogante et al., 2021). Summative assessments (SA) in simulation serve as a pivotal measure of student readiness for real-world clinical settings (Kardong-Edgren, 2016). In simulation, SA should reflect clinical practice and provide a meaningful assessment of competencies (Bauer et al., 2020). Kardong-Edgren (2016) highlighted critical steps when designing summative evaluations in SBE. Recommended steps include defining the objectives, knowledge and skills to be assessed, designing appropriate simulations, selecting or developing assessment tools, ensuring validity and reliability of ratings and training the assessors. It is proposed that following these steps will ensure SA of student proficiencies is accurate.

According to ASPiH (2023) expected student performance for summative assessments should be communicated explicitly during the learning experience, based on relevant curriculum and regulatory body standards.

The assessment of competencies in SBE has predominantly relied on quantitative metrics to gauge students' level of proficiencies (Gaba, 2004). However, there is a growing recognition with regards to the value

of qualitative assessments, which delve deeper into the nuances of learners' behaviours, decision-making processes and communication skills. A literature review by Lejonqvist, (2016) noted that qualitative methods of evaluating competence included the use of video-analysis and portfolios and are associated with longer learning periods. The use of portfolios enhances learning, demonstrates competence holistically and provides evidence of professional development and reflect on learning achievements (Green et al., 2014; Lejonqvist et al., 2016).

Quantitative competence evaluation methods strive for objectivity (Oermann et al., 2023). Lejonqvist, Erikson and Meretoja (2016) refer to the dominance of tools such as checklists to standardise assessments. Quantitative evaluations in SBE can stand alone as a summative assessment or sequential formative opportunities as part of a student portfolio. Such quantitative evaluations of a cross-section of a student's competence are often restricted to performance of specific skills. Good practice dictates assessments should be evaluated with tools that have proven validity and reliability. Whilst a competency checklist can provide basis for identifying strengths and weaknesses (Jarvis and Gibson, 1997), it requires constant maintenance to incorporate new changes in practice (McKinley et al., 2008). These often omit essential competencies targeting practical underpinning knowledge and disregard affective skills (McKinley et al., 2008). Walsh et al. (2009) propose that use of quantitative evaluation tools should be partnered with reflective discussions as competence goes beyond ability to perform, encompassing knowledge, skills and attitudes that are characteristic of human beings (Nascimento et al., 2021).

Simulation educators have investigated the integration of several techniques into the process of assessing competence (Decker et al., 2008). Dillon et al. (2004) identified challenges of using SBE in competency evaluation. These included cost and time commitments, development, validity and reliability testing of scoring rubrics and whether proficiencies demonstrated in simulation are transferable into practice.

The use of valid and reliable evaluation instruments is recognised as the foundation for a robust assessment of proficiencies (Bray et al., 2011) and paucity in the application of instruments providing valid and reliable data represents a challenge in measuring outcomes. Validity is the degree to which an instrument measures what it is intended to measure (Heale and Twycross, 2015). Validating a tool requires choosing appropriate methods and enrolling enough subjects to accurately produce a statistical analysis (Polit and Beck, 2021). Heale and Twycross (2015) determined content validity as the degree to which an instrument accurately measures all aspects of a construct. Reliability is essential to ensure quality and adequacy and it refers to the consistency, stability and dependability where instruments measure the target attribute (Polit and Beck, 2021).

Comprehensive assessment of proficiencies requires instruments to provide information on confidence, ability to think critically, knowledge, skills and expertise to deliver evidence-based care (Nursing & Midwifery Council, 2018). Kardong-Edgren, (2010) reviewed published evaluation instruments in Simulation and grouped them into learning domains of the Bloom's taxonomy. The review highlights cognitive instruments as most comprehensive but that most didn't report reliability and validity features. The study concluded that to capture all learning domains in simulation, facilitators required the use of multiple instruments. The ideal instrument for a comprehensive assessment should include measures of each domain of the taxonomy and educators should consider whether the instrument is appropriate for the population and fits the activity (Adamson et al., 2013). This review explores whether the practice of simulation assessment has developed.

1.1. Theoretical basis for the study

The Kirkpatrick's Evaluations Model classifies evaluations in SBE and guided the systematic search. The learning level addresses learning through professional standards, knowledge and skill acquisition by quantitative and qualitative indicators (Johnston, Coyer and Nash,

2018). This framework was used alongside the Bloom's Taxonomy (Bloom and Krathwohl, 1956; Anderson and Krathwohl, 2001) where learning is divided into three distinct domains: cognitive, psychomotor and affective, providing a comprehensive analysis of criteria to assess knowledge, technical skills and emotional intelligence in simulation literature. Revising evaluation criteria whilst combining both models, enables gathering and analysis of data on the effectiveness of assessments, promotes quality improvement of instruments and ensures students adhere to industry standards by thoroughly evaluating competency.

2. Methods

2.1. Objective

The specific review question was: In simulation-based education, what are the criteria on evaluating graduate nursing student's competency in obtaining a health history and performing patient assessment? Due to the considerable heterogeneity in the studies with respect to methodology, participants and design; conclusions are presented as a narrative review (Popay et al., 2006).

This review included studies describing the evaluation of simulation activities, structured with prebriefing, activity and debriefing; focused on evaluation of proficiencies of graduate nursing students undertaking health history and patient assessment academic courses, that were published from 2014 onwards.

Excluded papers focused on:

1. planning as these would not directly scrutinize specific aspects of competency evaluation;
2. simulations with no debriefing, as these omit a critical component of the simulation cycle reflecting possibilities of failure to capture learning and skill retention achieved through reflective discussions;
3. where medical or undergraduate nursing students agreed to participate, due to the differences in training scope;
4. simulations that didn't take place in academic contexts, maintaining relevance to academic settings and transferability to practice.

Papers focused solely on the efficacy of simulation instruments were initially considered as part of inclusion criteria, though none were relevant to the research question identified through the systematic literature search.

2.2. Research methods

To ascertain an answer to the research question, the word-search "simulation-based education" originated 13 Cochrane reviews that were not related to this study's theme.

A systematic literature search was conducted to identify primary research in the following databases: MEDLINE, CINAHL, ERIC and British Education Index. Keywords included: graduate nurs* student, health assessment, health history, patient assessment, simulat*, evaluation, competency, proficienc*. Studies were confined to the English language.

2.3. Data extraction and synthesis

A total of three-hundred studies were identified through database searching. Another study was identified through secondary searching, making a total of three-hundred-and-one articles recognised for screening. Of the three-hundred-and-one articles, thirteen were dissertations which were excluded due to insufficient quality evidence (Evans, 2002) and poor relevance to the topic. Of the remaining two-hundred-and-eighty-eighty, five records were duplicates making a total of two-hundred-and-eighty-three articles eligible for screening. Two-hundred-and-sixteen records were excluded based on abstract information. Sixty-seven records were assessed and fifty-six were removed based on this project's exclusion criteria. A total of eleven records were

assessed for methodological validity with the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Quasi-Experimental Studies (The Joanna Briggs Institute, 2017).

Extracted data includes study methodology, simulation modality, sample size, measurement tools referencing their reported validity and reliability, evaluation criteria and outcomes. Detail on the depth of assessment is presented in the format of a thematic analysis inspired by the 6-step guide of Braun and Clarke, 2006, with grounding on the theoretical frameworks that structured the project.

3. Results

3.1. Methodology and ethics

All eleven studies were published between 2014 and 2022 and were conducted in the USA with exception of Guerrero, (2022) in the UAE. All studies presented quasi-experimental methodology in 4 overarching designs; a one-group pre and post-test design (Haut et al., 2014; Jackson et al., 2022; Woroch and McNamara, 2021; Chang et al., 2019; Ndiwane et al., 2017; VanGraafeiland et al., 2022; Brown et al., 2020); a one-group post-test design (Carman et al., 2017); a two-group pre and post-test design (Ali et al., 2020; Bryant, Miller and Henderson, 2015); and a two-group with post-test and crossover design (Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022).

All studies provided a detailed evaluation protocol including integrity, transparency and professionalism assuring standards of SBE (Phrampus, 2018). Five studies were reviewed by University Boards and acquired exemption status from educational quality improvement (Haut et al., 2014; Woroch and McNamara, 2021; Ndiwane et al., 2017; VanGraafeiland et al., 2022; Ali et al., 2020). Students agreed to participate in all studies, with explicit informed consent obtained in four of the studies (Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022; Bryant, Miller and Henderson, 2015; Jackson et al., 2022; Ali et al., 2020). Jackson et al. (2022) highlighted the use of 3 sets of standards which included best practice for Male Urogenital Teaching Associates (MUTA) who were required to sign a consent form. One study obtained ethical approval and assured that all methods were performed in accordance with the Declaration of Helsinki (Guerrero et al., 2022). One study assumed the completion of test as indicative of consent (Chang et al., 2019), which can represent a threat to the validity of the informed consent.

3.2. Timescales, modalities and sampling

Six studies involved a single event (Haut et al., 2014; Jackson et al., 2022; Woroch and McNamara, 2021; Chang et al., 2019; Ndiwane et al., 2017; Carman et al., 2017). The rest described a two-day event (Brown et al., 2020; Ali et al., 2020), a three-day (Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022), a five-day (VanGraafeiland et al., 2022) and seven-week event (Bryant, Miller and Henderson, 2015). High-fidelity Simulation (HFS) was used in four studies (Haut et al., 2014; Ali et al., 2020; Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022; Brown et al., 2020). The Standardised Patient modality is evidenced in five studies (VanGraafeiland et al., 2022; Ndiwane et al., 2017; Chang et al., 2019; Woroch and McNamara, 2021; Jackson et al., 2022). Virtual simulation was used in two papers (Carman et al., 2017; Bryant, Miller and Henderson, 2015).

Two studies compared student outcomes following two types of debriefing (Ali et al., 2020; Guerrero et al., 2022). Bryant, (2015) and Haut et al. (2014) evaluated outcomes of simulation and compared them to standard course activities. A total of nine papers detailed their sample size and data suggests a lack of diversity with such small sampling. Most participants were female, with an average of 5 years of experience.

Table I
Methods to evaluate Student competency.

Criteria to evaluate competency in obtaining a Health History and perform Physical Assessment Simulation-based Education								
Domain (Bloom's Taxonomy)	Assessment Tool							
	Cognitive		Psychomotor	Affective				
Paper	Questionnaire (Knowledge)	Likert Scale (knowledge)	Checklist (Behavioural assessment)	Likert Scale (Preparedness)	Likert Scale (Self- Confidence)	Checklist (professionalism)	Likert Scale (autonomy)	Faculty and SP feedback
Haut et al. (2014)	X		x		x	x		x
Ali et al. (2020)	X		x		x			
Guerrero, Tungpalan-Castro and Pingue-Raguini (2022)	X		x					x
Brown et al. (2020)	X		x					
Jackson et al. (2022)	X			x	x			x
Woroch and McNamara (2021)	x			x	x		x	
Chang et al. (2019)	X		x		x	x		
Ndiwane et al. (2017)		x			x			
VanGraafeiland et al. (2022)		x		x	x			
Carman et al. (2017)			x					x
Bryant, Miller and Henderson (2015)			x			x		

3.3. Evaluation of competency

Eight studies evaluated competencies in obtaining a health history, to conduct physical assessment, competence to conduct a diagnostic work-up and recognising key findings and management of condition (Carman et al., 2017; Brown et al., 2020; Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022; Ali et al., 2020; Jackson et al., 2022; Haut et al., 2014, VanGraafeiland et al., 2022; Woroch and McNamara, 2021); two studies evaluated competency in obtaining a health history (Chang et al., 2019; Ndiwane et al., 2017) and one study evaluated competency in obtaining a health history and conducting physical assessment (Bryant et al., 2015). Table I presents the instruments used to determine competency.

Seven studies used questionnaires to appraise knowledge. These include multiple-choice (Haut et al., 2014; Brown et al., 2020; Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022), fill-in-the-blank (Jackson et al., 2022), surveys (Woroch and McNamara, 2021) and questions adapted from validated instruments (Chang et al., 2019). Ali et al. (2020) didn't specify the type of assessment. Seven studies used a behavioural assessment checklist (Haut et al., 2014; Ali et al., 2020; Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022; Brown et al., 2020; Chang et al., 2019; Carman et al., 2017; Bryant, Miller and Henderson, 2015). Items on these checklists included: the ability to gather clinical history information, perform physical assessment and interpret key examination findings; formulate lists of differential diagnosis, ordering investigations to support clinical impressions and response to changes in clinical situations. Three studies used checklists to evaluate professionalism (Haut et al., 2014; Chang et al., 2019; Bryant, Miller and Henderson, 2015). Items included discussions of clinical findings, management plans and alternative treatments, addressing ethical dilemmas and escalation. Likert scales were used to evaluate preparedness (Jackson et al., 2022; Woroch and McNamara, 2021; VanGraafeiland et al., 2021), self-confidence (Haut et al., 2014; Ali et al., 2020; Jackson et al., 2022; Woroch and McNamara, 2021; VanGraafeiland et al., 2022; Chang et al., 2019), self-perceived knowledge (VanGraafeiland et al., 2022; Ndiwane et al., 2017) and autonomy (Woroch and McNamara, 2021). Four studies included faculty and SP feedback in their criteria (Jackson et al., 2022; Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022; Haut et al., 2014; Carman et al., 2017).

Four studies evidenced Cronbach Alpha as a measurement of internal consistency and reliability of instruments (Guerrero, Tungpalan-Castro

and Pingue-Raguini, 2022; Chang et al., 2019; Ndiwane et al., 2017; Bryant, Miller and Henderson, 2015). Cronbach's Alpha assesses consistency between items within a measurement (Heo, Kim and Faith, 2015). Amirrudin, Nasution and Supahar (2021) found consistency in findings when Cronbach alpha applied standard deviations as a measurement, and this was observed in Bryant, Miller and Henderson (2015). Four papers cited content validity of their instruments (Haut et al., 2014; Ali et al., 2020; Chang et al., 2019; Brown et al., 2020). Four studies did not address the validity of their instruments (Jackson et al., 2022; Woroch and McNamara, 2021; VanGraafeiland et al., 2022; Carman et al., 2017).

The following methods for quantitative data analysis were observed. Three studies used software for statistical analysis (SPSS) (Ali et al., 2020; Ndiwane et al., 2017; Woroch and McNamara, 2021). Five studies used Wilcoxon signed rank tests for their non-parametric data (Haut et al., 2014; Brown et al., 2020; Jackson et al., 2022; Woroch and McNamara, 2021; Bryant, Miller and Henderson, 2015). Two papers used Wilcoxon rank-sum tests (Ali et al., 2020; Bryant, Miller and Henderson, 2015). These tests are used to compare two related samples or to conduct a paired difference test of repeated measurements on a single sample to assess how their population mean ranks differ (Xia, 2020). These tests are a non-parametric alternative to the unpaired/paired t-tests. Two papers used paired t-tests (Brown et al., 2020; Jackson et al., 2022) and one paper independent t-tests (Ali et al., 2020). Ndiwane et al. (2017) used analysis of variance (ANOVA) for their follow-up t-tests.

Two studies described a content analysis (Jackson et al., 2022; Ndiwane et al., 2017) and qualitative data is presented in four papers (Chang et al., 2019; Jackson et al., 2022; Haut et al., 2014; Carman et al., 2017).

4. Results

4.1. Cognition

Two studies examined the effect of HFS on student knowledge. Haut et al. (2014) noticed increase in knowledge with a 17% mean, but the result wasn't statistically significant ($p=0.09$). Brown et al. (2020) reported an increase in student knowledge (mean pre-test 35.2%, SD 12.1%; mean post-test 62.2%, SD 13.8%) with statistically significant results ($p=0.001$). A total of five studies evaluated the impact of SP

simulation on student knowledge. Jackson et al. (2022) documented a non-significant knowledge increase following the activity (mean pre-test 80.7%, SD 12.4%; mean post-test 82.2%, SD 12.1%). Woroch and McNamara (2021) also reported a non-significant increase in knowledge (mean pre-test 3.41%, mean post-test 3.78%). Chang et al. (2019) reported their results in the format of pre-test, knowledge post-teaching and knowledge after simulation with satisfactory results (mean pre-test 61.48%, SD 10.6%; mean post-teaching 70.51% SD 10.18% and mean post-simulation 72.42%, SD 9.99%). The authors highlighted non-significant increase in mean results from the post-test to after-simulation. VanGraafeilan et al. (2021) didn't have sufficient data to produce mean post-test scores for their second cohort but noticed a non-significant knowledge increase for the first (mean pre-test 1.2%, SD 0.62%; mean post-test 1.7%, SD 0.48%). Ndiwane et al. (2017) reported a pre-test mean of 2.44%, SD 0.75% and a post-test mean 2.97%, SD 0.59%. Two studies compared knowledge levels following debriefing with two different models. Ali et al. (2020) noted knowledge increase with video-assisted debriefing compared with verbal debriefing (mean post-test video-assisted debriefing (VAD)= 61%, SD 1.38%; mean post-test verbal debriefing= 8.9%, SD 0.1%). Guerrero, Tungpalan-Castro and Pingue-Raguini (2022) reported better knowledge results with the multiphase debriefing structure compared with the GAS model.

Qualitative feedback from Haut et al. (2014) reported variations in clinical decisions among participants. Chang et al. (2019) demonstrated that students value skill acquisition through SBE and Ndiwane et al. (2017) reported simulation increased knowledge on aspects of the clinical history. Carman et al. (2017) noted that students had difficulties diagnosing certain scenarios like GI bleed and distinguishing hyperosmolar hyperglycaemic syndrome from diabetic ketoacidosis.

4.2. Psychomotor

Results on the behavioural checklist by Haut et al. (2014) were reported in three subgroups, with Group One completing 73% of the behavioural tasks, Group Two 68% and Group Three 60%. Jackson et al. (2022) presented the results in paired t-tests for 25 students with $t(25)=2.69$; $p=0.01$ for obtaining the history, $t(25)=5.62$; $p=0.00$ for performing the exam and $t(25)=1.04$; $p=0.001$ for formulating a diagnosis. Ali et al. (2020) used paired t-tests to report $t(26)=1.91$; $p=0.006$. Chang et al. (2019) reported a post-intervention mean 3.0065%, SD 0.357%. Carman et al. (2017) reported that 75.9% of students displayed key behaviours for their hyperosmolar hyperglycaemic syndrome scenario, 86% for their GI bleeding scenario and 87% for their febrile neutropenic breast cancer scenario. Guerrero, Tungpalan-Castro and Pingue-Raguini (2022) documented better performance results with the multiphase debriefing structure compared with the GAS model.

4.3. Affective

Ali et al. (2020) reported increased self-efficacy in students that received verbal debriefing (mean 44.6%, SD 6.2%) compared with video-assisted debriefing (mean 41%, SD 5.6%). Woroch and McNamara (2021) observed an insignificant increase in preparedness post-activity (mean pre-test 3.41%, mean post-test 3.57%). VanGraafeilan et al. (2021) reported statistics for clinical escalation and confidence which were insignificant. Brown et al. (2020) reported the time-to-task improved from 93 to 64 seconds. Qualitative data from Haut et al. (2014) and Jackson et al. (2022) implies that students don't always feel confident making decisions. Ndiwane et al. (2017) reported student discomfort when asking about race or sexuality and Carman et al. (2017) addressed that students were focused on clinical management of conditions as opposed to addressing the emotional needs of patients.

5. Discussion

All evaluation instruments contain elements identified in domains of the Bloom's taxonomy. Haut et al. (2014) evaluated cognitive, psychomotor and affective domains with a behavioural checklist which was focused on the management and identification of events in rapidly changing situations. Cognition was tested with a questionnaire and proficiencies consolidated through debriefing. Their evaluation instrument incorporated knowledge aspects regarding the management of conditions (factual knowledge), therapeutic principles (conceptual knowledge) and problem solving with algorithm and sequential thinking for decision-making, accompanied with self-reflective discussions in debriefing (metacognitive knowledge) (Su and Osisek, 2011). Eight studies attempted to assess competency with a similar format (Bryant, Miller and Henderson, 2015; VanGraafeiland et al., 2022; Carman et al., 2017), although five studies used questionnaires and not checklists to test knowledge (Jackson et al., 2022; Ali et al., 2020; Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022; Ali et al., 2020; Brown et al., 2020). The former five studies and Haut et al. (2014) may have looked at knowledge with greater depth of detail with their questionnaire, although the data is too vague to establish whether conceptual and metacognitive knowledge were tested. These assessment tools strive to provide a comprehensive view of an individual's skills and abilities. However, the transferability of assessed skills into real-world scenarios remains limited due to limitations in simulation data with regards to accurately replicating all possible situations in standardised or high-fidelity scenarios. As Rudolph et al. (2006) suggest, the variations in fidelity/realism of simulated scenarios cannot guarantee a full replication of patient care settings, therefore constraining the transferability of assessed skills. Additionally, subjectivity in assessing behaviours by different evaluators may pose reliability limitations for these instruments.

Two studies (Ndiwane et al., 2017; Woroch and McNamara, 2021) were focused in obtaining a health history with evaluation instruments centred on factual knowledge. These instruments may lead to an incomplete understanding of how to obtain a health history, as these may not provide the required context to deduce the significance of information, such as subjective experiences and patient perspective. Furthermore, instruments focused on factual data may incite bias and lead to misinterpretation or overlook important aspects of the health history (DELETE).

In the study by Haut et al. (2014), students were expected to distinguish relevant information in the health history (analyse), make judgments on the importance of their examination findings (evaluate) and effectively manage a deterioration event (create). Six studies focused their evaluation instruments on student ability to synthesise data from health history and their competence in making judgments on the history and examination findings (Jackson et al., 2022; Ali et al., 2020; Chang et al., 2019; Brown et al., 2020; Bryant, Miller and Henderson, 2015; VanGraafeiland et al., 2022, Carman et al., 2017; Ndiwane et al., 2017; Woroch and McNamara, 2021). These studies correlate with Haut et al. (2014) on the importance of integrating and evaluating analytical skills but lack learning outcomes in the context of clinical decision-making. Furthermore, their data doesn't explore common challenges as well as variation in approaches to skill evaluation that would obtain valuable insight to refine evaluation instruments.

Four studies included the readiness to act on the simulation scenario in their evaluation criteria, mitigating limitations and the ability to use sensory cues to guide motor activity. Behavioural checklists were used for this, although faculty/SP feedback was used to complement assessments (Jackson et al., 2022; Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022; Woroch and McNamara, 2021; Haut et al., 2014). Evaluating readiness to act, recognition of limitations and sensory-guided motor activity is valuable, although data is insufficient to conclude whether the instrument factors the contextual challenges associated with realism in simulation that would determine whether

Table II
Proposed Evaluation Framework to standardise the assessment of competencies in obtaining a health history and perform patient assessment in Simulation-Based Education.

Proposed Evaluation Framework to standardise assessment of competencies in obtaining a Health History and perform Physical assessment in Simulation-based Education		Competency Domain (Bloom's Taxonomy)	
Criteria for Assessment	Competency Questionnaire (Knowledge)	Direct Questions to determine level of metacognitive knowledge.	Checklist (Assessment of psychomotor skills)
			Psychomotor Adaptation to patient's cognitive ability (communication)
		Ability to Synthesise information.	Ability to demonstrate clinical judgment
		Affective Readiness to act	Recognises Limitations to professional practice
			Faculty and SP feedback

readiness to act has real-world applicability. In addition, the subjectivity of feedback incorporated in the assessment can represent variability in perspectives from different evaluators which may compromise the integrity of the assessment.

More explicitly, the affective domain was evaluated at different levels. Jackson et al. (2022) assessed student capability to organise values into priorities through SP feedback. Ndiwane et al. (2017), Woroch and McNamra (2021) and VanGraafeilan et al. (2021) looked at the value students attributed to findings (valuing). Bryant, Miller and Henderson (2015) and Carman et al. (2017) assessed valuing and student ability to organise values into priorities through checklists. Haut et al. (2014), Ali et al. (2020), Guerrero, Tungpalan-Castro and Pingue-Raguini (2022); Chang et al. (2019) and Brown et al. (2020) evaluated student ability to organise and internalise values, as their behavioural checklists looked at how values control behaviour. The discrepancy in elements that assess the affective domain is a finding in this review. This is perhaps due to the limitations associated with intentional allowance for subjectivity when scoring student interpretation of scenarios. Monger (2014) highlighted four themes that provided greater insight into affective competency elements through simulation-research. These were the ability to communicate and adapt to the patient's cognitive ability, the ability to interpret environmental clues and perform accordingly, proactiveness and choreography of practice. Proactiveness was assessed by Brown et al. (2020) analysing time-to-task performance. Communication and performance based on interpretation of environmental clues were observed in behavioural checklists in seven studies (Haut et al., 2014; Ali et al., 2020; Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022; Brown et al., 2020; Chang et al., 2019; 2015), but all present a degree of subjectiveness. SP feedback was used to benchmark the degree of student preparedness (Jackson et al., 2022), the ability to interpret environmental clues and act accordingly (Haut et al., 2014) as well as the ability to communicate (Woroch and McNamara, 2021; Ndiwane et al., 2017).

The use of behavioural checklists presents a reductionist approach in the assessment of proficiencies. Five studies used a scaled behavioural checklist (Haut et al., 2014; Ali et al., 2020; Chang et al., 2019; Brown et al., 2020; Carman et al., 2017) which contain constructs that may not be directly observable and reflect higher levels of judgment in performance (Szmulowicz et al., 2010). These checklists may offer a more comprehensive assessment of competency, but the content not directly observable can lead to subjectivity and biased marking, despite coding interventions with performance expectations (Rosen and Pronovost, 2014). Three studies (Bryant, Miller and Henderson, 2015; Woroch and McNamara, 2021; Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022) used a performed or not performed checklist, which may not reflect higher levels of judgment behind performance. Bryant, Miller and Henderson (2015) calculated the reliability of their checklist with Cronbach alpha which lends strength to their conclusions. Reductionist assessments could be subject to evaluator's interpretation, leading to subject judgments and often focus on observable behaviours, neglecting cognitive processes behind actions, prioritising quantifiable metrics over qualitative aspects of competency.

Likert-scales were used to determine attitudes and for self-assessment. These use latent variables to describe items in the scale and attempt to quantify subjective data. Most studies used worded Likert-scales. Willits et al. (2016) justify that these provide interval data. The use of means to obtain standard deviations converts descriptive statistics into usable data and research transparency is essential to distinguish interval from ordinal data, as misuse of ordinal data as interval data skews the presentation of results.

This review found no consensus in the ideal evaluative strategy to assess competency in obtaining a health history and perform patient assessment among graduate nursing students. The limitations described demonstrate how instruments in simulation may not directly translate to student performance in the real-world. The realism, different modalities and the inability to address all variables in simulation-scenarios support

Table III
Summary of the included studies.

Study	Design	Modality	Purpose	Participants	Outcomes		
					Evaluation Criteria & Instruments	Validity and Reliability	Results
Haut et al. (2014) USA	Quasi-experimental one group pre and posttest design	High-fidelity Simulation (HFS)	Evaluate the feasibility of HFS as teaching pedagogy within ACPNP programme and evaluate student competency - comparing knowledge acquired to the traditional classroom.	N=10 Group size: 3-4	Evaluation: on how to obtain a health history, conduct physical assessment, ability to conduct diagnostic work-ups, recognising key findings and managing conditions. Instruments: SMARTER (behavioural checklist), MCQ (knowledge Questionnaire), Likert scales (professionalism & self-confidence), faculty feedback.	SMARTER: content validity by 3 faculty members. Interrater reliability = 78% agreement on 62 targeted responses.	Quantitative Knowledge (N=10) - Wilcoxon signed rank test Improved from mean pretest = 61% to mean Posttest = 78% p=0.09, No SD SMARTER – students performed majority of responses – group 1 (73%), group 2 (68%) and group 3 (60%). Self-Confidence: students strongly agreed experience increased their confidence. Qualitative Feedback stated students were not always confident making clinical decisions. There were variations between clinical decision-making among participants.
Jackson et al. (2022) USA	Quasi-experimental one group pre and posttest design	Standardised Patient (SP)	This study combined simulation and male urogenital teaching associate (MUTA) feedback to evaluate preparedness, knowledge, perceived learning and confidence in diagnosis/managing urogenital complaints.	N=26	Evaluation: on how to obtain a health history, conduct physical assessment, ability to conduct diagnostic work-ups, recognising key findings and managing conditions. Instruments: Fill-in-the-blank (Knowledge Questionnaire), Likert scales (preparedness, self-confidence), MUTA feedback.	All instruments were researcher designed.	Quantitative Paired t-tests (post-simulation): History t (25) mean=2.69% p=0.01; exam t(25) mean= 5.62 p=0.00; diagnosis t(25) mean=1.04 p=0.001 Qualitative MUTA feedback (confidence and preparedness): generalised as insufficient confidence levels and large variation in degrees of preparedness
Ali et al. (2020) USA	Quasi-experimental two group pre and posttest design	High-fidelity Simulation (HFS)	Test student knowledge acquisition and retention, competency and leadership skills in life threatening dysrhythmias following HFS with 2 types of debriefing: video assisted debriefing (experiment) and verbal debriefing (control).	N= 28 N (VAD) = 16 N (VD) = 12	Evaluation: on how to obtain a health history, conduct physical assessment, ability to conduct diagnostic work-ups, recognising key findings and managing conditions. Instruments: Knowledge Questionnaire, AHA/ ACLS (behavioural checklist), ACLS, likert scale (self-confidence).	Knowledge test: content validity established by 2 investigators. Test/re-test reliability = 0.94.	Quantitative: Statistics performed with SPSS software. Independent t-tests: Knowledge Group 1: mean posttest= 61%; SD 1.38%; Group 2 mean Posttest = 8,9%; SD 0.1% (p>0.05) Competency T(26)=1.91 p=0.006 Code team leader T(26)=0.45 p=0.66 Self-efficacy: Group 1: mean= 41%; SD 5.6% Group 2: mean = 44.6%; SD 6.2%
Guerrero, Tungpalan-Castro and Pingue-Raguini (2022) UAE	Quasi-experimental two group with posttest and crossover design	High-fidelity Simulation (HFS)	The aim of the study was to compare knowledge and skills acquired by the groups of postgraduate critical care nursing students following HFS with 2 types of debriefing.	N= 30 N (EAU) = 17 N (non-EAU)=13	Evaluation: on how to obtain a health history, conduct physical assessment, ability to conduct diagnostic work-ups, recognising key findings and managing conditions.	Cronbach's alpha Sim1: 0.86, Sim2: 0.89, Sim3: 0.88, Sim4: 0.82, Sim5: 0.87, Sim6: 0.91.	Quantitative: Knowledge test Phase I Group A Mean posttest= 79.83%; SD 1.37% Group B Mean Posttest = 90.83;

(continued on next page)

Table III (continued)

Study	Design	Modality	Purpose	Participants	Outcomes		
					Evaluation Criteria & Instruments	Validity and Reliability	Results
			Researchers were looking to find which method of debriefing was more effective in the acquisition of knowledge.		Instruments: MCQ (knowledge Questionnaire), 4-point Likert scale type checklist (behavioural checklist), faculty and feedback.		SD 01.66% Mean difference 11.0% p=0.001 Phase II Group A Mean posttest= 90.75%; SD 0.75% Group B Mean Posttest = 81.1%; SD 0.76% Mean difference -9.58% p=0.000 Behavioural Checklist Phase I Group A Mean posttest= 87.39%; SD 2.03% Group B Mean Posttest = 96.91; SD 1.36% Mean difference 9.51% p=0.002 Phase II Group A Mean posttest= 88.53%; SD 1.60% Group B No data Mean difference -9.17% p=0.002
Woroch and McNamara (2021) USA	Quasi-experimental one group pre and posttest design	Standardised Patient (SP)	Evaluate whether simulation with SP would perceive confidence and competence in the assessment and management of disclosure of intimate partner violence.	N= 50 N (MHNP) = 12 N (FNP) = 38 Group Size= 4	Evaluation: on how to obtain a health history, conduct physical assessment, ability to conduct diagnostic work-ups, recognising key findings and managing conditions. Instruments: PREMIS adapted version (knowledge survey), Likert scales (autonomy, preparedness, self-confidence).	Adapted version not validated.	Quantitative: Statistics performed with SPSS software. Independent t-tests: Knowledge: T(37) mean pretest = 2.24%; mean Posttest = 3.46% Preparedness T (37) = mean pretest = 3.41%, mean Posttest = 3.78% Awareness of resources T(37)=mean pretest = 2.46% mean Posttest = 3.57%
Chang et al. (2019) USA	Quasi-experimental one group pre and posttest design	Standardised Patient (SP)	Examine the effect of motivational interviewing (MI) education in students who received MI training and SP simulation to address prescription of opioid abuse in the older adults living with chronic pain.	N=31	Evaluation: On how to obtain a health history. Instruments: MIKAT (knowledge and professionalism), BECCI (behavioural checklist), Likert scale (Self-confidence).	Cronbach's alpha MIKAT 0.84 Likert scale: reviewed by panel of experts for content validity. BECCI: low internal consistency but high inter-rater reliability with a moderate level of reliability across time.	Quantitative: MIKAT Mean pretest = 61.48% SD 10.60% Mean Posttest = 70.51% SD 10.18% Mean posttest after simulation= 72.42% SD 9.99% BECCI Mean after simulation= 3.0065% SD 0.357% Qualitative: Students found value in simulation education and felt they have acquired new skills.
Brown et al. (2020) USA	Quasi-experimental one group pre and posttest design	High-fidelity Simulation (HFS)	Determine student knowledge regarding early recognition and treatment of acute deteriorations.	N=25	Evaluation: on how to obtain a health history, conduct physical assessment, ability to conduct diagnostic work-ups, recognising key findings and managing conditions. Instruments: MCQ (knowledge	Knowledge test content validity by team of paediatric intensivists and ANP.	Quantitative Paired t-tests. Wilcoxon-signed rank test Knowledge T(23): mean pretest = 35.2% SD 12.1%, mean Posttest = 62.2% SD 13.8% (p=0.001) Time to task: Improved

(continued on next page)

Table III (continued)

Study	Design	Modality	Purpose	Participants	Outcomes		
					Evaluation Criteria & Instruments	Validity and Reliability	Results
Bryant, Miller and Henderson (2015) USA	Quasi-experimental two group pre and posttest design	Virtual Simulation (VS)	Explore how clinical proficiencies in health assessment skills could be enhanced using virtual simulations compared to standard course activities.	N=65 N (full-time) =43 N (part-time) =13	Questionnaire), behavioural checklist. Evaluation: on how to obtain a health history and conduct physical assessment Instruments: NLN tool (adapted) (behavioural checklist, professionalism checklist).	Cronbach's alpha NLN (adap) 0.89	from 93 seconds to 64 seconds. Quantitative No significant difference in course grades and IPP scores between intervention and control group. Mean difference between groups too small to be calculated. Power of the study < 80%
Ndiwane et al. (2017) USA	Quasi-experimental one group pre and posttest design	Standardised Patient (SP)	Determine whether SP simulation increased student knowledge on cultural issues when obtaining a health history	N=63	Evaluation: On how to obtain a health history Instruments: Likert scales (knowledge and self-confidence).	Cronbach's alpha Satisf 0.88; Conf 0.78; criticat thi 0.82	Quantitative Paired t-tests Knowledge T (51): mean Pretest=2.44% SD 0.75%; mean posttest= 2.97% SD 0.59% Satisfaction T(51): mean=3.83%; SD 0.67% Self confidence T(51): mean= 3.19% SD 0.78% Critical thinking T(51): mean= 3.59%, SD 0.76%
VanGraafeiland et al. (2022) USA	Quasi-experimental one group pre and posttest design	Standardised Patient (SP)	Evaluate confidence, perception of readiness, and knowledge following implementation of simulation mock telephone triage experience.	N=39 N (cohort 1) = 20 N (cohort 2)= 19	Evaluation: on how to obtain a health history, conduct physical assessment, ability to conduct diagnostic work-ups, recognising key findings and managing conditions. Instrument: Likert scales (knowledge, self-confidence, preparedness).	Instrument developed by 2 faculty members	Quantitative Knowledge Cohort 1 (N=20) Mean Pretest=1.2% SD 0.62% Mean posttest= 1.7% SD 0.48% Cohort 2 (N=19) Mean Pretest=1.5% SD 0.50% Mean posttest= no data Referrals/escalation Cohort 1 (N=20) Mean Pretest=1.4% SD 0.50% Mean posttest= 1.3% SD 0.53% Cohort 2 (n=19) Mean Pretest=1.4% SD 0.50% Mean posttest= 1.3% SD 0.57% Confidence decision-making Cohort 1 (N=20) Mean Pretest=1.2% SD 0.32% Mean posttest= 1.5% SD 0.51% Cohort 2 (N=19) Mean Pretest=1.1% SD 0.32% Mean posttest= 1.4% SD 0.50%
Carman et al. (2017) USA	Quasi-experimental one group posttest design	Virtual Simulation (VS)	Describe the use of virtual simulations in a distance-based ACNP programme and assess student performance.	N=23	Evaluation: on how to obtain a health history, conduct physical assessment, ability to conduct diagnostic work-ups, recognising key findings and managing conditions. Instrument:	key behaviours on checklist ensured interrater reliability.	Quantitative Demonstration of behaviours from % of students in the following scenarios: Hyperglycaemia hyperosmolar syndrome – 75.9% GI bleeding – 86% Breast cancer with

(continued on next page)

Table III (continued)

Study	Design	Modality	Purpose	Participants	Outcomes		
					Evaluation Criteria & Instruments	Validity and Reliability	Results
					Behavioural checklist, feedback		febrile neutropaenia – 87.1% Qualitative Students had difficulty diagnosing GI bleed and recognising/ differentiating hyperosmolar hyperglycaemia state from ketoacidosis. There was focus on management of conditions but there was lack of addressing emotional needs of patients (80% recognised septic shock but none demonstrated concern for the emotional needs of patients).

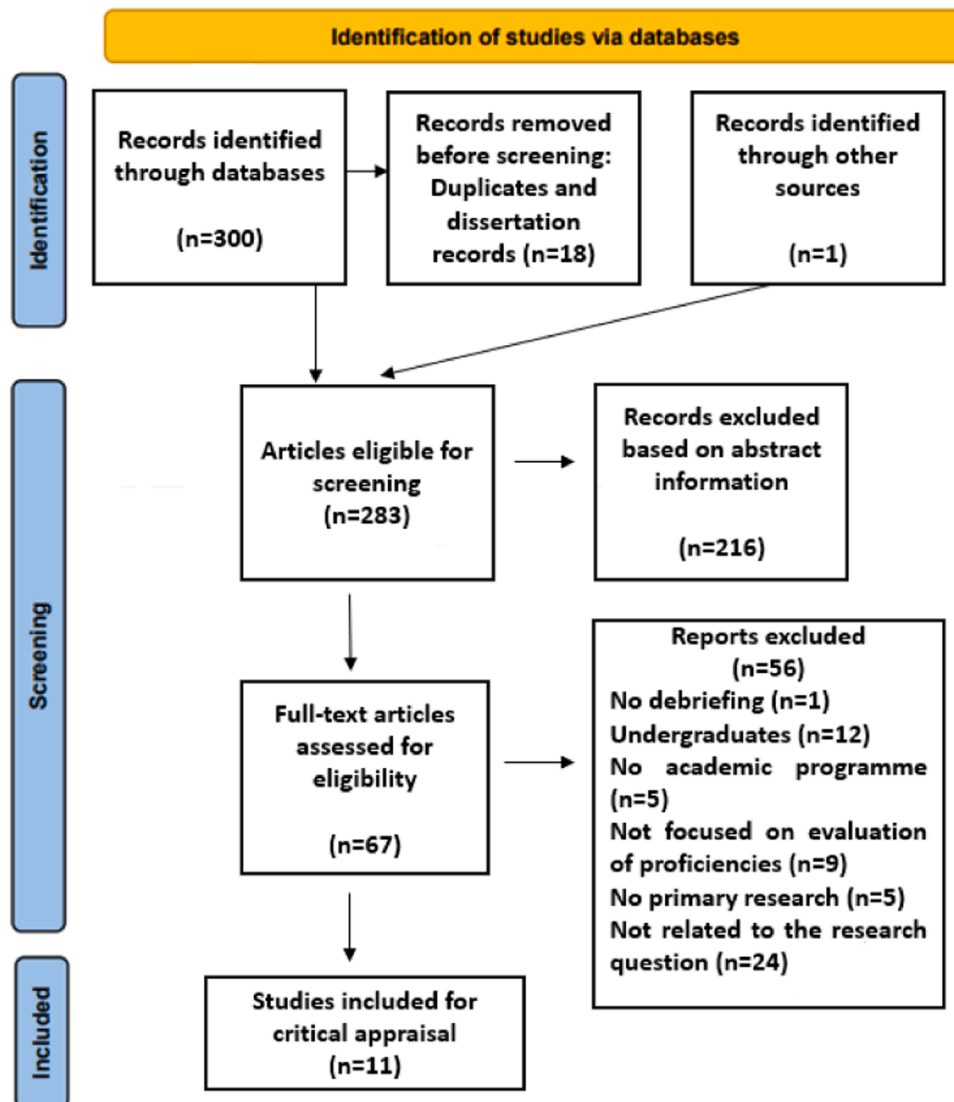


Fig. 1. PRISMA flow diagram.

the argument that instruments can't guarantee whether competency acquired in simulation is transferable to a patient care setting, or whether it will improve health outcomes. Studies included in this review, generally provided outcomes resulting from a single, two, three, five-day simulation events or a seven-week event. There were also different time allocations for prebriefing, activity and debriefing, so it is not possible to deduce the frequency and duration of SBE for greater benefit. Furthermore, the evaluation criteria in a single event will differ from a 7-week-event and in either case, it may not represent an accurate picture of competency development and acquisition. [Ndiwane et al. \(2017\)](#) found a deterioration in post-test scores in their follow-up reviews, which suggests students may benefit from repetitive SBE encounters to refine knowledge. Embedding simulation in academic programmes, incorporating formative strategies like portfolios where students document and reflect their achievements may be the key to addressing some assessment gaps identified in the literature; as experience plays a key role in the mastering of competencies ([Benner, 1984](#)).

There is paucity in validity and reliability of evaluation instruments which represents a limitation to study results. Knowledge, confidence, self-efficacy, preparedness and critical thinking measurements were described with statistical evidence, but the methods described, despite ascribed as being of good quality, may not be most appropriate to use with small samples ([Weissgerber et al., 2018](#)), which may weaken the reliability of results. Studies with item-scales included means and standard deviations to ensure equal and proportionate weight to final scores, promoting correlations in item-scales which increased internal consistency ([Liaw et al., 2015](#)). Most knowledge tests were researcher designed with no reliability testing, with the exception of [Chang et al. \(2019\)](#) who used Cronbach alpha to calculate internal consistency (0.84). Two other studies used the same method to validate their assessment instrument ([Guerrero, Tungpalan-Castro and Pingue-Raguini, 2022](#); [Bryant, Miller and Henderson, 2015](#)) which added confidence to the quality of their measurements. Three studies had their instruments informally validated by faculty members ([Haut et al., 2014](#); [Ali et al., 2020](#); [Brown et al., 2020](#)), which despite efforts may introduce bias to the research. Instruments with poor reliability may compromise validity, accuracy and credibility of the data collected, potentially having an impact on the overall integrity of the assessment. Furthermore, calculating Cronbach's alphas helps identify inconsistent items in an instrument and consistently measures intended constructs to an acceptable level ([Polit and Beck, 2021](#)).

Standardising assessments in simulation remains challenging due to the subjective nature of the affective domain, calling for transparency in research with ongoing refinement and adaptation of valid and reliable instruments that encompass the full spectrum of competencies.

5.1. Recommendations to practice

Throughout the process of narrative synthesis, the recurrent finding is the need for transparency in simulation research. Researchers should make it evident whether simulation activities are linked to key domains and whether activities are aligned with learning outcomes. The assessment criteria should be designed to focus on the above and incorporate elements of the Bloom's taxonomy for a comprehensive assessment. This review highlights questionnaires as a preferred method to assess factual and procedural knowledge, checklists to assess technique and include affective elements such as ability to adapt to patient's cognitive ability, communication efficacy, synthesis of history data, readiness-to-act, clinical judgement and recognition of professional limitations. Qualitative information such as feedback should also be considered to complement evaluations. Combining these elements would help form a comprehensive assessment of competencies in SBE. [Tables II and III](#) represents our proposed framework. Debate remains to whether instruments help translate how students will perform in clinical settings, but the use of comprehensive, valid and reliable instruments will improve simulation research [Fig. 1](#).

Instruments that allow data triangulation of knowledge and performance-checklists with qualitative feedback from SPs or faculty members, creates transparency in the assessment process, garnering insight on aspects that have not been factored in reductionist assessments, minimising limitations uncovering nuances on performance, enhancing validity and robustness of instruments. Furthermore, it is important to highlight that rubrics require ongoing evaluation and remodelling to best reflect simulation outcomes ([Cormack et al., 2018](#)) and that future research should always promote transparency in the consent process, such as implementing verbal or written consent procedures and providing clear explanations of the study purpose and procedures to participants.

5.2. Limitations

This narrative review includes 11 primary research papers which all had limitations in their research methodology and/or sample sizes. Most studies were conducted in the USA and despite being in alignment with the ASPiH standards, it is possible that due to cultural differences in healthcare systems, instruments in this review may not adequately capture the nuances and specifics of role expectations in the UK.

6. Conclusion

Bloom's taxonomy was useful to understand and categorise criteria in assessment of competency in obtaining a health history and performing patient assessment. This narrative synthesis highlights poor validity and reliability of assessment instruments in simulation, as well as the discrepancy and lack of standardisation of affective skills assessment. These skills encompass attitudes, behaviours and communication abilities, which pose a significant challenge when standardising assessment due to their subjective nature. This leads to discrepancies in drawing reliable conclusions about the trustworthiness of assessment outcomes. Poor reliability of evaluation instruments suggests inconsistent assessments and in SBE, is detrimental in capturing student skill acquisition.

Future studies that systematically address validity, reliability and standardisation of assessment instruments should incorporate simulation activities and longitudinally measure outcomes with formative strategies such as portfolios. This would ultimately better prepare healthcare professionals for the patient care setting as the mastering of competencies can only be achieved with experience.

Funding

This study was completed as a component of a Master of Science in Advanced Clinical Practice and funded by Southampton Solent University. There was no involvement of Solent University on the project.

CRedit authorship contribution statement

Luis Alexandre Costa: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Eloise Jane Monger:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of Competing Interest

The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Adamson, K.A., Kardong-Edgren, S., Willhaus, J., 2013. An updated review of published simulation evaluation instruments. *Clin. Simul. Nurs.* 9 (9), 405.

- Ali, A., et al., 2020. The impact of debriefing modalities on nurse practitioner students' knowledge and leadership skills in managing fatal dysrhythmias: a pilot study. *Nurse Educ. Pract.* 42.
- Amirrudin, M., Nasution, K., Supahar, S., 2021. Effect of variability on cronbach alpha reliability in research practice. *J. Mat. Stat. Dan. Komputasi* vol. 17 (2), 223–230.
- Anderson, L.W. and Krathwohl, D.R., 2001. *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Allyn & Bacon. Boston, MA.
- Arrogante, O., et al., 2021. Comparing formative and summative simulation-based assessment in undergraduate nursing students: nursing competency acquisition and clinical simulation satisfaction. *BMC Nurs.* 20 (1).
- Association for Simulated Practice in Healthcare, 2023. *Simulation-based Education in Healthcare Standards Framework and Guidance*. Practice in Health and Care. London: Health Education England.
- Bauer, D., Lahner, F.-M., Schmitz, F.M., Guttormsen, S., Huwendiek, S., 2020. An overview of and approach to selecting appropriate patient representations in teaching and summative assessment in medical education. *Swiss Med. Wkly.* 150, w20382.
- Benner, P., 1984. *From novice to expert: Excellence and power in clinical nursing practice*. Addison-Wesley. California: Menlo Park.
- Bloom, B.S., Krathwohl, D.R., 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners*. NY: Longmans, Green.
- Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. *Qual. Res. Psychol.* 3 (2), 77–101.
- Bray, B.S., et al., 2011. Assessment of human patient simulation-based learning. *Am. J. Pharm. Educ.* (10), 208
- Brown, K.M., et al., 2020. Pediatric critical care simulation curriculum: training nurse practitioners to lead in the management of critically ill children. *J. Pediatr. Health Care* 34 (6), 584–590.
- Bryant, R., Miller, C.L., Henderson, D., 2015. Virtual clinical simulations in an online advanced health appraisal course. *Clin. Simul. Nurs.* 11 (10), 437–444.
- Cant, R.P., Cooper, S.J., 2011. The benefits of debriefing as formative feedback in nurse education. *Aust. J. Adv. Nurs.*, 29 (1), 37–47.
- Carman, M., et al., 2017. Use of a virtual learning platform for distance-based simulation in an acute care nurse practitioner curriculum. *Dimens. Crit. Care Nurs.: DCCN* 36 (5), 284–289.
- Chang, Y.-P., et al., 2019. Motivational interviewing training with standardized patient simulation for prescription opioid abuse among older adults. *Perspect. Psychiatr. care* 55 (4), 681–689.
- Collier-Sewell, F., et al., 2023. Competencies and standards in nurse education: the irresolvable tensions. *Nurse Educ. Today* 125.
- Cormack, C.L., et al., 2018. The 360-degree evaluation model: a method for assessing competency in graduate nursing students. a pilot research study. *Nurse Educ. Today* 64, 132–137.
- Decker, S., et al., 2008. The evolution of simulation and its contribution to competency. *J. Contin. Educ. Nurs.* 39 (2), 74–80.
- Dillon G.F., et al., 2004. *Simulations in the United States Medical Licensing Examination™ (USMLE™), BMJ Quality & Safety*.
- Donaldson, L., 2009. *150 years of the Chief Medical Officer's Annual Report*. London: Department of Health.
- Evans, D., 2002. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J. Clin. Nurs.* 11 (4), 520–526.
- Gaba, D.M., 2004. The future vision of simulation in healthcare. *Simul. Healthc.* 1 (1), 2–14.
- Green, J., Wyllie, A., Jackson, D., 2014. Electronic portfolios in nursing education: a review of the literature. *Nurse Educ. Pract.* 14 (1), 4–8.
- Guerrero, J.G., Tungpalan-Castro, G.M., Pingué-Raguini, M., 2022. Impact of Simulation Debriefing Structure on Knowledge and Skill Acquisition for Postgraduate Critical Care Nursing Students: Three-Phase Vs. Multiphase. *BMC Nurs.* 21 (1).
- Haut, C., et al., 2014. Using high-fidelity simulation to teach acute care pediatric nurse practitioner students. *J. Nurse Pract.* 10 (10).
- Heale, R., Twycross, A., 2015. Validity and reliability in quantitative studies. *Evid. Based Nurs.* 18 (3), 66–67.
- Heo, M., Kim, N., Faith, M.S., 2015. Statistical power as a function of Cronbach alpha of instrument questionnaire items. *BMC Med. Res. Methodol.* 15 (1), 1–9.
- Jackson, D.C., et al., 2022. Combining simulation and male urogenital teaching associates to increase np student's clinical knowledge and confidence. *Clin. Simul. Nurs.* 72, 15–20.
- Jarvis P., Gibson S., 1997. *The Teacher Practitioner and Mentor in Nursing, Midwifery, Health Visiting and the Social Services*. Second edition. Nelson Themes, Cheltenham.
- Jeffries, P.R., 2005. A framework for designing, implementing and evaluating simulations used as teaching strategies in nursing. *Nurs. Educ. Perspect.* 26 (2), 96–103.
- Kardong-Edgren, M., 2016. Summative simulated-based assessment in nursing programs. *J. Nurs. Educ.* 55 (6), 320–324.
- Kardong-Edgren, S., Adamson, K.A., Fitzgerald, C., 2010. A review of currently published evaluation instruments for human patient simulation. *Clin. Simul. Nurs.* 6 (1), 35.
- Lejonqvist, G.B., Eriksson, K., Meretoja, R., 2016. Evidence of clinical competence by simulation, a hermeneutical observational study. *Nurse Educ. Today* 38, 88–92.
- Liaw, S.Y., et al., 2015. Designing and evaluating an interactive multimedia Web-based simulation for developing nurses' competencies in acute nursing care: randomized controlled trial. *J. Med. Internet Res.* 17 (1), 38–53.
- Livesay, K., Lawrence, K., 2018. Staff Perception of the Sustainability of a Mature Simulation Program in Nursing and Midwifery Education: a phenomenological analysis. *Nurse Educ. Today* 71, 145–150.
- Mahoney, J., 2002. Improving the advanced health assessment course. *Nurse Pract.* 27 (3), 85–86.
- McGaghie, W.C., et al., 2014. A critical review of simulation-based mastery learning with translational outcomes. *Med. Educ.* 48 (4), 375–385.
- McKinley, R.K., et al., 2008. Checklists for assessment and certification of clinical procedural skills omit essential competencies: a systematic review. *Med. Educ.* 42 (4), 338–349.
- Meretoja, R., Isoaho, H., Leino-Kilpi, H., 2004. Nurse competence scale: development and psychometric testing. *J. Adv. Nurs.* 47 (2), 124–133.
- Monger, E.J., 2014. *Video-view-point: video analysis to reveal tacit indicators of student nurse competence*. dissertation. Original typescript.
- Nascimento, J. da S.G., et al., 2021. Development of clinical competence in nursing in simulation: the perspective of bloom's taxonomy. *Rev. Bras. De. Enferm.* 74 (1).
- Ndiwane, A.N., et al., 2017. Use of Simulation to Integrate Cultural Humility into Advanced Health Assessment for Nurse Practitioner Students. *J. Nurs. Educ.* 56 (9), 567–571.
- Nursing & Midwifery Council, 2018. *Future Nurse: Standards of Proficiency for Registered Nurses*.
- Oermann, M.H., 2023. Competency-based education in nursing: getting started with program evaluation. *Nurse Educ.* 48 (2), 108–109.
- Phrampus, P.E., 2018. Simulation and integration into patient safety systems. *Simul. Healthc.* 13 (4), 225–226.
- Polit, D.F., Beck, C.T., 2021. *Nursing research: generating and assessing evidence for nursing practice*. Eleventh edn. Philadelphia, PA: Wolters Kluwer.
- Popay, J., et al., 2006. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version, Lancaster: Lancaster University [viewed 15 June 2023]. Available from: (<https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesi%20guidanceVersion1-April2006.pdf>).
- Rosen, M.A., Pronovost, P.J., 2014. Advancing the use of checklists for evaluating performance in health care. *Acad. Med.* 89 (7), 963–965.
- Rudolph, J.W., Simon, R., Dufresne, R.L., & Raemer, D.B., 2006. There's no such thing as "nonjudgmental" debriefing: a theory and method for debriefing with good judgment. *Simulation in Healthcare – Realism and Fidelity of Scenarios*.
- Rushforth, H., 2008. Reflections on a study tour to explore history taking and physical assessment education. *Nurse Educ. Pract.* 8 (1), 31–40.
- Su, W.M., Osisek, P.J., 2011. The revised bloom's taxonomy: implications for educating nurses. *J. Contin. Educ. Nurs.* 42 (7), 321–327.
- Szmulowicz, E., et al., 2010. Improving residents' end-of-life communication skills with a short retreat: a randomized controlled trial. *J. Palliat. Med.* 13, 439–452.
- The Joanna Briggs Institute Levels of Evidence and Grades of Recommendation Working Party (2017). *Checklist for Quasi-experimental Studies*.
- VanGraafeiland, B., et al., 2022. A mock telephone triage experience for nurse practitioner students. *J. Nurse Pract.* 18 (3), 320–323.
- Walsh, M., Bailey, P.H., Koren, I., 2009. Objective structured clinical evaluation of clinical competence: an integrative review. *J. Adv. Nurs.* 65 (8), 1584–1595.
- Weissgerber, T.L., et al., 2018. Meta-research: why we need to report more than 'data were analyzed by t-tests or ANOVA. *Elife* 7, 361–363.
- Willits, F.K., Theodori, G.L., Luloff, A.E., 2016. Another look at Likert scales. *J. Rural Soc. Sci.* 31 (3), 6.
- Woroch, R.A., McNamara, M., 2021. Intimate Partner Violence Standardized Patient Simulation for Nurse Practitioner Students," *Journal of the American Psychiatric Nurses Association*.
- Xia, Y., 2020. Correlation and association analyses in microbiome study integrating multiomics in health and disease. *Prog. Mol. Biol. Transl. Sci.* 171, 309–491.
- Xu, C., Koh, K.W.L. and Zhou, W., 2022. The Development of Advanced Practice Nurses in Singapore, *International nursing review*.