Methods for Measuring Healthcare Worker Competence: An Integrative Literature Review

Eric Nzirakaindi Ikoona

ikoonae@yahoo.com

National Public Health Agency, Sierra Leone

Lucy Namulemo

Uganda Counselling and Support Services

Ronald Kaluya

Uganda Counselling and Support Services

Rebecca Ikoona

MUJHU Research Collaboration

Foday Sahr

National Public Health Agency, Sierra Leone

Systematic Review

Keywords: competency assessment, healthcare workforce, clinical vignettes, OSCE, workplace-based assessment, simulation, entrustable professional activities, nurse manager competencies, psychometrics, COSMIN

Posted Date: December 16th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-8360481/v1

Additional Declarations: No competing interests reported.

Abstract

Background

The measurement of clinical competence is fundamental to ensuring patient safety, optimizing health outcomes, and strengthening healthcare systems. As healthcare delivery increasingly relies on task-shifting, competency-based approaches, and expanded roles for nursing leadership, rigorous assessment methods are essential for workforce development and quality assurance.

Objective

This integrative review synthesizes evidence on methods for measuring healthcare worker competence, examining theoretical foundations, methodological approaches, psychometric properties, and practical applications across diverse healthcare settings, including emerging evidence on nurse manager competencies.

Methods

An integrative literature review was conducted following Whittemore and Knafl's methodology, guided by PRISMA reporting guidelines. Electronic databases including PubMed, CINAHL, Scopus, Web of Science, and Embase were searched from January 2000 to January 2025. A total of 2,847 records were identified, with 65 studies (18 systematic reviews and 47 primary studies) meeting inclusion criteria after screening. Quality appraisal used AMSTAR 2 for systematic reviews and COSMIN guidelines for primary studies.

Results

The review identified multiple assessment methodologies including medical record review, direct observation, standardised patients, clinical vignettes, Objective Structured Clinical Examinations, workplace-based assessments, and simulation-based training. Contemporary evidence supports programmatic assessment approaches integrating multiple methods within competency-based frameworks such as Entrustable Professional Activities. Recent systematic reviews applying COSMIN guidelines reveal variable but improving psychometric evidence across instruments. For nurse managers specifically, twelve instruments measuring 477 competencies across four domains—strategic management, operational management, clinical competencies, and human resource management—have been identified.

Conclusions

No single method provides complete competency assessment; triangulation across multiple approaches is essential. The gap between competence and performance remains a central challenge requiring workplace-based assessment integration. Critical research gaps persist regarding validation in resource-limited settings, emerging competency domains including digital health and interprofessional collaboration, and standardization of terminology for managerial competencies.

Introduction

The measurement of healthcare worker competence represents a cornerstone of quality assurance in health systems worldwide (Al Jabri et al., 2021; Filomeno et al., 2024). Competence in healthcare encompasses a complex array of abilities, including clinical skills, diagnostic reasoning, decision-making, communication, and interprofessional collaboration—all of which are fundamental to ensuring patient safety and optimizing health outcomes (Rusli et al., 2024). As healthcare systems face increasing demands to demonstrate quality of care, the rigorous assessment of healthcare professionals' competencies has risen to prominence in strategic healthcare planning and workforce development.

The evolution of competency assessment reflects broader transformations in medical education and healthcare delivery. Traditional assessment methods, characterized by time-based progression and knowledge-focused examinations, have given way to competency-based approaches that emphasize the demonstration of observable skills and behaviours required for safe clinical practice (Dong, Low & Yan, 2024). This paradigm shift has been accompanied by the development of sophisticated assessment instruments, theoretical frameworks, and quality standards that guide the measurement of healthcare worker competence across diverse settings and disciplines. Healthcare organizations increasingly need to define roles and assess competencies in light of recent global health developments, driving the development of multiple measurement instruments across hierarchical management levels (Filomeno et al., 2024; Billiau et al., 2025).

The importance of competency measurement extends beyond educational settings to encompass critical policy decisions regarding task-shifting, scope of practice, and healthcare workforce optimization. In resource-limited settings where physician shortages necessitate expanded roles for nurses and other non-physician providers, robust competency assessment mechanisms are essential for ensuring quality care while maintaining patient safety (Fairall et al., 2012). Contemporary healthcare is challenged by aging populations, increased complexity of care demands, continuously changing regulations, and limited resources (Billiau et al., 2025). In response, organizations increasingly rely on flexible, interconnected systems of agile teams, with effective management of these interconnected ecosystems requiring well-defined and measurable competencies.

This integrative literature review aims to synthesize current evidence on methods for measuring healthcare worker competence. The specific objectives are: (1) to examine the theoretical frameworks underpinning competency assessment; (2) to evaluate the validity, reliability, and practical applicability of

different measurement methods; (3) to synthesize recent systematic review evidence on psychometric properties of competency instruments; (4) to examine available instruments for assessing nurse manager and leadership competencies; and (5) to identify critical gaps and future research directions.

Methods

Study Design

This study employed an integrative literature review methodology following the framework described by Whittemore and Knafl (2005). The integrative review approach was selected because it allows for the combination of diverse methodologies—including theoretical literature, empirical research, and systematic reviews—to provide a comprehensive understanding of a complex phenomenon. Unlike systematic reviews that typically focus on intervention effectiveness using homogeneous study designs, integrative reviews can incorporate experimental and non-experimental research to capture the full complexity of competency assessment methods. The review was guided by PRISMA guidelines for reporting (Page et al., 2021) and incorporated principles from the Arksey and O'Malley (2005) scoping review framework for mapping the breadth of literature.

Operational Definitions

To ensure conceptual clarity and facilitate consistent interpretation of findings, the following operational definitions guided this review:

Competence is defined as the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and the community being served (Epstein & Hundert, 2002). This definition encompasses both the cognitive and behavioural dimensions of professional capability and recognizes competence as developmental, impermanent, and context-dependent.

Clinical competence refers specifically to the ability to perform clinical tasks to a defined standard, integrating knowledge, skills, and professional attitudes in the delivery of patient care (Miller, 1990). It represents what a healthcare professional is capable of doing under optimal or standardized conditions.

Performance is defined as what a healthcare professional actually does in day-to-day practice (Rethans et al., 2002). Performance may differ from competence due to contextual factors including workload, resources, supervision, organizational culture, and patient characteristics. The gap between competence and performance—the 'know-do gap'—represents a central challenge in competency assessment.

Competency assessment refers to the systematic process of gathering and evaluating evidence about a healthcare professional's abilities in order to make judgments about their capability to practice safely

and effectively (Lockyer et al., 2017). This may include formative assessment (for learning) and summative assessment (for decision-making).

Entrustable Professional Activity (EPA) is a unit of professional practice that can be fully entrusted to a trainee once sufficient specific competence has been demonstrated to allow for unsupervised execution (ten Cate, 2013). EPAs describe units of work that are independently executable, observable, and measurable in their process and outcome, making them suitable for entrustment decisions.

Head nurse/Nurse manager refers to a registered nurse with 24-hour first-line organizational and hierarchical responsibility for one or two clinical units in a hospital setting, encompassing operational management, unit-level policy development, human resources management, and patient care oversight (Billiau et al., 2025). This definition distinguishes head nurses from middle-level and top-level nurse managers.

Psychometric properties refer to the measurement characteristics of an assessment instrument, including validity (the degree to which an instrument measures what it intends to measure), reliability (the consistency of measurement across time, raters, or items), and responsiveness (the ability to detect change over time), as defined by COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) guidelines (Mokkink et al., 2018).

Programmatic assessment refers to an integrated assessment system that combines multiple assessment methods, aggregates data over time, and uses competence committees or equivalent bodies to make holistic judgments about learner progress and readiness for independent practice (van der Vleuten et al., 2012).

Search Strategy

Electronic databases including PubMed, CINAHL (EBSCOhost), Scopus, Web of Science, and Embase (Ovid) were searched from January 2000 to January 2025. The search strategy combined three concept blocks using Boolean operators: (1) healthcare workers (nurse* OR physician* OR doctor* OR clinician* OR "healthcare professional*" OR "health worker*" OR "head nurse*" OR "nurse manager*"); (2) competence (competenc* OR skill* OR performance OR ability OR proficiency OR capability); and (3) assessment (assess* OR measur* OR evaluat* OR instrument* OR tool* OR scale* OR psychometric* OR valid* OR reliab*). Database-specific subject headings (MeSH, CINAHL headings) were incorporated where applicable.

Additionally, the following nursing and healthcare management journals were hand-searched: Journal of Nursing Administration, Journal of Nursing Management, Journal of Healthcare Management, International Journal of Healthcare Management, Nursing Management, Nursing Administration Quarterly, Nursing Outlook, Medical Education, Academic Medicine, and Advances in Health Sciences Education. Reference lists of included systematic reviews were screened to identify additional relevant studies. Grey literature was searched through OpenGrey and ProQuest Dissertations and Theses.

Eligibility Criteria

Inclusion criteria were defined according to the Population, Concept, and Context (PCC) framework:

Population

Healthcare professionals including physicians, nurses (all levels including nurse managers and head nurses), and allied health professionals working in clinical settings.

Concept

Methods, instruments, or frameworks for measuring clinical competence or professional competencies, including studies focusing on development, validation, or psychometric testing of assessment tools.

Context

Clinical practice settings including hospitals, primary care facilities, and community health settings; medical and nursing education programs.

Exclusion criteria included: (1) studies focusing exclusively on undergraduate student assessment without relevance to professional practice; (2) studies addressing patient self-assessment or patient-reported outcomes only; (3) conference abstracts, editorials, and commentaries without original data; (4) studies not published in English; and (5) studies with insufficient methodological detail for quality appraisal.

Study Selection and Screening

All retrieved records were exported to Rayyan systematic review software (Ouzzani et al., 2016) for duplicate removal and blinded screening. Two reviewers independently screened titles and abstracts against eligibility criteria. Full-text articles were retrieved for potentially eligible studies and independently assessed by both reviewers. Disagreements were resolved through discussion and, where necessary, consultation with a third reviewer. The study selection process is presented in the PRISMA flow diagram (Fig. 1).

Data Extraction

A standardized data extraction form was developed and pilot-tested on a sample of five studies. For systematic reviews, extracted data included: author(s), year, review type, databases searched, number of included studies, focus/population, key instruments identified, psychometric properties reported, quality assessment framework used, and main findings. For primary studies, extracted data included: author(s), year, country, study design, sample characteristics, assessment method(s) evaluated, competency domains assessed, psychometric properties (validity, reliability, responsiveness), and key findings. Data extraction was performed by one reviewer and verified by a second reviewer.

Quality Appraisal

Systematic reviews were appraised using the AMSTAR 2 (A MeaSurement Tool to Assess systematic Reviews) checklist (Shea et al., 2017). Primary studies reporting instrument development or validation were appraised using the COSMIN Risk of Bias checklist (Mokkink et al., 2018), which evaluates the methodological quality of studies on measurement properties. Consistent with integrative review methodology, no studies were excluded based on quality; however, quality ratings informed the synthesis and interpretation of findings.

Data Synthesis

Given the heterogeneity of included studies in terms of design, populations, and assessment methods, narrative synthesis was employed. The synthesis was organized thematically by: (1) theoretical frameworks; (2) assessment methods; (3) psychometric evidence; (4) nurse manager competencies; and (5) evidence from task-shifting contexts. Inductive thematic analysis was used to identify competency domains and subthemes across instruments, with preliminary themes assessed for internal homogeneity and external heterogeneity. Convergent and divergent findings across studies were identified and reported.

Results

Study Selection

The PRISMA flow diagram (Figure 1) summarizes the study selection process. The initial database search yielded 2,847 records (2,456 from electronic databases and 391 from other sources including hand-searching and reference list screening). After removing 623 duplicates, 2,224 records were screened by title and abstract, of which 2,089 were excluded as not meeting eligibility criteria. Full-text articles were assessed for 135 potentially eligible studies. Of these, 70 were excluded for the following reasons: wrong population (n=18), wrong concept/outcome (n=22), wrong study design (n=14), insufficient methodological detail (n=9), and duplicate reports (n=7). A total of 65 studies were included in the final synthesis, comprising 18 systematic reviews and 47 primary studies.

Characteristics of Included Systematic Reviews

Table 1 presents the characteristics of key systematic reviews included in this synthesis. Reviews were published between 2016 and 2025, with the majority (n=12) published from 2021 onwards, reflecting growing attention to competency assessment methodology. Reviews addressed diverse populations including general healthcare professionals, nurses, nurse managers at various hierarchical levels, and physicians. Quality appraisal using AMSTAR 2 indicated that most reviews were of moderate to high

methodological quality, with common limitations including incomplete reporting of excluded studies and lack of protocol registration.

Table 1. Characteristics of Included Systematic Reviews

Author (Year)	Review Type	Focus/Population	Databases	N	Key Findings
Billiau et al. (2025)	Scoping review	Head nurses (24-hr responsibility)	PubMed, Embase, CINAHL	12	477 competencies identified; 4 domains (strategic, operational, clinical, HR); 20 subthemes; quality/safety most consistently measured
Filomeno et al. (2024)	Systematic review (COSMIN)	Nurse managers (all levels)	PubMed, Scopus, CINAHL, PsycINFO	10	Chase instrument most comprehensive; CASHN highest quality (Grade A); variable psychometric evidence
Mainz et al. (2024)	Scoping review	Health professionals (digital competence)	ScienceDirect, Scopus, PubMed, EBSCOhost	46	76% studies post-2020; strong focus on technical skills; self-assessment predominant; variable statistical quality
Robinson et al. (2024)	Systematic review	Health workers in LMICs (simulation)	Multiple databases	38	Simulation effective for training; limited Kirkpatrick Level 4 outcomes; cost analyses sparse
Cheng et al. (2024)	Systematic review	Healthcare simulation (debriefing)	Multiple databases	70	Debriefing critical for learning; video-assisted debriefing shows benefits in some studies
Rusli et al. (2024)	Hybrid systematic review	Home-based care nurses	PubMed, CINAHL, Scopus	15	Identified competency instruments specific to home care context; highlighted contextual factors
Al Jabri et al. (2021)	Systematic review (COSMIN)	Healthcare professionals (core competencies)	CINAHL, Scopus, PubMed	9	9 instruments identified; gaps in quality improvement, safety, communication, health IT domains
Anderson et al. (2021)	Scoping review	WBA implementation	MEDLINE, CINAHL, ERIC, Scopus	82	Multiple methods needed; programmatic assessment supported; assessor training essential

Laakkonen et al. (2024)	Mixed-methods SR	Healthcare leaders (digital competence)	Multiple databases	23	Growing importance of digital leadership competencies; limited validated instruments

Abbreviations: COSMIN = COnsensus-based Standards for selection of health Measurement INstruments; WBA = Workplace-Based Assessment; LMICs = Low- and Middle-Income Countries; HR = Human Resources; SR = Systematic Review; N = Number of included studies

Summary of Assessment Methods

Table 2 provides a comparative summary of the principal methods for measuring healthcare worker competence identified in this review, organized by their position within Miller's pyramid framework.

Table 2. Comparison of Competency Assessment Methods

Method	Miller's Level	Strengths	Limitations	Resources
Medical Record Review	Does	Low cost; retrospective; unobtrusive; large samples possible	Underestimates care quality; documentation bias; cannot assess undocumented actions	Low
Direct Observation	Does	Real-time assessment; captures actual practice behaviours	Hawthorne effect; inter-rater variability; expensive; requires trained observers	High
Standardised Patients	Shows How / Does	Standardized cases; can assess rare conditions; replicable across providers	Cannot simulate all conditions; single encounter unreliable; training costs	Moderate-High
Clinical Vignettes	Knows How	Low cost; standardized; large-scale feasible; controls case-mix variation	Assesses knowledge not practice; know-do gap; may not predict performance	Low
OSCE	Shows How	Standardized; samples broadly; assesses multiple skills; enables fair comparison	Artificial environment; resource intensive; may not predict real performance	High
Workplace-Based Assessment	Does	Authentic setting; integrates with practice; formative feedback; captures actual performance	Reliability concerns; assessor training needed; sampling challenges; time-intensive	Moderate
Simulation-Based Assessment	Shows How	Safe environment; repeatable; immediate feedback; crisis management; procedural skills	Fidelity concerns; transfer to practice uncertain; equipment costs; faculty training	High
EPA-Based Assessment	Does	Work-based; entrustment decisions; integrates competencies; developmental progression	Requires cultural change; infrastructure needs; faculty development; implementation complexity	Moderate-High

Note: Miller's levels refer to the pyramid of clinical competence: Knows (knowledge), Knows How (applied knowledge), Shows How (demonstration), Does (performance in practice). OSCE = Objective Structured Clinical Examination; EPA = Entrustable Professional Activity.

Theoretical Frameworks for Competency Assessment

Donabedian's Structure-Process-Outcome Framework. Avedis Donabedian's seminal framework for evaluating healthcare quality provides a foundational lens through which competency assessment can be understood (Donabedian, 1966; Ayanian & Markel, 2016). The framework categorizes quality

measurement into three interconnected dimensions: structure, process, and outcomes. In the context of competency assessment, structure encompasses the attributes of the healthcare setting, including facilities, equipment, and human resources—particularly the qualifications and competencies of healthcare workers themselves. Process refers to the transactions between patients and providers throughout the delivery of care, including diagnostic procedures, treatment protocols, and patient education. Outcomes represent the effects of healthcare on the health status of patients and populations (De Rosis et al., 2024).

The Donabedian framework remains influential in contemporary healthcare quality assessment (Yang et al., 2025). Recent applications have integrated patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) within the framework, reflecting the growing emphasis on patient-centred care (De Rosis et al., 2024). The framework's flexibility allows for application across diverse healthcare settings and levels within delivery systems, from individual clinical encounters to population-level health outcomes. Importantly, Donabedian emphasized that improvements in structure should lead to enhanced processes, which in turn yield better outcomes—a causal chain that underscores the importance of competency assessment as a structural prerequisite for quality care.

Miller's Pyramid of Clinical Competence. George Miller's framework for assessing clinical competence, introduced in 1990, provides a hierarchical model that has profoundly influenced medical education and assessment for nearly three decades (Miller, 1990; Al-Hashimi et al., 2023). The pyramid organizes clinical competence into four ascending levels: 'knows' (knowledge), 'knows how' (applied knowledge), 'shows how' (demonstration of competence), and 'does' (performance in practice). This framework recognizes that knowledge alone is insufficient for clinical practice; healthcare professionals must be able to apply knowledge, demonstrate skills, and ultimately perform competently in real clinical settings.

Traditional assessment methods such as multiple-choice examinations target the lower levels of Miller's pyramid—'knows' and 'knows how'—focusing on cognitive knowledge and theoretical understanding (Dewan, Khalil & Gupta, 2024). In contrast, performance-based assessments such as OSCEs and workplace-based assessments target the higher levels—'shows how' and 'does'—evaluating the ability to demonstrate and perform clinical tasks. The goal of assessment is to be educational, formative in nature, and to offer a reliable measurement of capacity while predicting future clinical performance (Giesey et al., 2020). The pyramid has been fundamental in shifting medical education away from purely theoretical modes of assessment toward clinical performance-based methods that more accurately predict competence in practice.

Competency-Based Medical Education and Entrustable Professional Activities. Competency-based medical education (CBME) represents a transformative approach that emphasizes the demonstration of predefined competencies rather than time-based progression through training (ten Cate, 2013; Lockyer et al., 2017). CBME organizes the educational experience around competencies, emphasizes performance outcomes, promises greater accountability to patients and society, and is flexible and learner-centred (Warm et al., 2014). Major frameworks including the Accreditation Council for Graduate

Medical Education (ACGME) competencies in the United States and CanMEDS in Canada define core competency domains that healthcare professionals must master.

Entrustable Professional Activities (EPAs), developed by ten Cate (2005), emerged as a practical mechanism for translating competency frameworks into clinical practice (ten Cate, 2013; Mink et al., 2022). EPAs are defined as units of professional practice that can be fully entrusted to a trainee once sufficient specific competence has been demonstrated to allow for unsupervised execution. Unlike competencies, which describe attributes of physicians, EPAs describe units of work—they are independently executable, observable, and measurable in their process and outcome, making them suitable for entrustment decisions (ten Cate & Carraccio, 2019). Recent implementations demonstrate large-scale application of EPA-based assessment. Canada's Competence by Design (CBD) program represents postgraduate specialist medical education's transformative change to a competency-based model centred around EPAs (Humphrey-Murto et al., 2024). The Association of American Medical Colleges (AAMC) has identified thirteen core EPAs for entering residency, and recent research demonstrates that EPA-based assessments correlate positively with OSCE performances (Chen et al., 2024). The American Board of Surgery launched EPAs for general surgery residency programs in July 2023.

Methods for Measuring Healthcare Worker Competence

Healthcare worker competence can be measured through a variety of methods, each with distinct strengths and limitations. This section reviews the principal approaches to competency assessment, examining the evidence for validity, reliability, and practical applicability of each method.

Patient Outcome Measures. Patient outcomes—including cure rates, complications, mortality, and readmission rates—have traditionally been used as proxy measures of healthcare worker competence (Mant, 2001; Donabedian, 2005). The logic is straightforward: competent providers should produce better patient outcomes. However, patient outcome measures are frequently confounded by factors beyond the control of individual healthcare workers, including patient comorbidities, socioeconomic determinants of health, and health system characteristics (Peabody et al., 2000). Consequently, where scientific evidence establishes a clear link between the process of care and patient outcomes, process measures are often preferred over outcome measures for competency assessment (Brook & Lohr, 1985). Contemporary approaches increasingly recognize the complementary nature of structure, process, and outcome measures, with integrated assessment systems drawing on multiple data sources.

Medical Record Review. Medical record review has long been employed to measure the technical quality of care provided by healthcare workers (Luck et al., 2000). This method offers several advantages: medical records are ubiquitous, can be obtained after each patient encounter, provide a temporal relation of clinical events, and can be reviewed retrospectively at relatively low cost without alerting healthcare providers to the evaluation. Despite these advantages, medical record review has significant limitations. Records are typically generated for purposes other than documenting the process of care—such as legal

protection or obtaining reimbursement—and may lack crucial clinical details. Studies using standardised patients found that one-third to one-half of procedures performed were not recorded, with counselling activities particularly under-documented (Norman et al., 1985). Medical records systematically underestimate the quality of care (Luck et al., 2000).

Direct Observation. Direct observation and recording of patient encounters represents a commonly used approach to competency assessment, particularly in developing countries (Nolan et al., 2001). This method allows assessors to observe clinical performance in real time, potentially capturing aspects of care that may not be documented in medical records. However, direct observation faces significant methodological challenges. The method is expensive, requiring trained observers who must be present during clinical encounters. More problematically, direct observation is highly susceptible to the Hawthorne effect—healthcare workers' behaviour may change substantially when they are aware of being observed, introducing participation bias that compromises validity. A single demonstration of competence is not sufficient to ensure adequate performance in everyday practice—the so-called competence-performance gap (Rethans et al., 2002). Additionally, there is substantial variation in judgment between observers, even when standardized assessment tools are used.

Standardised Patients. Standardised patients (SPs) are individuals trained to provide a reproducible and unbiased presentation of a patient case during clinical encounters (Tamblyn et al., 1991; Nawathe & Herrington, 2024). SPs may be either real patients or healthy actors trained to portray symptoms and problems for healthcare education, evaluation, and research. The SP methodology has been widely validated for assessing history-taking, physical examination, communication skills, and clinical decision-making. SPs may be announced or unannounced ('mystery patients'). Research demonstrates that experienced physicians cannot reliably differentiate real patients from unannounced SPs, and that clinical performance is similar for both types (Peabody et al., 2000). Studies show that SPs can be trained to accurately and consistently evaluate provider performance, achieving agreement rates exceeding eighty percent with faculty physician observers. A key advantage of SPs is their ability to provide a replicable case for multiple healthcare providers, enabling direct comparison of performance. However, clinical performance by individual providers is not always consistent across different patients and presentations, meaning that a single SP encounter does not provide a reliable estimate of overall provider competence.

Clinical Vignettes. Clinical vignettes are brief, written case histories of simulated patients based on typical clinical situations, accompanied by questions exploring what a healthcare provider would do if presented with the actual patient (Veloski et al., 2005; Peabody et al., 2004). Vignettes are validated competence measurement tools that simulate actual patient conditions to evaluate the potential competence of healthcare providers to deliver required care. Their major advantage lies in controlling for case-mix variation—the same set of vignettes can be presented to multiple providers, eliminating the need for case-mix adjustment when comparing performance. Well-constructed clinical vignettes typically include sections representing the sequential phases of a clinical encounter: history taking, physical examination, laboratory tests, diagnosis, and treatment planning. Validation studies comparing

vignettes with standardised patient visits have demonstrated that vignettes reflect actual clinical practice and possess greater predictive validity than abstracted medical records (Peabody et al., 2000; Leonard & Masatu, 2005). Vignettes are particularly valuable in resource-limited settings due to their low cost and ease of administration. Nevertheless, while vignettes assess what providers know to do, they do not necessarily capture what providers actually do in practice—the 'know-do gap' (Das & Hammer, 2005).

Objective Structured Clinical Examinations. The Objective Structured Clinical Examination (OSCE) has become a global standard for assessing clinical competence in health professions education (Harden & Gleeson, 1979; Al-Hashimi et al., 2023; Dewan et al., 2024). Originally developed by Ronald Harden in 1975, the OSCE consists of a circuit of stations—typically five to fifteen minutes each—through which candidates rotate, completing standardized clinical tasks while being assessed by examiners using predetermined criteria. Stations may involve history-taking, physical examination, counselling, interpretation of diagnostic data, or procedural skills. The OSCE format offers several advantages: standardization enables fair peer comparison, the structured nature ensures systematic assessment of specific competencies, complex procedures can be evaluated without endangering patient health, and the multi-station format samples broadly across clinical domains, enhancing content validity. Recent systematic reviews have examined the psychometric properties of OSCEs across various healthcare disciplines. Vhora et al. (2024) evaluated OSCE implementation in psychology programs, finding variable but generally positive evidence for validity and reliability when standardised protocols were followed. Despite widespread adoption, OSCEs face criticisms regarding construct validity and cost. The COVID-19 pandemic accelerated adoption of virtual OSCE formats, expanding accessibility while introducing new validity considerations (Dewan et al., 2024).

Workplace-Based Assessment. Workplace-based assessment (WBA) encompasses a range of methods designed to evaluate healthcare professionals' performance in actual clinical settings (Norcini & Burch, 2007; Anderson et al., 2021). Unlike OSCEs or written examinations that assess what trainees can do under standardized conditions, WBA targets what they actually do in day-to-day practice—the apex of Miller's pyramid. Common WBA methods include the mini-clinical evaluation exercise (mini-CEX), direct observation of procedural skills (DOPS), case-based discussion (CBD), and multi-source feedback (360-degree evaluation). Dong et al. (2024) synthesize key principles for effective WBA implementation: validity (alignment with authentic practice), reliability (consistency through multiple observations), fairness (standardization of expectations), flexibility (adaptation to clinical context), and feasibility (integration with workflow). The evidence supports using WBA within programmatic assessment frameworks that aggregate multiple observations over time rather than relying on single high-stakes assessments.

Simulation-Based Assessment. Simulation-based training and assessment has emerged as a major innovation in competency development (Ahmed et al., 2024; González García et al., 2024). Simulation encompasses diverse modalities including mannequin-based simulation, standardised patients, virtual reality, screen-based simulation, and hybrid approaches. Simulation offers unique advantages: it

provides a safe environment where errors do not harm patients, enables deliberate practice of rare or high-stakes scenarios, and allows for standardized assessment across learners. Robinson et al. (2024) systematically reviewed simulation-based education of health workers in low- and middle-income countries (LMICs), screening 27,738 records. They found growing evidence of effectiveness at Kirkpatrick Level 4 (patient outcomes), though high-quality studies demonstrating sustainable patient outcome improvements remain limited. Meta-analyses demonstrate that simulation with deliberate practice yields superior skill acquisition and retention compared to traditional methods (Ahmed et al., 2024). Debriefing is recognized as a critical component of simulation-based learning; a systematic review of 70 studies found that video-assisted debriefing improved performance outcomes in some studies (Cheng et al., 2024). Artificial intelligence is increasingly reshaping simulation-based education, enabling adaptive feedback and personalized learning (Pearson, 2024).

Psychometric Standards and Instrument Quality

The quality of competency assessment depends fundamentally on the psychometric properties of assessment instruments and methods. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative has established rigorous guidelines for evaluating measurement properties including content validity, structural validity, internal consistency, reliability, measurement error, hypothesis testing for construct validity, cross-cultural validity, and responsiveness (Mokkink et al., 2018). Recent systematic reviews applying COSMIN guidelines reveal variable psychometric evidence across competency instruments.

Al Jabri et al. (2021) conducted a systematic review of healthcare professionals' core competency instruments, identifying nine validated tools addressing competencies in professionalism, ethical and legal issues, research and evidence-based practice, and personal attributes. The review noted significant gaps in psychometric evidence across instruments and populations. Few instruments addressed competencies in quality improvement, safety, communication, or health information technology. Filomeno et al. (2024) applied COSMIN guidelines to evaluate nurse manager competency instruments across first-, middle-, and top-level managers, finding that the Chase Nurse Manager Competency Instrument was most comprehensive while the CASHN questionnaire scored highest on methodological quality, achieving Grade A according to COSMIN criteria.

A scoping review of digital competence measurement among health professionals (Mainz et al., 2024) identified substantial growth in publications from 2020 onward, with seventy-six percent of included studies published after 2020, reflecting increasing attention to technology-related competencies. The review found a strong focus on technical skills and knowledge in both definitions and measurement tools, with less attention to broader digital competency dimensions including methodological, social, and personal competences. Measurement instruments mainly used self-assessment as indicators of competence and differed greatly in their statistical quality. Laakkonen et al. (2024) synthesized evidence on healthcare leaders' digital competence, demonstrating the value of mixed-methods systematic

reviews in understanding complex competency constructs. These reviews underscore the dynamic nature of competency requirements and the need for instruments that evolve with changing healthcare practices.

Instruments for Measuring Nurse Manager Competencies

A distinct body of literature has emerged addressing competency assessment for nurse managers and head nurses, reflecting the unique demands of first-line nursing leadership positions. Head nurses act as the cornerstone between multiple stakeholders, manage first-level nursing services, translate the organization's culture and strategic goals at the operational level, and hold 24-hour organizational and hierarchical accountability for patient care (Billiau et al., 2025). Their job content encompasses operational management, unit-level policy development, human resources management, and patient care, requiring them to embody multiple roles including leaders, liaisons, monitors, distributors, intrapreneurs, crisis managers, resource allocators, and negotiators.

A recent scoping review by Billiau et al. (2025) identified twelve instruments published between 2006 and 2024 that measure head nurses' competencies. These include the Chase Nurse Manager Competency Instrument (53 items), the Nurse Managers' Leadership and Management Competencies Scale (194 items), the Assessment Tool for Head Nurses' Managerial Competencies (78 items), the Nurse Manager Competency Instrument for Research (26 items), the Leadership and Management Inventory (28 items), the Nursing Leadership and Management Competency Inventory (23 items), the Nurse Manager Performance Assessment (43 items), the Nurse Manager Competency Inventory (93 items), the Competency Assessment Scale for Head Nurses (52 items), the Indonesian First-Line Nurse Managers' Managerial Competence Scale (43 items), the Nurse Manager Competency Model (22 items), and the Nurse Manager Evidence-Based Practice Competency Scale (16 items). Data originated from diverse countries including the United States (n=4), Indonesia, Finland, Iran, Sweden, Taiwan, Turkey, Thailand, and China.

Across these twelve instruments, Billiau et al. (2025) identified a total of 477 competencies. Thematic analysis revealed four key competency domains encompassing 20 related subthemes. First, strategic management competencies (addressed by 11/12 instruments) involve systemic thinking, research and evidence-based practices, change management, creative thinking, external liaison, and possessing an overall organizational view and political savviness. Second, operational management competencies (12/12 instruments) focus on quality and safety management, process management, technology and information management, financial management, resource management, and staff planning. Third, clinical competencies (8/12 instruments) encompass direct patient care, case management, customer service commitment, and diversity and ethical practices. Fourth, human resources competencies (12/12 instruments) include communication and interpersonal competencies, general staff management, professional development, and individual staff management.

Quality and safety management emerged as the most consistently measured competency domain, addressed by all twelve instruments. Other frequently assessed domains included process management, systemic thinking, individual staff management, and communication and interpersonal competencies. In contrast, competencies such as providing direct patient care, customer service commitment, external liaison, and resource management were addressed by fewer instruments. Nine instruments permitted self-assessment, two facilitated 360-degree assessment, and one enabled assessment by the head nurses' superiors. Ten instruments applied a five-point Likert scale, whereas two instruments used a four-point Likert scale.

The development of head nurses' competencies has been associated with important organizational outcomes including increased job satisfaction among staff, higher levels of patient satisfaction, more substantial professional commitment, heightened workplace engagement, better adherence to safety protocols, improved organizational productivity and profitability, and encouragement of innovative work practices (Billiau et al., 2025). However, the literature reveals significant inconsistencies in the terminology and descriptions of the head nurse position, as well as the competencies outlined in instruments. Terms such as nurse managers, nurse supervisors, and nurse leaders are often used interchangeably although they may not be equivalent, complicating the selection of appropriate instruments and cross-study comparisons.

Evidence on Quality of HIV Clinical Services Provided by Nurses

The measurement of healthcare worker competence has particular relevance in the context of task-shifting policies that expand clinical roles for non-physician providers. A substantial body of evidence has examined the quality of HIV clinical services provided by nurses, responding to concerns that antiretroviral treatment (ART) is complex and requires investigation of care quality when delivered by non-physicians (Hopkins et al., 1996; Philips et al., 2008). Studies across sub-Saharan Africa have consistently demonstrated that nurses, when adequately trained and supported, provide ART with patient outcomes—including virologic suppression, adherence rates, and retention in care—comparable to those achieved by physicians (Bedelu et al., 2007; Zachariah et al., 2007; Fairall et al., 2012).

A landmark randomized controlled trial in South Africa (Fairall et al., 2012) assessed the effects of nurse task-shifting on mortality, viral suppression, and quality indicators for people living with HIV. The study found no difference in time to death or viral load suppression rates between nurse-led and doctor-led groups, with nurses safely carrying out ART initiation and re-prescription while improving patient health outcomes and quality of care. Suzan-Monti et al. (2015) reported that patients seen by nurses in Cameroon did better than those seen by doctors on all WHO Quality of Life elements.

Critically, positive outcomes in nurse-led care have been attributed to health systems having strong supportive supervision mechanisms, continuous education, simplified standardized guidelines, and

systematic on-site clinical mentorship (Chopra et al., 2008). This evidence demonstrates that competence is not solely an individual attribute but is shaped by structural and organizational factors that enable or constrain effective practice. The literature identifies a need for research examining nurses' competencies in their natural work settings, as opposed to controlled experimental conditions, to understand the real-world translation of competence into quality care (Crowley & Mayers, 2015).

Discussion

This integrative review synthesized evidence from 65 studies including 18 systematic reviews to provide a comprehensive analysis of methods for measuring healthcare worker competence. Several key themes emerge from the synthesis of theoretical frameworks, methodological approaches, and empirical evidence that carry important implications for practice, policy, and research.

First, there is clear recognition that no single method provides a complete assessment of competence. Each approach—from medical record review to simulation-based assessment—captures different facets of clinical competence and operates at different levels of Miller's pyramid. Medical record review and patient outcomes are readily available but face validity limitations; standardised patients and OSCEs offer standardization but assess 'shows how' rather than 'does'; workplace-based assessment targets authentic performance but faces challenges of reliability and resource intensity. Contemporary best practice emphasizes programmatic approaches that triangulate multiple methods to generate comprehensive, valid evidence for competency decisions (Anderson et al., 2021; Dong et al., 2024; Humphrey-Murto et al., 2024). The integration of formative and summative assessment, with robust feedback mechanisms, supports both learning and accountability. The 12 nurse manager instruments identified demonstrate that even within a specific population, multiple approaches with varying comprehensiveness are needed.

Second, the gap between competence and performance remains a central challenge in the field. Clinical vignettes and OSCEs assess what providers know and can demonstrate, but may not fully predict what they actually do in practice. The 'know-do gap' has been documented across multiple settings and has important implications for quality improvement (Das & Hammer, 2005). Workplace-based assessment addresses this gap by evaluating performance in authentic clinical settings, but faces challenges of standardization, observer bias, and resource requirements. EPAs offer a promising framework for linking competency assessment to entrustment decisions about actual clinical responsibilities, with national-scale implementations demonstrating feasibility (Humphrey-Murto et al., 2024), though implementation requires substantial infrastructure and cultural change.

Third, context matters profoundly. The evidence base for competency assessment methods is concentrated in high-resource settings, with limited validation research from low- and middle-income countries where the need for efficient, scalable assessment approaches is greatest. Clinical vignettes have demonstrated feasibility in resource-limited settings (Leonard & Masatu, 2005), and simulation-based education shows promise for health worker training in LMICs (Robinson et al., 2024), but more

research is needed on the psychometric properties and practical implementation of various methods across diverse healthcare contexts. The impact of structural factors—supervision, training, equipment, workload—on the translation of competence into quality care requires greater attention in competency assessment frameworks, as highlighted by the task-shifting literature.

Fourth, emerging competency domains present new measurement challenges. Digital competence, interprofessional collaboration, leadership, and quality improvement skills are increasingly recognized as essential for contemporary healthcare practice, but validated assessment instruments for these domains remain limited (Mainz et al., 2024; Laakkonen et al., 2024). The rapid pace of technological change in healthcare requires adaptive assessment approaches that can evolve with changing competency requirements. Artificial intelligence may reshape both the competencies required of healthcare workers and the methods available for assessing them (González García et al., 2024).

Fifth, for nurse managers specifically, the review of available instruments reveals the multifaceted competencies required for effective leadership while highlighting significant terminological and conceptual inconsistencies. The 477 competencies identified across twelve instruments underscore the diverse and complex nature of head nurses' job content (Billiau et al., 2025). However, competencies are not sufficiently defined, and international literature lacks a clear distinction between competencies and tasks, making it difficult to assess head nurses' competencies consistently across settings and to compare findings across studies. Existing measurement instruments are often created in specific national or institutional contexts, and have rarely been validated against a broader, more generalizable framework.

Several limitations of this review should be acknowledged. The integrative review methodology, while appropriate for synthesizing diverse evidence, does not permit the quantitative pooling of effect sizes possible in systematic reviews of intervention studies. The focus on English-language publications may have excluded relevant research from non-English sources. Additionally, the rapidly evolving nature of competency assessment means that new evidence continues to emerge that may not be captured in this review.

Conclusion

The measurement of healthcare worker competence is fundamental to ensuring patient safety, optimizing health outcomes, and strengthening health systems. This integrative review has synthesized evidence on the theoretical foundations, methodological approaches, and practical applications of competency assessment in healthcare. Donabedian's structure-process-outcome framework and Miller's pyramid of clinical competence provide enduring conceptual foundations, while competency-based education frameworks and EPAs offer practical mechanisms for implementing competency assessment in clinical training and practice.

The evidence supports the use of multiple complementary methods—including clinical vignettes, OSCEs, standardised patients, and workplace-based assessment—within programmatic assessment systems

that aggregate diverse evidence to inform robust competency decisions. Each method has distinct strengths and limitations, and the selection of appropriate methods depends on the purpose of assessment, available resources, and the specific competencies being evaluated. National-scale implementations of EPA-based programmatic assessment demonstrate feasibility and provide models for other contexts. Future research should prioritize validation of assessment instruments in diverse settings, development of measures for emerging competency domains including digital health and interprofessional collaboration, standardization of terminology for managerial competencies, and investigation of the link between competency assessment and patient outcomes.

For health systems, particularly in resource-limited settings, the findings underscore the importance of investing in assessment infrastructure, training for assessors, and systems for aggregating and interpreting competency data. The evidence on nurse-led HIV care demonstrates that with appropriate support—including training, supervision, and simplified protocols—non-physician providers can deliver high-quality care, highlighting the potential of task-shifting policies when accompanied by robust competency assurance mechanisms. For nursing leadership, the identification of four key competency domains—strategic management, operational management, clinical competencies, and human resource management—provides a comprehensive framework that can support both future instrument development and practical competency assessment. As healthcare systems worldwide continue to evolve, the rigorous assessment of healthcare worker competence will remain essential for ensuring quality, safety, and equity in health service delivery.

Declarations

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

ENI conceptualized and designed the study, conducted the literature search, performed data extraction and analysis, and drafted the manuscript. LN contributed to study design, participated in screening and data extraction, and critically revised the manuscript. RK contributed to quality appraisal and data synthesis. RI contributed to data extraction and manuscript revision. FS provided supervision and

critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgement

The authors thank the staff of the National Public Health Agency, Sierra Leone, for their support during the conduct of this review. We also acknowledge the contributions of colleagues who provided feedback on earlier drafts of this manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article. The search strategies, data extraction forms, and quality appraisal tools are available from the corresponding author upon reasonable request.

References

1. Ahmed, R., et al. (2024). The impact of simulation-based training in medical education: A review. Medicine, 103(27), e38813.
2. Al Jabri, Y. M., Kvist, T., Azimirad, M., & Turunen, H. (2021). A systematic review of healthcare professionals' core competency instruments. Nursing & Health Sciences, 23(1), 87–102.
3. Al-Hashimi, M., et al. (2023). Formative objective structured clinical examinations as an assessment tool in UK undergraduate medical education: A review of its utility. Cureus, 15(5), e38676.
4. Anderson, H. L., Kurtz, J., & West, D. C. (2021). Implementation and use of workplace-based assessment in clinical learning environments: A scoping review. Academic Medicine, 96(11S), S164–S174.
5. Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology, 8(1), 19–32.
6. Ayanian, J. Z., & Markel, H. (2016). Donabedian's lasting framework for health care quality. New England Journal of Medicine, 375(3), 205–207.
7. Bedelu, M., Ford, N., Hilderbrand, K., & Reuter, H. (2007). Implementing antiretroviral therapy in rural communities: The Lusikisiki model of decentralized HIV/AIDS care. Journal of Infectious Diseases, 196(Suppl 3), S464–S468.
8. Billiau, L., Taghon, D., Duprez, V., Eeckloo, K., & Malfait, S. (2025). Instruments for measuring head nurses' competencies in a hospital setting: A scoping review. BMC Nursing, 24, 1069.
9. Brook, R. H., & Lohr, K. N. (1985). Efficacy, effectiveness, variations, and quality: Boundary-crossing research. Medical Care, 23(5), 710–722.

10. Chen, Y. C., et al. (2024). Can routine EPA-based assessments predict OSCE performances of undergraduate medical students? Medical Teacher, 46(9), 1134–1143.
11. Cheng, A., et al. (2024). Debriefing methods for simulation in healthcare: A systematic review. Simulation in Healthcare, 19(1S), S112–S121.
12. Chopra, M., et al. (2008). Achieving the health Millennium Development Goals for South Africa: Challenges and priorities. The Lancet, 374(9694), 1023–1031.
13. Colliver, J. A., & Williams, R. G. (1993). Technical issues: Test application. Academic Medicine, 68(6), 454–463.
14. Crowley, T., & Mayers, P. (2015). Trends in task shifting in HIV treatment in Africa: Effectiveness, challenges and acceptability to the health professions. African Journal of Primary Health Care & Family Medicine, 7(1), 807.
15. Das, J., & Hammer, J. (2005). Which doctor? Combining vignettes and item response to measure clinical competence. Journal of Development Economics, 78(2), 348–383.
16. Das, J., & Leonard, K. (2006). Using vignettes to measure the quality of health care. World Bank Policy Research Working Paper.
17. De Rosis, S., et al. (2024). Performance measurement and user-centeredness in the healthcare sector. International Journal of Health Planning and Management, 39(4), 1172–1182.
18. Dewan, S., Khalil, A. N., & Gupta, A. (2024). OSCEs in 2024: Navigating current challenges and enhancing future assessment strategies. Korean Journal of Medical Education, 36(2), 135–142.
19. Donabedian, A. (1966). Evaluating the quality of medical care. Milbank Memorial Fund Quarterly, 44(3), 166–206.
20. Donabedian, A. (2005). Evaluating the quality of medical care. Milbank Quarterly, 83(4), 691–729.
21. Dong, C., Low, S. C., & Yan, C. C. (2024). Workplace-based assessment in healthcare: Key concepts and best practices. Proceedings of Singapore Healthcare, 33, 20101058241310618.
22. Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. JAMA, 287(2), 226–235.
23. Fairall, L., et al. (2012). Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH). The Lancet, 380(9845), 889–898.
24. Filomeno, L., et al. (2024). Systematic review and psychometric properties analysis of first-, middle-, and top-level nurse manager's core competencies instruments. Journal of Nursing Management, 2024, 2655382.
25. Giesey, R. L., Chen, O., & Engelman, D. (2020). Assessment of clinical competence in competency-based education. Journal of Dental Hygiene, 94(5), 27–34.
26. González García, A., et al. (2024). Use of artificial intelligence in the design and implementation of simulation scenarios in health. BMC Medical Education, 24, 414.
27. Harden, R. M., & Gleeson, F. A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE). Medical Education, 13(1), 41–54.

28. Hopkins, H., Talisuna, A., Whitty, C. J., & Staedke, S. G. (1996). Impact of home-based management of malaria on health outcomes in Africa: A systematic review of the evidence. Malaria Journal, 6, 134.
29. Humphrey-Murto, S., et al. (2024). Design and implementation of a national program of assessment model – Integrating EPA assessments in Canadian specialist postgraduate medical education. Perspectives on Medical Education, 13, 95–102.
30. Katz, H. P., et al. (1996). The role of the medical record in quality measurement. Medical Care, 34(4), 346–359.
31. Laakkonen, N., et al. (2024). Digital competence among healthcare leaders: A mixed-methods systematic review. Journal of Nursing Management, 2024, 8435248.
32. Lawthers, A. G., et al. (1993). Measuring quality of care: A review. Quality Assurance in Health Care, 5(4), 251–258.
33. Leonard, K. L., & Masatu, M. C. (2005). The use of direct clinician observation and vignettes for health services quality evaluation in developing countries. Social Science & Medicine, 61(9), 1944–1951.
34. Lockyer, J., et al. (2017). Core principles of assessment in competency-based medical education. Medical Teacher, 39(6), 609–616.
35. Luck, J., Peabody, J. W., Dresselhaus, T. R., Lee, M., & Glassman, P. (2000). How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. American Journal of Medicine, 108(8), 642–649.
36. Mainz, A., Nitsche, J., Weirauch, V., & Meister, S. (2024). Measuring the digital competence of health professionals: Scoping review. JMIR Medical Education, 10, e55737.
37. Mant, J. (2001). Process versus outcome indicators in the assessment of quality of health care. International Journal for Quality in Health Care, 13(6), 475–480.
38. Miller, G. E. (1990). The assessment of clinical skills/competence/performance. Academic Medicine, 65(9 Suppl), S63–S67.
39. Mink, R. B., et al. (2022). Entrustable professional activities and subspecialty training. Pediatrics, 150(6), e2022058953.
40. Mokkink, L. B., et al. (2018). COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. Quality of Life Research, 27(5), 1171–1179.
41. Nawathe, A. C., & Herrington, C. M. (2024). How I Do It: A pediatric surgery standardized patient program. Journal of Pediatric Surgery Open, 9, 100138.
42. Nolan, T., et al. (2001). Quality of hospital care for seriously ill children in less-developed countries. The Lancet, 357(9250), 106–110.
43. Norcini, J. J., & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. Medical Teacher, 29(9–10), 855–871.

44. Norman, G. R., Neufeld, V. R., Walsh, A., Woodward, C. A., & McConvey, G. A. (1985). Measuring physicians' performances by using simulated patients. Journal of Medical Education, 60(12), 925–934.
45. Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. Systematic Reviews, 5(1), 210.
46. Page, M. J., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71.
47. Peabody, J. W., Luck, J., Glassman, P., Dresselhaus, T. R., & Lee, M. (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. JAMA, 283(13), 1715–1722.
48. Peabody, J. W., Tozija, F., Munoz, J. A., Nordyke, R. J., & Luck, J. (2004). Using vignettes to compare the quality of clinical care variation in economically divergent countries. Health Services Research, 39(6p2), 1951–1970.
49. Pearson, D. (2024). AI-enhanced simulation in healthcare education: Current applications and future directions. Simulation in Healthcare, 19(2), 89–95.
50. Philips, M., Zachariah, R., & Venis, S. (2008). Task shifting for antiretroviral treatment delivery in sub-Saharan Africa: Not a panacea. The Lancet, 371(9613), 682–684.
51. Régent, A., Thampy, H., & Singh, M. (2023). Assessing clinical reasoning in the OSCE: Pilot-testing a novel oral debrief exercise. BMC Medical Education, 23, 718.
52. Rethans, J. J., Norcini, J. J., et al. (2002). The relationship between competence and performance: Implications for assessing practice performance. Medical Education, 36(10), 901–909.
53. Robinson, S. J. A., et al. (2024). Simulation-based education of health workers in low- and middle-income countries: A systematic review. Global Health: Science and Practice, 12(6), e2400187.
54. Rusli, K. D. B., et al. (2024). A hybrid systematic narrative review of instruments measuring home-based care nurses' competency. Journal of Advanced Nursing, 80(7), 2647–2671.
55. Shea, B. J., et al. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews. BMJ, 358, j4008.
56. Stillman, P. L., Swanson, D. B., et al. (1986). Assessing clinical skills of residents with standardized patients. Annals of Internal Medicine, 105(5), 762–771.
57. Suzan-Monti, M., Blanche, J., et al. (2015). Quality of life conditions in antiretroviral-treated patients from Cameroon. Health and Quality of Life Outcomes, 13, 178.
58. Tamblyn, R. M., Klass, D. J., Schnabl, G. K., & Kopelow, M. L. (1991). The accuracy of standardized patient presentation. Medical Education, 25(2), 100–109.
59. ten Cate, O. (2013). Nuts and bolts of entrustable professional activities. Journal of Graduate Medical Education, 5(1), 157–158.
60. ten Cate, O., & Carraccio, C. (2019). Envisioning a true continuum of competency-based medical education, training, and practice. Academic Medicine, 94(9), 1283–1288.

61. van der Vleuten, C. P., et al. (2012). A model for programmatic assessment fit for purpose. Medical Teacher, 34(3), 205–214.
62. Veloski, J., Tai, S., Evans, A. S., & Nash, D. B. (2005). Clinical vignette-based surveys: A tool for assessing physician practice variation. American Journal of Medical Quality, 20(3), 151–157.
63. Vhora, A., Davies, R. L., & Rice, K. (2024). The psychometric quality of objective structured clinical examinations within psychology programs: A systematic review. Teaching of Psychology, 51(2), 145–158.
64. Warm, E. J., et al. (2014). The next steps in competency-based medical education: Milestones, entrustable professional activities and observable practice activities. Journal of General Internal Medicine, 29(8), 1090–1092.
65. Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. Journal of Advanced Nursing, 52(5), 546–553.
66. Wu, A. W., & Ashton, C. M. (1997). An overview of quality assessment and assurance. In K. N. Lohr (Ed.), Medicare: A strategy for quality assurance (Vol. 1). National Academies Press.
67. Yang, J., et al. (2025). Application of Donabedian three-dimensional model in outpatient care quality: A scoping review. Journal of Nursing Management, 2025, 6893336.
68. Zachariah, R., Teck, R., et al. (2007). Community support is associated with better antiretroviral treatment outcomes in a resource-limited rural district in Malawi. Transactions of the Royal Society of Tropical Medicine and Hygiene, 101(1), 79–84.