Revista Española
de Educación Médica

UNIVERSIDAD DE MURCIA

Methods of clinical competency assessment in Medical Internships: review of the scope of current practices and emerging trends.

Métodos de evaluación de competencias clínicas en Internados Médicos: revisión de alcance de prácticas actuales y tendencias emergentes.

Nicolás Lavados-Toro [1] *, Joaquín Salgado-González [2], Silvana Jiménez-Vera [3], Álvaro Herrera-Alcaíno [2], Oscar Jerez-Yañez [4,5]

[1] Faculty of Medicine, University of Chile, Chile, nicolas lavados@ug.uchile.cl, https://orcid.org/0009-0005-4223-3288 ; [2] Faculty of Medicine, University of Chile, Chile, joaquin.salgado.g@ug.uchile.cl, https://orcid.org/0009-0005-3307-1440 ; [3] Faculty of Medicine, University of Chile, Chile, silvanajimenez@ug.uchile.cl, https://orcid.org/0000-0003-3730-7447 ; [4] Faculty of Medicine, San Sebastián University, Chile, levarito@uchile.cl, https://orcid.org/0009-0007-4861-2144 ; [5] Department of Health Sciences Education, Faculty of Medicine, University of Chile, Chile, ojerez@uchile.cl, https://orcid.org/0000-0003-0869-5938

* Correspondence: nicolaslavados@ug.uchile.cl

Received: 21/12/25; Accepted: 28/1/26; Published: 29/1/26

Summary.

The aim of this scoping review was to map the methods used to assess clinical competence during medical internships and to identify emerging trends associated with the integration of educational technologies. A search was conducted in PubMed, Web of Science, and Scopus (October 2024), including studies published in the last 10 years that assessed clinical competence using technology-supported, hands-on methods. Twenty-six studies were selected, involving 5,749 medical students in diverse clinical settings. The methods identified included Objective Structured Clinical Examinations (OSCE), Mini -CEX, and other Clinical Environment-Based Assessments, including assessments based on Trusted Professional Activities (EPA). Emerging technologies integrated included high-fidelity simulation, virtual reality, telemedicine, and digital feedback systems. Overall, the studies reported improvements in clinical skills, diagnostic reasoning, communication, and self-confidence, particularly when immediate feedback was used in formative assessments. Challenges persist related to inter-rater variability, methodological heterogeneity, and the need for adequate teacher training and technological resources. In conclusion, the findings demonstrate a transition toward assessment models characteristic of competency-based medical education, highlighting the importance of direct observation, structured feedback, and the strategic use of technology. It is recommended to strengthen the standardization of assessment practices and develop research to evaluate their long-term impact.
Keywords: clinical competence, medical internship, OSCE, Mini -CEX, formative assessment, educational technology.

Abstract.

This scoping review aimed to map current methods used to assess clinical competence during medical clerkships and to identify emerging trends related to the integration of educational technologies. Searches were conducted in PubMed, Web of Science, and Scopus (October 2024), including studies from the past 10 years evaluating clinical competence through practical, technology -supported assessments. A total of 26 studies involving 5,749 medical students across

diverse clinical settings were included. Identified assessment methods comprised Objective Structured Clinical Examinations (OSCEs), Mini -Clinical Evaluation Exercises (Mini -CEX), and multiple Workplace -Based Assessments, including Entrustable Professional Activity (EPA) -based evaluations. Emerging technologies—high -fidelity simulation, virtual reality, telemedicine, and digital feedback systems—were increasingly incorporated. Overall, studies showed improvements in clinical skills, diagnostic reasoning, communication, and self --confidence, particularly when immediate formative feedback was provided. Persistent challenges included evaluator variability, methodological heterogeneity, and the need for faculty training and adequate technological resources. In conclusion, findings reflect a shift toward competence -based medical education frameworks emphasizing direct observation, structured feedback, and strategic technological integration. Strengthening standardization of assessment practices and generating longitudinal evidence on long -term impacts are recommended.
Keywords: clinical competence, medical clerkship, OSCE, Mini -CEX, formative assessment, educational technology.

1. Introduction

The assessment of clinical competence during medical internships has undergone significant evolution, reflecting broader transformations in medical education over the past few decades (1). Traditionally, assessment processes relied on written and oral examinations, along with performance observation by supervising physicians—approaches that are useful but limited in capturing the complexity of clinical practice (2–3). In response, competency-based medical education (CBME) has gained momentum, integrating assessments situated in the real clinical setting grouped under the Workplace -Based Assessments (WBA) framework, which includes tools such as the Mini Clinical -Evaluation Exercise (Mini -CEX), Direct Observation of Procedural Skills (DOPS), and assessments based on Entrustable Professional Activities (EPA) (4–7). Furthermore, simulation is complemented by methods such as the Objective Structured Clinical Examination (OSCE), in accordance with widely used competency frameworks (e.g., ACGME) that organize domains such as patient care, medical knowledge, interpersonal and communication skills, professionalism, practice-based learning and improvement, and systems-based practice (3,4,8). In parallel, the incorporation of technologies—digital platforms, high-fidelity simulators, virtual reality (VR), and telemedicine—has expanded assessment possibilities, generating standardized and safe experiences and enhancing reflective learning with timely feedback (9–13).

However, its effective adoption requires adequate infrastructure, teacher training, and scaling strategies (14). To ensure international transferability and contextual adaptation, this work is framed within the World Federation for Medical Education (WFME) global standards for basic medical education (2020 edition), which are principles- -based and non-prescriptive, proposing cross-cutting domains (curriculum, assessment, resources, quality assurance, and governance) as levers for improvement applicable to different institutional and regional realities (15). Given this rapid transformation—driven, in part, by technological innovations—it becomes necessary to systematically map current methodologies for assessing clinical competencies in internships.

This scoping review seeks to answer the following question: What assessment methods are used to support the development of clinical competence in medical students during internships, and what trends are emerging in the implementation of technology in these methods? The specific objectives are: (i) to explore the existing literature on clinical competence assessment methods used in medical internships; (ii) to identify emerging trends in technological tools integrated into these assessments; and (iii) to characterize methods that articulate formative and summative processes, identifying strengths, weaknesses, and gaps in the literature. For the purposes of this manuscript, the term medical internship refers to the final clinical phase of undergraduate training, equivalent to medical clerkship in English-speaking countries, characterized by the student's active participation in supervised clinical activities.

2. Methods

This review was designed to map and characterize the assessment methods used to measure clinical competence during medical internships, following the PRISMA-SCR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines (14). Studies published within the last 10 years that included methods for both formative and summative assessment of clinical competence during medical internships using technological tools were included. Studies that were theoretical or lacked application in clinical practice, as well as studies conducted in preclinical or postgraduate settings, were excluded. The search was conducted in October 2024 in PubMed, Web of Science, and Scopus, using specific keyword combinations, as shown in Table 1.

Table 1. Search strategy used in each database.

Databases	Terms used in the search	Combinations used
Web of Science	#1 ALL=(("medical student" OR "medical students") AND ("clinical clerkship" OR "medical clerkship" OR "clinical rotation")	#4: #1 AND #2 AND #3
	#2 ("competence assessment" OR "competency evaluation" OR "clinical competence assessment" OR "formative assessment" OR "summative evaluation" OR "feedback")	
	#3 ("workplace-based assessment" OR "portfolio assessment" OR "OSCE" OR "Mini-CEX" OR "Direct Observation of Procedural Skills" OR "DOPS" OR "simulation-based assessment" OR "digital assessment tools" OR "virtual reality" OR "high-fidelity simulators"))	
PubMed	#1 (("medical student"(All Fields) OR "medical students"(All Fields)) AND ("clinical clerkship"(All Fields) OR "medical clerkship"(All Fields) OR "clinical rotation"(All Fields))	#4: #1 AND #2 AND #3
	#2 ("competence assessment"(All Fields) OR "competency evaluation"(All Fields) OR "clinical competence assessment"(All Fields) OR "formative assessment"(All Fields) OR "summative evaluation"(All Fields) OR "feedback"(All Fields))	
	#3 ("workplace-based assessment"(All Fields) OR "portfolio assessment"(All Fields) OR "OSCE"(All Fields) OR "Mini-CEX"(All Fields) OR "Direct Observation of Procedural Skills"(All Fields) OR "DOPS"(All Fields) OR "simulation-based assessment"(All Fields) OR "digital assessment tools"(All Fields) OR "virtual reality"(All Fields) OR "high-fidelity simulators"(All Fields)))	
Scopus	#1 TITLE-ABS-KEY (("medical student" OR "medical students") AND ("clinical clerkship" OR "medical clerkship" OR "clinical rotation")	#4: #1 AND #2 AND #3
	#2 ("competence assessment" OR "competency evaluation" OR "clinical competence assessment" OR "formative assessment" OR "summative evaluation" OR "feedback")	

	#3 ("workplace-based assessment" OR "portfolio assessment" OR "OSCE" OR "Mini-CEX" OR "Direct Observation of Procedural Skills" OR "DOPS" OR "simulation-based assessment" OR "digital assessment tools" OR "virtual reality" OR "high-fidelity simulators"))

OSCE: Objective Structured Clinical Examinations; MINI CEX: Mini Clinical Evaluation Exercise; DOBS: Direct Observation of Procedural Skills. The search was limited to articles published up to October 27, 2024.

Selection process

Initially, the titles and abstracts of all identified studies were evaluated. Duplicates were removed using Rayyan software (16). Subsequently, the studies considered relevant underwent full text review. Each article was evaluated independently, and in cases of discrepancies or doubts, a reconsideration was carried out to ensure the consistent application of the criteria.

Data collection process

Data from the selected studies were extracted using a predefined table in Google Sheets, which included information such as authors, year, country, sample size, participants' level, area of competence assessed, assessment method, technology and tools used, and innovative elements. A second table was also used to record whether the assessment was formative or summative, the feedback mechanism, the main findings, student perceptions, conclusions, and implications for practice.

Data elements

For this review, data were collected across a wide range of outcomes, including: Assessment methods—WBAs and their component tools (Mini -CEX, DOPS, EPA -based assessments), as well as other methods such as OSCEs and simulations; Competency domains—clinical reasoning, communication, professionalism, teamwork, technical skills; Feedback mechanisms—type (formative or summative) and delivery channels; Technological tools—digital platforms, virtual reality, high -fidelity simulators, telemedicine; Innovative components—unique or distinctive elements of each study; Student perceptions and practical recommendations.

Risk assessment of bias in studies

Although scoping reviews do not require a formal assessment of the risk of bias, the MERSQI was included to describe the methodological quality of the studies and contextualize the interpretation of the findings, providing a deeper understanding of the strength of the available evidence. The methodological quality of the studies was assessed using the Medical Education Research Study Quality Instrument (MERSQI) (17). This assessment helps mitigate potential biases present in the studies, particularly those related to reporting bias.

Measures of effect

Measures of effect were systematically documented for each of the outcomes studied, considering means, standard deviations, correlation coefficients, reliability indices (Cronbach's alpha), and significance tests (ANOVA, Wilcoxon tests). Measures were classified according to their statistical purpose, such as assessing reliability, correlation, or thematic consistency in the findings. This approach facilitated a comparison of clinical outcomes and assessment methods.

Synthesis methods

To determine the eligibility of studies for the synthesis, the characteristics of the interventions and outcomes were tabulated in Google Sheets and compared against predefined criteria, allowing for coherent grouping of studies. Eligible studies were rigorously processed to standardize data formats. The results were then presented using thematic summaries, enabling comparative groupings and cross-sectional analysis.

Statement

During this work, the authors used Microsoft Copilot 365 for data extraction and ChatGPT4o to synthesize the extracted results and subsequently to improve the writing and translate the original article into Spanish. After using this tool, the authors reviewed and edited the content as needed and assume full responsibility for the content of the publication. Furthermore, iterative human reviews were applied to minimize potential biases, ensure conceptual coherence, and maintain fidelity to the original data extracted from the included studies.

Statement of Ethics

This study is a literature-based review, therefore, it did not require approval from an institutional ethics committee or informed consent.

3. Results

One hundred and twenty-seven studies were identified (Figure 1) by searching the three electronic databases. After removing duplicates, 45 articles were selected for full-text review. Finally, 26 studies met the eligibility criteria for data extraction and analysis (18–43).

Characteristics of the studies

Studies conducted in several countries were included, mostly from the USA (14 studies: 18, 20, 24, 25, 27, 31, 32, 34, 36, 38–40, 42, 43), followed by Japan (4 studies: 22, 23, 29, 41), Norway (21, 28), Switzerland (33, 37), Pakistan (35), Australia (30), Chile (19), and China (26) (Figure 2). The 26 studies included a total of 5,749 medical students, with sample sizes ranging from 13 participants in pilot studies (32) to 1,810 participants in large cross-sectional studies (40). Most were published between 2020 and 2024, representing 65.4% of the reviewed literature.

Methodological designs included randomized controlled trials (RCTs) (26, 28, 35), quasi-experimental studies such as pre-post studies (24, 33, 38), non-randomized clinical trials (27), observational designs, such as cross-sectional studies (18, 30, 31, 39, 40), longitudinal studies (19, 22), prospective cohort studies (23, 29), prospective observational studies (41) and mixed studies (42), in addition to pilot studies (20, 25, 32, 36), qualitative studies (21, 34), a multilevel analysis (37) and a validation study (43).

The studies focused primarily on specific clinical residencies, such as Emergency Medicine (18, 27), Internal Medicine (19, 33, 43), and specialized residencies such as Obstetrics and Gynecology (OB/GYN) (36), Psychiatry (25), and Surgery (26, 31, 42). Some multi-setting studies (24, 37) offered different perspectives by evaluating tools in various clinical environments and disciplines (Figure 3). A complete characterization of the studies is presented in the Supplementary Material (Table S1). The marked heterogeneity in designs (RCTs, cross-sectional studies, pilot studies, qualitative studies), sample sizes, and types of outcomes assessed precluded direct comparisons between studies and limits the possibility of synthesizing robust quantitative trends.

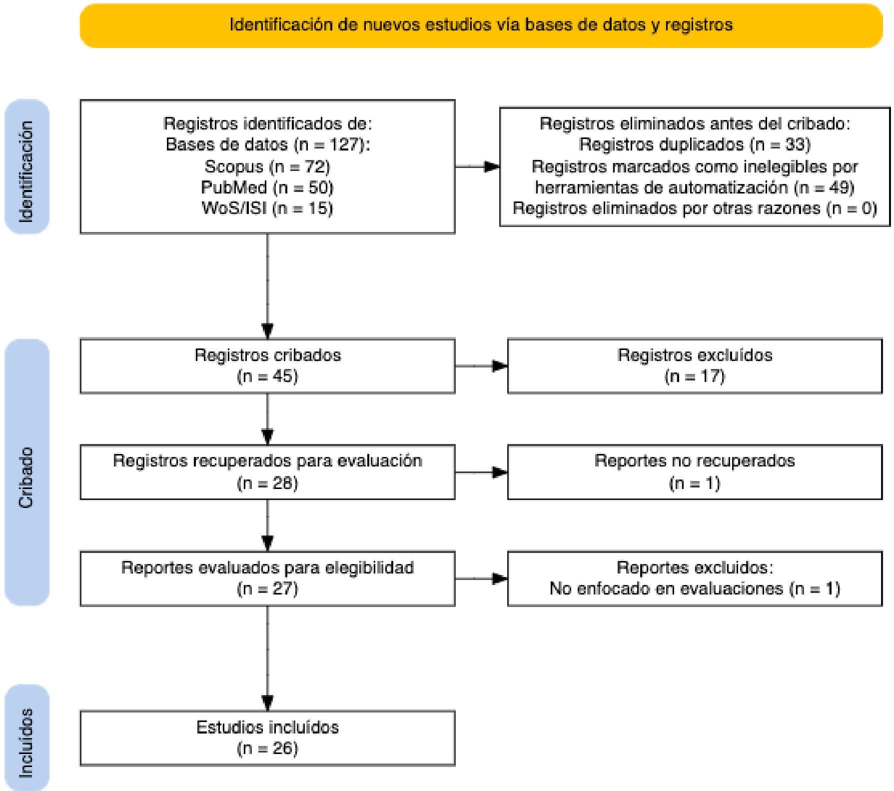

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the selected studies.

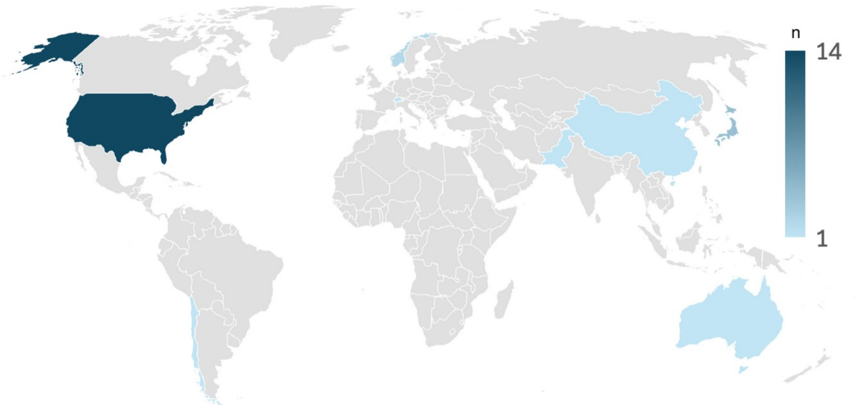

Figure 2. Map of the distribution of studies by country.

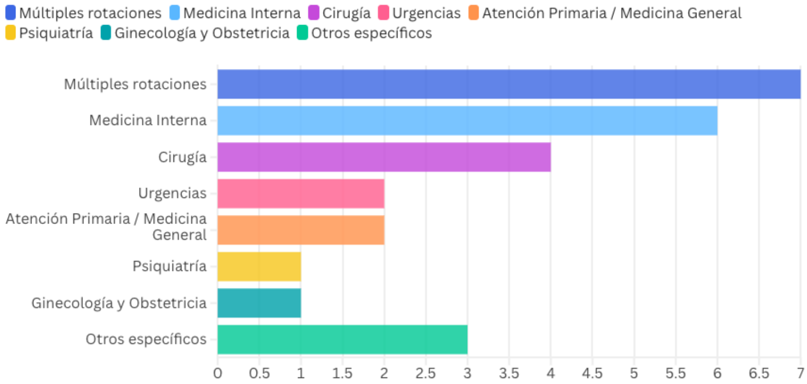

Figure 3. Clinical scenario. *Includes scenarios such as medical interviews in clinical practice, specific outpatient units, or inpatient stays not categorizable in a specific specialty.

Risk of bias

MERSQI scores ranged from 6 to 18 points, calculated across the six assessed domains. The mean score across the included studies was 12.1 (range: 10.5 to 18). Seven studies recruited participants from multiple institutions, most achieving a response rate above 75%, suggesting moderate to high methodological quality, particularly in studies with experimental approaches and well-structured designs. However, some studies and pilot studies had limitations in instrument validity, sample size, and scope.

Individual outcomes: effect measures

For all outcomes, summary statistics and effect measures with their respective precision were extracted, which are presented in Table 2. For example, Bord et al. (2015) reported a mean OSCE score of 70.5% (SD = 7.2%), with a mean point-biserial correlation of 0.24, indicating moderate discriminatory ability among students (18). Bozzo et al. (2020) observed significant improvements in clinical skills during internal medicine residency, with a Spearman correlation of 0.739 ($p < 0.0001$) and a Cronbach's alpha of 0.8 (19). Costich et al. (2024) demonstrated an increase in the delivery of specific, task-oriented feedback ($p = 0.006$) in assessments based on reliable professional activities in a pediatric primary care setting (20). Shikino et al. (2023) used a voice recognition system to generate feedback, achieving an increase in diagnostic accuracy from 51.3% to 89.7%, with improvements in Mini-CEX scores and checklists (41). Luo et al. (2023) observed improvements in self-confidence and clinical competencies, although procedural skills did not improve significantly (26). These results show how various educational interventions positively influenced the development of clinical competence.

Areas of competence

The studies assessed a range of competencies in medical training, highlighting not only traditional clinical skills, but also professional and communicational aspects (Figure 4).

- Clinical performance and professionalism: studies such as those by Kasai et al. (2020) and Shikino et al. (2023) addressed clinical performance and professionalism, emphasizing decision-making and professional reasoning (23, 41).
- Clinical skills: Research such as that by Bord et al. (2015), Klapheke et al. (2022) and Kim et al. (2016) explored general clinical competencies, while Malone et al. (2024) and Martinsen et al. (2021) delved into specific skills such as emergency management, history taking and physical examination (18, 24, 25, 27, 28).
- Communication and interpersonal interaction skills in sensitive settings, such as end-of-life care, were addressed by the studies of Qureshi & Zehra (2020) and Parikh et al. (2015).

Empathy and communication were also highlighted in the work of Bozzo et al. (2020) and Patel et al. (2024) (19, 31, 32, 35).

- Teamwork and collaboration: the study by Olupeliyawa et al. (2014) highlighted the importance of collaborative skills (30).
- Feedback and supervision: studies such as those by Gran et al. (2016) and Phinney et al. (2022) highlighted the value of structured feedback and self-reflection as fundamental mechanisms for clinical learning (21, 34).
- Reliable Professional Activities: were used as a reference in studies such as Costich et al. (2024) and Ryan et al. (2021, 2024), linking specific clinical activities with the development of observable and measurable competencies (20, 39, 40).
- Emerging competencies: For example, Reid et al. (2021) incorporated menopause management through telemedicine, expanding the scope of clinical skills (36).

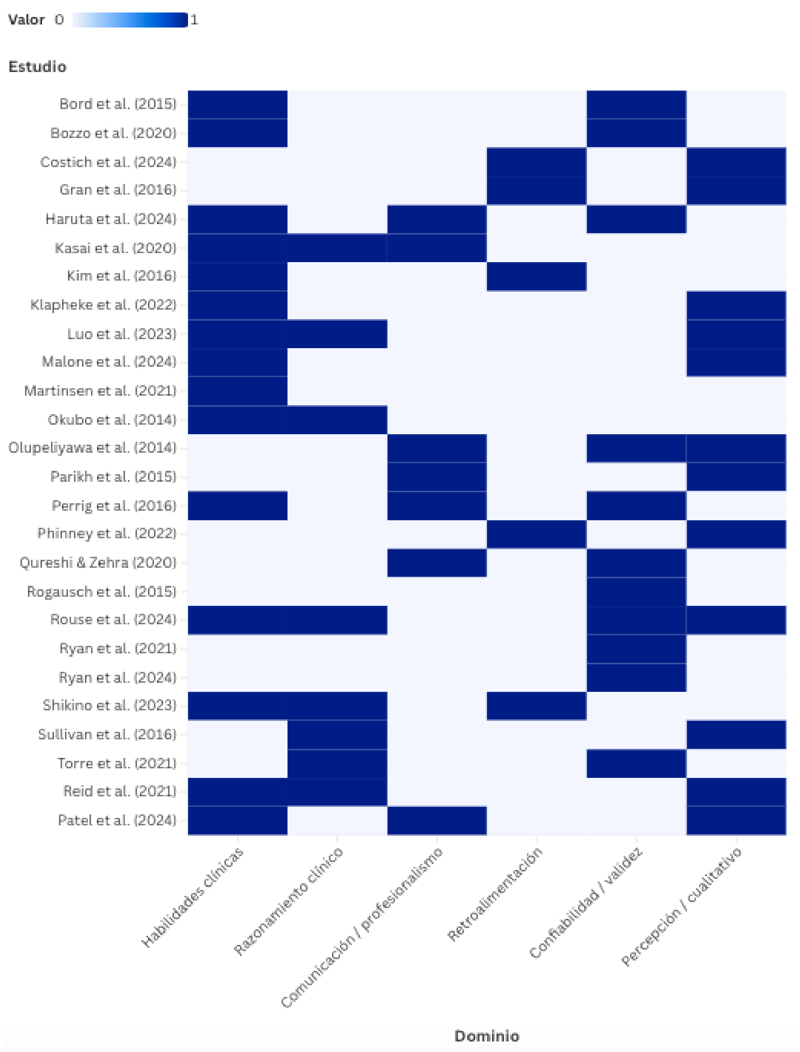

Figure 4. Heatmap of assessed competency areas.

Table 2. Results in terms of outcomes and effect measures of the included studies.

Study	Result	Measure(s) of the effect
Bord et al. (2015)	Development of an OSCE for the assessment of clinical skills in the emergency internship	Average score: 70.5% (SD = 7.2%), Item difficulty: Low (>80% correct), Medium (50-80% correct), High (<50% correct), Item discrimination: Good (r_pb > 0.3), Fair (r_pb = 0.1-0.3), Poor (r_pb < 0.1), Point biserial correlation (r_pb): 0.24 (average)
Bozzo et al. (2020)	Improvement in clinical skills in internal medicine	Measures of central tendency, normality tests, paired Wilcoxon tests (p < 0.05), multiple linear regression models, Spearman correlation (r = 0.739, p < 0.0001), Cronbach's alpha (0.8), p-values (p ≤ 0.05)
Costich et al. (2024)	Implementation and teacher perception of EBA based on EPA in outpatient primary care during pediatric internship	Teacher feedback: Increased delivery of specific and task-oriented feedback (p = 0.006), greater satisfaction with feedback opportunities (not significant), more feedback within 24 hours (not significant). Interview topics: Benefits of EBA, barriers to feedback, suggestions for improvement
Gran et al. (2016)	Feedback experiences during the internship	Thematic analysis to identify key themes and feedback experiences, consistency of themes across interviews
Haruta et al. (2024)	Validation of the Simulated Patient Assessment Tool (SPAT) to evaluate clinical performance with simulated patients	Exploratory factor analysis revealed two factors (communication and medical performance), explaining 60.47% of the variance. Internal consistency: overall Cronbach's alpha 0.929. One-way ANOVA showed significant differences between simulated patients (F(34,760)=16.79, p<0.001) and scenarios (F(20,774)=11.39, p<0.001). Convergent validity showed a moderate correlation (r=0.212, p<0.05) between SPAT and OSCE post-CC
Kasai et al. (2020)	Improvement in clinical performance and professionalism in internship students	Mini-CEX: Significant improvements in medical history taking (pre: 5.52 ±1.05, post: 6.57 ±0.75, p<0.001), physical examination (pre: 5.38 ±0.82, post: 6.78 ±0.87, p<0.001), professionalism (pre: 5.87 ±1.00, post: 7.13 ±0.81, p<0.001), clinical judgment, counseling, organization/efficiency, and overall competence. P-MEX: Improvements in doctor-patient relationship (pre: 3.00 ±0.34, post: 3.38 ±0.24, p<0.001) and reflective skills (pre: 3.15 ±0.32, post: 3.50 ±0.26, p<0.001). Wilcoxon signed-rank test (p<0.05)
Kim et al. (2016)	Implementation of mini-CEX in all third-year internships and its impact on direct observation and clinical skills	Compliance: 92% of forms completed, 78% with specific feedback. Student report: significant increase in direct observation of physical examination (e.g., surgery 49%→87%, p<0.0001). AAMC GQ: more faculty observation in history taking and physical examination. OSCE: decrease in failure rate from 12% to 2% (p=0.0046). p-values: <0.0001 (observation in surgery), 0.0046 (OSCE)
Klapheke et al. (2022)	Pilot study of EBA using EPAs and RIME model in psychiatric internship	Averages in EPAs and RIME: profile >4.0 in EPA1, EPA6, and reporter role in RIME. Student feedback: 44% considered it useful, 32.2% considered the evaluations fair. Teacher feedback: ease of completing EPA = 3.89 (SD=1.27), clarity of explanation = 4.33 (SD=0.5)
Luo et al. (2023)	Improvement in self-confidence and clinical competence in surgical interns	Self-Confidence Assessment (SCA): improvements at all stations (e.g., pre-anamnesis: 2.68 ± 0.82, post-anamnesis: 3.47 ± 0.84, p < 0.01; pre-physical exam: 2.47 ± 0.84, post-physical exam: 3.42 ± 0.77, p < 0.01; CPR, clinical reasoning, surgical asepsis). Mini-CEX: significant improvement (OSCE group: 6.59 ± 0.62 vs. control: 5.31 ± 1.09, p < 0.01). DOPS: no

		significant difference (OSCE group: 5.15 ± 0.58 vs. control: 4.96 ± 0.63, p > 0.05).
Malone et al. (2024)	Skills in urgent and emergency care	Wilcoxon signed-rank test (T = 337.5, p = 0.001), McNemar test (p ≤ 0.001), thematic analysis for qualitative data, significance determined by p ≤ 0.05
Martinsen et al. (2021)	Clinical skills measured by mini-CEX assessments	ANOVA (F = 3.603, p = 0.066), ANCOVA (F = 1.884, p = 0.179), average scores (3.5–3.6 out of 4), standard deviations (0.55–0.63)
Okubo et al. (2014)	Evaluation of an outpatient clinical internship program to improve clinical reasoning in students	Mini-CEX: higher in all areas for students in the program. SCT: significantly higher post-course (pre: 77.4 ±10.5; post: 86.8 ±10.2). OSCE: higher scores for intervention group vs. control group
Olupeliyawa et al. (2014)	Evaluation of the educational impact of the T-MEX on student collaboration in healthcare teams	Generalization coefficient: 0.62 with three forms, 0.80 predicted with eight forms. Content analysis of feedback and reflections. Thematic analysis of focus groups and interviews.
Parikh et al. (2015)	Communication, empathy and trust in end-of-life care	Descriptive statistics of average scores (mean = 89.0%, SD = 6.7%), standard deviations, correlation with confidence scores (r = 0.325, p < 0.01) and communication skills (r = 0.383, p < 0.01)
Perrig et al. (2016)	Improvement in musculoskeletal examination skills	Friedman test (p < 0.001 for CS, p < 0.001 for MSES, p < 0.01 for IPS), Wilcoxon rank-sum test (CS: p < 0.001, MSES: p < 0.001, IPS: p < 0.001), Mann-Whitney U test (CS: p < 0.01, MSES: p < 0.01, IPS: p < 0.01), Cronbach's alpha (0.47–0.83), p-values (p ≤ 0.05)
Phinney et al. (2022)	Feedback and self-reflection in clinical settings	Thematic analysis to identify key themes and tensions, consistency of themes across interviews
Qureshi & Zehra (2020)	Communication skills using feedback from simulated patients	Pearson correlation coefficient (r = 0.83) for internal consistency of LCSAS (Cronbach's alpha = 0.83), improvement of pre/post-test scores (mean difference = 1.5, p ≤ 0.05)
Rogausch et al. (2015)	Influence of prior clinical skills and contextual characteristics on mini-CEX scores in internships	Regression analysis: Clinical position of the instructor was the most influential predictor (regression coefficient = 0.55, p < 0.001 for residents vs. department heads). Task complexity and clinic size were significant predictors. OSCE was not a significant predictor. Correlation was weak between OSCE and mini-CEX (r = 0.26 overall, r = 0.27 per domain). p-values: p < 0.001 for instructor position, p < 0.05 for complexity and size
Rouse et al. (2024)	Improved OSCE to neutralize grade inflation and provide a more comprehensive assessment of clinical skills	Comparison of OSCE pre and post-intervention (pre: M = 94.25%, SD = 5.65%, post: M = 81.00%, SD = 6.88%), student surveys (average: 4.4 Reporter, 4.4 Interpreter, 4.2 Manager, 3.5 time allotted, 3.5 difficulty)
Ryan et al. (2021)	Central EPAs measured by O-SCORE scale	Generalization theory for reliability (Phi coefficient = 0.19–0.44), average scores (3.48–3.62), variance components (student = 3.5%–8%, evaluator = 29.6%–50.3%)

Ryan et al. (2024)	EBA reliability assessment for summative confidence decisions in medical education	Phi coefficient: >0.7, acceptable reliability threshold. Variance attributed to the learner: <10% in most analyses. Number of observations required: range 3 to >560, median 60
Shikino et al. (2023)	Evaluation of the effectiveness of SRS feedback in clinical reasoning during simulated encounters	Diagnostic accuracy: SRS group higher (pre: 51.3%, post: 89.7%) vs IC recorder (pre: 57.5%, post: 67.5%) (p=0.037). Mini-CEX: significant improvements in history taking, physical examination, professionalism, organization/efficiency, and overall competence (p<0.001). Checklist: total score higher in SRS (pre: 12.2, post: 16.1) vs IC recorder (pre: 13.1, post: 13.8) (p<0.001). Feedback time: SRS 22.6 ±2.1 min vs IC 27.7 ±2.1 min (p=0.04)
Sullivan et al. (2016)	Improvement in clinical decision-making skills	Repeated measures ANOVA (acute diverticulitis: Pillai's Trace = 0.807, F(2,36) = 75.279, p<0.000; GI hemorrhage: Pillai's Trace = 0.822, F(2,19)=43.941, p<0.000), thematic analysis, p-values (p ≤ 0.05)
Torre et al. (2021)	Validation of the Multistep Exam (MSX) to evaluate analytical clinical reasoning in internal medicine internships	Correlation: MSX showed a significant positive correlation with Step 2 CS ICE (r=0.26, p<0.01). Multiple linear regression: MSX was a significant predictor of Step 2 CS ICE (β=0.19, p<0.001), explaining an additional 4% of variance beyond NBME Medicine and OSCE Medicine. Reliability: Cronbach's alpha 0.70–0.80
Reid et al. (2021)	OSCE evaluation of telemedicine for menopause management in Gynecology/Obstetrics internships	Post-meeting grade scores (median: 20/45), student surveys (78% discomfort with telemedicine, 66% excellent or higher educational value), differential diagnosis accuracy (100% identified menopause/perimenopause, 84% hyperthyroidism)
Patel et al. (2024)	Dual coaching assessment of anamnesis skills in internship students	Minicard scores (no significant improvement), student surveys (average 1.43; 1=Excellent, 5=Poor), patient surveys (average 1.23; 1=Excellent, 5=Poor), teacher surveys (average 1.69; 1=Excellent, 5=Poor)

OSCE: Objective Structured Clinical Examinations; **EPA:** Entrustable Professional Activities; **SPAT:** Simulated Patient Assessment Tool; **Mini -CEX:** Mini Clinical Evaluation Exercise; **RIME:** Reporter Interpreter Manager Educator; **SCA:** Self -Confidence Assessment; **ANOVA:** Analysis of Variance; **ANCOVA:** Analysis of Covariance; **SRS:** Student Response Systems; **MSX:** Multistep Exam; **WBA:** Workplace -Based Assessments

Evaluation methods

The reviewed studies applied a variety of evaluative approaches, tailored to specific learning objectives and skills (Figure 5).

- Objective Structured Clinical Examination (OSCE): widely adopted to assess clinical skills, communication, and professionalism. Studies such as Bord, Bozzo, and Parikh used it with standardized patients to increase realism and clinical relevance (18, 19, 31). Some, such as Rouse et al. (2024), employed multi-scenario formats to assess reasoning and procedural skills (38).
- Within the framework of Workplace -Based Assessments (WBA), tools such as Mini CEX were identified -, a key tool for brief and structured observations of real-time clinical performance used by Kasai, Kim, Martinsen, among others, to assess interaction with patients, diagnostic reasoning and professionalism (23, 24, 28), DOPS and EPA-based assessments.
- Studies that implemented WBA in a broad sense—including EPA -based assessments and real-time feedback—were especially relevant in Costich, Ryan, and Phinney (20, 34, 40).
- Specialized tools: such as the Simulated Patient Assessment Tool validated by Haruta et al. (2024) for clinical interviews, or the virtual surgical cases in Sullivan et al. (2016) that simulate complex clinical decisions (22, 42).
- Collaborative skills assessments: the T-MEX instrument, applied by Olupeliyawa et al., specifically assessed teamwork in clinical contexts (30).

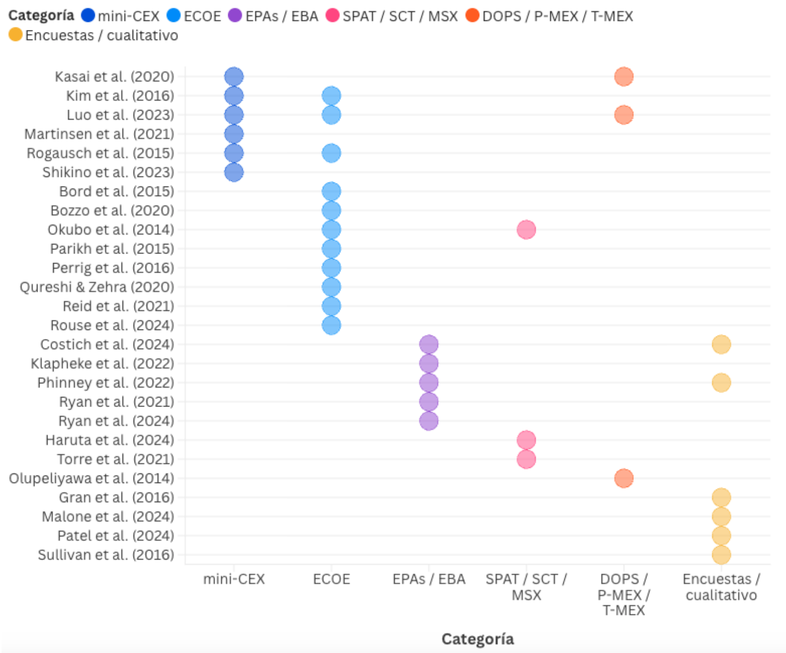

Figure 5. Dotplot of Evaluation Methods used.

Innovative technologies and components

The studies used technologies to strengthen the evaluation processes, focusing on active learning, personalization, and immediate feedback.

- Simulations and virtual tools: highlighting high-fidelity simulators and virtual reality. For example, Malone et al. (2024) compared both types of simulation for emergency training; Bord combined mannequins and simulated patients with immediate feedback in the emergency department (18, 27).

- Role-playing, peer review, and standardized patients (SPs): Kasai used clinical dramatizations for peer feedback. Qureshi and Bozzo integrated SPs to train communication skills (19, 23, 35).

- Voice recognition and digital platforms: Shikino et al. (2023) used an automatic recognition system to generate interview transcripts and provide targeted feedback. Phinney and Ryan used mobile platforms with QR codes to facilitate real-time WBA assessments (34, 39, 41).

- Telemedicine and remote learning: Reid and Patel incorporated tools for virtual clinical encounters, assessing skills in digital communication and remote care (32, 36).

- Structured feedback and self-reflection: Gran implemented the StudentPEP tool to guide reflection after clinical encounters. Luo used the "sandwich" method of immediate feedback after each OSCE station (21, 26).

- Innovative components: from Klapheke's EPA/RIME model to Costich's use of performance-oriented training, these innovations point to a more objective and consistent assessment (20, 25).

Formative/summative assessments, feedback mechanisms, and key findings

The studies used a wide variety of assessment methods, both summative and formative, each incorporating specific feedback methods to reinforce student learning.

- Formative assessments with immediate feedback. The use of formative assessments was a practice adopted to promote real-time learning. Bord et al. (2015) and Luo et al. (2023) provided immediate feedback following OSCE stations, which resulted in significant improvements in clinical skills (18, 26). Kasai et al. (2020) integrated peer feedback during clinical role-plays, observing marked improvements in communication, medical interviewing, and professionalism (23).

- Feedback for professional development. It adopted different formats. Gran et al. (2016) emphasized the importance of timely and trust-based feedback between teachers and students in primary care settings (21). Kim et al. (2016) and Haruta et al. (2024) used structured forms (Mini-CEX and SPAT) to provide specific task-oriented feedback, strengthening clinical skills-based learning (22, 24).

- Innovative feedback approaches. Phinney et al. (2022) introduced mobile WBA assessments with real-time feedback and auto-completion (34). Costich et al. (2024) employed open and task-oriented feedback, improving the specificity and timeliness of comments, increasing student satisfaction (20).

- Summative assessments to measure comprehensive skills. Although less frequent, some tools were used for summative purposes: Rouse et al. (2024) developed an improved OSCE format with video review, which reduced grade inflation and increased assessment reliability (38). Bozzo et al. (2020) combined immediate formative feedback with structured written summative assessments by SPs, resulting in sustained improvements in clinical competence (19).

- Key findings on clinical competence and diagnostic accuracy. Studies showed that formative assessments with immediate feedback lead to consistent improvements in clinical skills. Shikino et al. (2023) used a speech recognition system to deliver precise feedback that significantly improved students' diagnostic accuracy (41). Sullivan et al. (2016) showed that instructor-led discussions improved clinical reasoning, although some errors persisted (42).

- Impact on self-confidence and skills retention. Several studies have shown that feedback also enhances student self-confidence. Luo et al. (2023) reported an increase in self-confidence after formative OSCEs, although with less impact on complex procedural skills (26). Perrig et al. (2016) showed that musculoskeletal skills improved steadily after structured interventions, while interpersonal skills showed initial improvements that declined over time (33).

- Student perceptions. Students highly valued assessments with structured and immediate feedback: Bozzo et al. (2020) and Luo et al. (2023) found that they especially appreciated OSCEs with realistic clinical scenarios and direct feedback, strengthening their confidence and competence (19,26). Kasai et al. (2020) highlighted that they recognized the value of receiving multiple perspectives, especially in peer review sessions (23). Martinsen et al. (2021) noted that, although they valued Mini-CEX sessions, they expressed a desire for more detailed suggestions (28).

- Conclusions on the development of clinical skills. Studies conclude that formative assessments with immediate feedback are key to the progressive development of skills. Bord et al. (2015) highlighted that OSCEs were useful for identifying strengths and weaknesses in emergency medicine, offering feedback (18). Kim et al. (2016) recommended implementing Mini-CEX across all internships to improve direct observation and feedback (24).

- Implications for practice. Several studies offer practical recommendations for integrating these assessment tools into medical training. Bozzo et al. (2020) suggested incorporating OSCEs with standardized patients (SPs) to provide both formative and summative feedback, enhancing clinical competence in high-demand settings (19). Furthermore, Costich et al. (2024) recommended strengthening faculty development to support work-based assessments (WBAs) in the outpatient setting, highlighting the need to train professionals in providing feedback (20). Patel et al. (2024) and Reid et al. (2021) proposed implementing more frequent feedback sessions and incorporating telemedicine training from the early stages of medical training to adapt to contemporary clinical environments (32, 36).

The summary of data extracted from technology and tools, innovative components, key findings, and practical implications of the included studies is presented in Table 3. The full version of this table can be reviewed in Tables S6 and S7 of the supplement.

Table 3. Technology and Tools, Innovative Components, Key Findings and Practical Implications of the included studies.

Author(s)	Technology & Tools	Innovative Components	Key findings	practical implications
Kasai et al. (2020)	Role-playing, peer review	Integration of multiple roles	Role-playing and peer review improve clinical performance in communication, medical interviews, physical examinations, and professionalism.	An effective educational strategy with limited resources, promoting comprehensive patient care.
Shikino et al. (2023)	MR	Automatic transcriptions	Feedback via SRS leads to greater diagnostic accuracy and better clinical performance with higher scores on Mini-CEX and SRS.	SRS-based feedback is effective and efficient, recommended for improving clinical training.
Haruta et al. (2024)	-	Development and validation of SPAT	The validity and reliability of the SPAT was confirmed, although standardization is required in the assessments of standardized patients.	Standardize the process of evaluating standardized patients, consider the selection of scenarios for high-demand exams.
Bord et al. (2015)	OSCE with simulation	Scenarios with stable and unstable phases	The OSCE effectively discriminates between high and low performing students, with a wide distribution of grades.	Recommended in emergency medicine to assess and improve clinical skills; can be adapted to other institutions.
Malone et al. (2024)	HF and VR simulations	Comparison between HF and VR	The simulations in HF were better rated and more successfully achieved than those in VR.	Caution is advised when using VR for summative assessment; additional practice in VR environments is required.
Olupeliyawa et al. (2014)	T-Mex	Structured reflection,	The use of T-Mex resulted in an improvement in collaborative skills, effective feedback, and self-assessment.	Recommended for assessing and developing teamwork skills in clinical settings.
Qureshi & Zehra (2020)	LCSAS + SPs	SP Feedback	Feedback from standardized patients is effective in improving communication skills.	SPs feedback should be integrated into the OSCE of clinical rotations for communication training.
Gran et al. (2016)	StudentPEP	Guided reflection	StudentPEP can improve mutual trust and provide timely feedback.	Recommendations for feedback mechanisms.
Bozzo et al. (2020)	SP	multi-scenario iterative design	OSCEs effectively improve competence in demanding clinical rotations. High inter-observer correlation.	Regular integration of OSCE with feedback improves clinical competence; supports the use of standardized patients for real-time formative feedback.
Rouse et al. (2024)	OSCE multi-scene	Marco RIME, video presentation	The OSCE multi-scene with the RIME framework provided a more comprehensive assessment, reduced grade inflation, and improved grade distribution.	Consider additional adjustments to time allocation and station content; ongoing evaluation is required.
Torre et al. (2021)	MSX	Structured form	The MSX is useful for evaluating and providing feedback on clinical reasoning.	More studies on the subject are needed.
Perrig et al. (2016)	Small group training	Multi-source feedback	Significant improvement in clinical and musculoskeletal skills immediately after the intervention and sustained during follow-up.	Structured feedback-oriented sessions during clinical rotations improve skill retention, particularly in practical exams such as musculoskeletal skills.
Patel et al. (2024)	Zoom, audio and video recordings	Double coaching	There was no significant improvement in anamnesis skills, but positive feedback was received from students and patients.	Incorporate more frequent and longitudinal feedback sessions; consider video review for deeper discussions.
Reid et al. (2021)	Zoom, checklist SP	Digital feedback	OSCE in telemedicine is effective for practicing with patients and managing menopause, but challenging according to students.	Consider integrating telemedicine training early and providing menopause instruction prior to the meeting.

Okubo et al. (2014)	SNAPPS, mini-CEX, 1 minute preceptor	Integration of SNAPPS and mini-CEX	Improvement in clinical reasoning; higher scores in mini-CEX, SCT and OSCE.	Recommended for use in clinical rotations; adaptable to other educational environments.
Costich et al. (2024)	Qualtrics™, modified Chen confidence scale	PDT and FORT for teachers	WBAs are feasible in the outpatient setting, with better specificity, timeliness, and satisfaction with feedback.	Further teacher development and training are needed; it is recommended to explore students' perspectives on the impact of WBAs.
Klapheke et al. (2022)	EPA/RIME Supervisory Scale	Integration of EPAs and RIME model	The EPA/RIME framework was successful with minimal additional time commitment from teachers.	Further skills development is recommended in EPAs, with more assessments and teacher instruction.
Parikh et al. (2015)	Simulation in end-of-life care	Integration of CP into surgery through SP	Positive correlation between OSCE performance and trust/communication and empathy scores.	OSCEs in end-of-life care during clinical rotations can enhance early professional skills in challenging settings.
Sullivan et al. (2016)	Virtual surgical scenarios	Sequential VSPCs	VSPCs are beneficial for clinical reasoning and decision making.	More varied cases and decision points are recommended for improved learning.
Luo et al. (2023)	-	"Feedback sandwich"	The OSCE training with immediate feedback enhances self-confidence and clinical competence, but not procedural skills.	Recommended for pre-clinical rotation training, in order to improve performance and confidence.
Kim et al. (2016)	-	Cross-cutting implementation	The mini-CEX is feasible in all clinical rotations, improves direct observation and clinical skills, with a decrease in OSCE failure rates.	Recommended for use in all clinical rotations to enhance observation, feedback, and clinical skills.
Martinsen et al. (2021)	Observational feedback	Multiple assessments with standardized feedback	The mini-CEX is feasible and valued, but no significant educational impact on scores was observed, despite some improvement in self-reported skills.	Continued use of the mini-CEX with further training of evaluators is suggested to achieve impact.
Phinney et al. (2022)	QR codes, mobile access	Two iterations of WBA, using CHAT	EPA -RIME supports formative assessment; it requires adjusting the feedback culture.	Need for a cultural change to consider WBAs as low-risk assessments; continuous support from the supervisor and greater autonomy for students in their completion.
Rogausch et al. (2015)	None	Multilevel analysis	Mini-CEX scores were more influenced by the characteristics of the context than by prior clinical skills.	Consider focusing on narrative feedback or improving the design of the Mini-CEX (and WBA) to increase the validity of the scores.
Ryan et al. (2021)	WBA system compatible with mobile devices	Mobile WBA with frequent feedback	Low reliability of WBAs for summative decisions.	More research is needed to develop reliable instruments for summative decisions; use WBAs to improve feedback.
Ryan et al. (2024)	None	WBA multi-institution	Modest reliability in the use of the O-SCORE; high variability attributed to the evaluators.	More robust training for evaluators is suggested, and possibly reducing the number of evaluators or implementing new, specific scales.

OSCE: Objective Structured Clinical Examinations; **SRS:** Student Response Systems; **EM Milestone:** Emergency Medicine Milestone; **PEP:** Peer Education Programs; **EPA:** Entrustable Professional Activities; **SPAT:** Simulated Patient Assessment Tool; **Mini -CEX:** Mini Clinical Evaluation Exercise; **RIME:** Reporter Interpreter Manager Educator; **MSX:** Multistep Exam; **WBA:** Workplace -Based Assessments; **PMEX -**: Professionalism Mini -**Evaluation** Exercise; **TMEX: -**Teamwork Mini -Clinical Evaluation Exercise **VR:** Virtual Reality (Reality Virtual); **HF:** High Fidelity Emergency Care; **SP:** Simulated Patient; **VSCP:** Virtual Surgical Patient Cases; **CHAT:** Cultural -Historical Activity Theory -; **CS:** Clinical Skill; **MSES:** Musculoskeletal Examination Skills; **PDT:** Performance -Driven Training; **FORT:** Frame -of -Reference Training for Faculty.

4. Discussion

Interpretation

This review provides an updated overview of clinical competency assessment methods during medical internships, highlighting a sustained trend toward approaches consistent with competency-based medical education (CBME), which prioritize direct observation, timely feedback, and the selective integration of digital technologies (6–8, 15). Taken together, the included studies show that clinical performance assessment should encompass both technical and transversal dimensions (communication, professionalism, clinical reasoning, teamwork), in line with frameworks widely used in medical education (3, 6–8). This shift also reflects transformations in the healthcare environment following the COVID- -19 pandemic, which accelerated adjustments in learning assessment and the use of virtual modalities (1).

Situated assessment and formative feedback as drivers of learning

The findings agree that Workplace- -Based Assessments (WBAs)—a framework that includes tools such as Mini -CEX, DOPS, and EPA-based assessments—contribute to the progressive development of competencies in real-world clinical practice by facilitating structured observation and immediate, performance-focused feedback (9, 23–24, 28, 34, 37, 40–41). This formative value is consistent with literature that positions feedback as a central element for bridging gaps between knowledge and clinical performance, fostering self-reflection, and sustaining improvement trajectories (7–9). However, the evidence suggests that feedback alone does not guarantee uniform improvements across all areas: certain complex procedural skills require repeated, deliberate practice and opportunities for authentic performance, which limits the impact of purely technological or one-off interventions (23, 26, 33). Likewise, the need to align assessment with milestone progression frameworks and to measure competencies in a valid way is linked to conceptual and operational challenges that must be considered when designing and using WBA in undergraduate studies (44).

Technological integration and telemedicine: opportunities and implementation conditions

The incorporation of simulation, virtual reality, and telemedicine expands assessment possibilities by creating safe, standardized, and replicable scenarios, and by preparing students for increasingly digitalized clinical contexts (10, 11, 27, 32, 36, 42). In curricular terms, this transformation is supported by active learning strategies with technology (45), by evidence linking simulation to relevant educational outcomes (46), and by the advancement of digital/AI tools with formative and evaluative potential (47). This momentum was reinforced after the pandemic (48) and aligns with calls to consolidate the CBME (Continuous Clinical Medicine and Evaluation) with criteria of quality and curricular coherence (49). Within this framework, virtual reality and simulations of varying fidelity offer complementary pathways whose adoption should be contextualized according to pedagogical objectives, resources, and workload (50–52). In Latin American contexts, progressive and cost-conscious implementation -is key: Mini -CEX in paper format or simple mobile apps, locally trained simulated patients, and structured feedback based on narrative guides can be prioritized; in addition, low-fidelity simulation and asynchronous virtual case studies constitute cost- -effective alternatives to more complex technologies (46, 50–52). The development of reflective thinking as a transversal competency positions feedback and guided reflection as elements that transform evaluative information into learning (53–56). Telemedicine acquires particular relevance as an area of evaluation and learning: it allows for the assessment of technology-mediated communication, reasoning with distributed information, and decision-making in remote environments (57–59). Student acceptance and willingness, as well as institutional preparation and scaling models, influence its sustainable implementation (60–61). At the undergraduate level, the planned incorporation of telemedicine training and its structured evaluation show good feasibility and perceived usefulness (62–63).

Methodological heterogeneity and quality: implications for interpreting the findings

The heterogeneity of designs, sample sizes, contexts, and outcomes hinders direct comparisons and limits robust causal inferences, which is to be expected in scope mapping and in the field of undergraduate clinical assessment (3,4,7,8). Inter-rater variability and the modest reliability observed in some tools within the WBA suite—particularly in multicenter studies and observational scales—underscore the need for evaluator training and instrument refinement (20,34,39,40). To contextualize the strength of the evidence, methodological quality was considered using the MERSQI; while the scores suggest moderate to high quality in some of the studies, limitations persist in instrument validity, statistical power, and standardization of measures, which must be considered when interpreting the data set (17, 18–43).

Implications for curriculum and institutional policy

The results reinforce the suitability of low-risk, development-centered, longitudinal assessment ecosystems that combine direct observation, narrative feedback, and strategic use of technology for clear educational purposes (7–9, 24, 34). At the institutional level, standardization of WBA practices (definitions, rubrics, expectations, minimum frequency) is recommended, along with systematic teacher training—e.g., frame -of -reference training and performance- -driven training—to improve the consistency, specificity, and timeliness of feedback (20, 39, 40, 49).

Furthermore, the integration of new assessment tools must align with public health priorities (aging, mental health) and the increasing digital literacy of the healthcare team. Incorporating frameworks such as those of the AAMC and ACGME can guide competencies in digital literacy, interprofessional collaboration, and patient-centered virtual care (48, 50–52). Given that much of the literature is based on US regulatory frameworks, it is pertinent to link these findings with global standards (e.g., WFME) to promote transferability and adaptation in Latin American settings, where institutional, cultural, and resource conditions differ (15).

Knowledge gaps and research priorities

Based on the mapping carried out, priority gaps are identified: 1) Long-term impact of formative assessments on clinical performance and patient outcomes; 2) Validity and reliability of emerging tools (including scales and digital platforms) in different contexts and languages; 3) Standardization of WBA taxonomies and frameworks (minimum requirements for observations and evaluator training) for informed educational decisions; 4) Equity and scalability of technologies (simulation and telemedicine) in resource-limited settings; and 5) Contextual evidence in Latin America and Spanish-speaking countries, where technological adoption and the organization of internships differ from the prevailing normative frameworks (9, 20, 34, 39, 40, 45–52, 57–63). Addressing these gaps requires multicenter studies with more homogeneous designs and greater power, explicit evaluator calibration strategies, and economic and implementation evaluations that inform educational policy decisions (7–9, 39, 40, 49–52).

Limitations

This review presents several limitations that affect the interpretation and scope of the findings. First, the heterogeneity of the included studies—in terms of design (RCTs, quasi- -experimental, observational, qualitative), sample sizes, contexts, and outcomes—makes direct comparisons difficult and limits the possibility of establishing robust causal relationships, something to be expected in the field of undergraduate clinical assessment and scope mapping (3, 4, 7, 8). This diversity is also reflected in the variability among evaluators and the modest reliability observed in some tools within the Workplace -Based Assessments (WBA) suite, which underscores the need for systematic teacher training and instrument refinement to improve consistency and educational use (20, 34, 39, 40). Furthermore, the mix of objective and self-reported measures (e.g., perceptions, self-confidence) and short-term outcomes adds uncertainty about the sustainability of the educational impact in the medium and long term (23, 26, 33).

Second, although MERSQI was applied to describe the methodological quality of the studies, its use in a scoping review is not intended to exclude evidence, but rather to contextualize the overall robustness of the data. Limitations persist in instrument validity, statistical power, and standardization of measures, which must be considered when interpreting the results (17, 18–43). Furthermore, the heterogeneity of conceptual frameworks and reported evaluative practices—sometimes lacking uniform implementation criteria—limits the quantitative synthesis and reinforces the descriptive nature of this review (3, 6–8).

Third, publication bias may exist, since studies with positive or significant results could be overrepresented, while interventions with no effect receive less dissemination, a phenomenon frequently described in educational research (3, 6–8). In addition, there is a possible language bias, as the search prioritized literature in English and Spanish; works in other languages may not have been captured, affecting the comprehensiveness of the mapping (16).

Fourth, the geographical predominance of evidence in developed countries (mainly the United States, Japan, and Switzerland) limits the transferability of certain findings. In particular, the applicability of high-fidelity technologies or virtual reality may be reduced in contexts with fewer resources, as is the case in many Latin American medical schools, where infrastructure, technical support, and the availability of teaching time are more restricted (46, 50–52). This situation reinforces the need to promote research in Spanish-speaking and Latin American contexts, considering different institutional, cultural, and resource realities, to facilitate a contextual adaptation of competency-based assessment models (15).

Fifth, regarding technological integration, studies show the benefits of simulation and VR in learning and assessment; however, implementation barriers, access gaps, and costs persist, limiting their scalability, especially outside of high-complexity centers (46, 50–52). Similarly, telemedicine is emerging as an indispensable area for clinical training and assessment, but its effective adoption depends on institutional preparedness, student acceptance, and curriculum planning—factors that vary widely across settings and can affect equity in its implementation (57–63). These asymmetries imply that some of the positive effects observed in more technologically prepared environments cannot be directly extrapolated to Latin American institutions without phased strategies and cost--effective designs (50–52, 60–61).

Sixth, although the adoption of frameworks such as AAMC/ACGME can guide curricular alignment toward competencies in digital literacy, interprofessional collaboration, and patient-centered virtual care, much of the literature analyzed is based on US regulations, which necessitates linking these findings with global standards (e.g., WFME) to improve their transferability and relevance in Latin America (15, 48–52). In the absence of comparable local studies and cross-cultural validations, adoption processes should be accompanied by implementation evaluations and economic analyses that consider the reality of each academic -healthcare system (49–52, 60–61).

Finally, methodological decisions inherent to the design—time window (last 10 years), set of databases consulted, and inclusion/exclusion criteria—may have left out relevant earlier literature or literature from non-indexed sources, which constitutes another limitation inherent to this type of mapping (16). Taken together, these limitations indicate that the results should be interpreted primarily as a characterization of emerging practices and trends, and not as definitive evidence of comparative effectiveness; their main value lies in guiding curricular decisions and research priorities under a contextualized reading of methodological quality and the feasibility of adoption in different scenarios (7–9,17,46,50–52,57–63).

5. Conclusions

- A review of 26 studies and over 5,700 students confirms a global shift toward competency-based medical education, where tools such as standardized assessment methods like OSCE and tools grouped under the WBA framework—including Mini -CEX, DOPS, and EPA-based assessments—are becoming the current standard for evaluating real-world clinical performance. Furthermore, there is a growing integration of technologies such as virtual reality, telemedicine, and speech recognition systems to standardize and enhance these processes.
- The fundamental value of these methodologies lies in their ability to provide immediate and structured feedback, which positively impacts the intern's self-confidence and diagnostic accuracy. However, the effective implementation of these innovations is not automatic; it requires overcoming significant gaps in teacher training and ensuring the sustainability of institutional technological resources.
- In summary, for these innovations to be transformative, faculties must adequately balance formative and summative processes, and incorporate technology as a facilitator of reflective clinical learning.

Supplementary material: Annex 1. Search strategy for each database; Items extracted; Characteristics of the studies; Quality assessment; Results and effect measurement; Area of Competence, Evaluation Method, Technology and Tools, Innovative Components, Key Findings, Conclusions, Practical implications; Type of feedback, Feedback Mechanism, Students' perception.
Funding: There has been no funding.
Declaration of conflict of interest: The authors declare that they have no conflict of interest.
Authors' contributions: Conceptualization: NLT, SJV, AHA, Data procurement: NLT, Methodology/formal analysis/validation: NLT, Project management: NLT, AHA, Writing - original draft: NLT, JSG, SJV, Writing - revision and editing: NLT, JSG, AHA, OJY.

6. References.

1. Hauer K, Lockspeiser T, Chen H. The COVID -19 pandemic as an imperative to advance medical student assessment: three areas for change. *Academic Med.* **2021,** 96, 182–185. https://doi.org/10.1097/ACM.0000000000003764
2. Ten C, Regehr G. The power of subjectivity in the assessment of medical trainees. *Academic Med.* **2019,** 94, 333–337. https://doi.org/10.1097/ACM.0000000000002495
3. Whitehead C, Kuper A, Hodges B, Ellaway R. Conceptual and practical challenges in the assessment of physician competencies. *Med. Teach.* **2015,** 37, 245–251. https://doi.org/10.3109/0142159X.2014.993599
4. Witheridge A, Ferns G, Scott-Smith W. Revisiting Miller's pyramid in medical education: the gap between traditional assessment and diagnostic reasoning. *Int. J. Med. Educ.* **2019,** 10, 191–192. https://doi.org/10.5116/ijme.5d9b.0c37
5. Kang D, Siddiqui S, Weiss H, Sifri Z, Krishnaswami S, Nwomeh B, Price R, Tarpley J, Finlayson S, Swaroop M. Are we meeting ACGME core competencies? A systematic review of literature on international surgical rotations. *Am. J. Surg.* **2018,** 216, 782–786. https://doi.org/10.1016/j.amjsurg.2018.07.048
6. Ginsburg S, McIlroy J, Oulanova O, Eva K, Regehr G. Toward authentic clinical evaluation: pitfalls in the pursuit of competition. *Academic Med.* **2010,** 85, 780–786. https://doi.org/10.1097/ACM.0b013e3181d73fb6

7. Lockyer J, Carraccio C, Chan M, Hart D, Smee S, Touchie C, Holmboe E, Frank J, on behalf of the ICBME Collaborators. Core principles of assessment in competence-based medical education. *Med. Teach.* **2017,** 39, 609–616. https://doi.org/10.1080/0142159X.2017.1315082
8. Harris P, Bhanji F, Topps M, Ross S, Lieberman S, Frank J, Snell L, Sherbino J, on behalf of the ICBME Collaborators. Evolving concepts of assessment in a competition-based world. *Med. Teach.* **2017,** 39, 603–608. https://doi.org/10.1080/0142159X.2017.1315071
9. Lörwald A, Lahner F, Nouns Z, Berendonk C, Norcini J, Greif R, Huwendiek S. The educational impact of Mini -Clinical Evaluation Exercise (Mini -CEX) and Direct Observation of Procedural Skills (DOPS) and its association with implementation: a systematic review and meta --analysis. *PLOS ONE* **2018,** 13, e0198009. https://doi.org/10.1371/journal.pone.0198009
10. McGrath J, Taekman J, Dev P, Danforth D, Mohan D, Kman N, Crichlow A, Bond W. Using virtual reality simulation environments to assess competence for emergency medicine learners. *Academic Emerg. Med.* **2018,** 25, 186–195. https://doi.org/10.1111/acem.13308
11. Park J, Kwon H, Chung C. Innovative digital tools for new trends in teaching and assessment methods in medical and dental education. *J. Educ. Eval. Health Prof.* **2021,** 18, 13. https://doi.org/10.3352/jeehp.2021.18.13
12. Zackoff M, Real F, Cruse B, Davis D, Klein M. Medical student perspectives on the use of immersive virtual reality for clinical assessment training. *Academic Pediatrician* **2019,** 19, 849–851. https://doi.org/10.1016/j.acap.2019.06.008
13. O'Leary M, Scully D, Karakolidis A, Pitsia V. The state -of -the -art in digital technology -based assessment. *Eur. J. Educ.* **2018,** 53, 160–175. https://doi.org/10.1111/ejed.12271
14. Kassutto S, Baston C, Clancy C. Virtual, augmented, and alternate reality in medical education: socially distanced but fully immersed. *Sch.* **2021,** 2, 651–664. https://doi.org/10.34197/ats-scholar.2021-0002RE
15. World FFME. Basic medical education: WFME global standards for quality improvement. The 2020 review. *WFME* **2020,** https://wfme.org/standards/bme/
16. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst. Rev.* **2016,** 5, 210. https://doi.org/10.1186/s13643-016-0384-4
17. Al A, Haque M, Parle J. A modified medical education research study quality instrument (MMERSQI) developed by Delphi consensus. *BMC Med. Educ.* **2023,** 23, 63. https://doi.org/10.1186/s12909-023-04033-6
18. Bord S, Retezar R, McCann P, Jung J. Development of an objective structured clinical examination for assessment of clinical skills in an emergency medicine clerkship. *West. J. Emerg. Med.* **2015,** 16, 866–870. https://doi.org/10.5811/westjem.2015.9.27307
19. Bozzo N, Arancibia S, Contreras R, Pérez G. Description and analysis of OSCE with simulated patients in Internal Medicine internship 2016–2017 at the Faculty of Medicine, University of Chile. *Rev. Méd. Chile* **2020,** 148, 810–817. https://doi.org/10.4067/S0034-98872020000600810
20. Costich M, Friedman S, Robinson V, Catallozzi M. Implementation and faculty perception of outpatient medical student workplace-based assessments. *Clin. Teach.* **2024,** 21, e13751. https://doi.org/10.1111/tct.13751
21. Gran S, Brænd A, Lindbæk M, Frich J. General practitioners' and students' experiences with feedback during a six -week clerkship in general practice: a qualitative study. *Scand. J. Prim. Health Care* **2016,** 34, 172–179. https://doi.org/10.3109/02813432.2016.1160633
22. Haruta J, Nakajima R, Monkawa T. Development of a validated assessment tool for medical students using simulated patients: an 8 -year panel survey. *BMC Med. Educ.* **2024,** 24, 399. https://doi.org/10.1186/s12909-024-05386-2
23. Kasai H, Ito S, Tajima H, Takahashi Y, Sakurai Y, Kawata N, Sugiyama H, Asahina M, Sakai I, Tatsumi K. The positive effect of student -oriented clinical clerkship rounds employing

role -play and peer review on the clinical performance and professionalism of clerkship students. *Med. Teach.* **2020,** 42, 73–78. https://doi.org/10.1080/0142159X.2019.1656330
24. Kim S, Willett L, Noveck H, Patel M, Walker J, Terregino C. Implementation of a Mini -CEX requirement across all third -year clerkships. *Teach. Learn. Med.* **2016,** 28, 424–431. https://doi.org/10.1080/10401334.2016.1165682
25. Klapheke M, Abrams M, Cubero M, Zhu X. Aligning medical student workplace -based assessments with entrustable professional activities and the RIME model in a psychiatry clerkship. *Academic Psychiatry* **2022,** 46, 283–288. https://doi.org/10.1007/s40596-022-01614-3
26. Luo P, Shen J, Yu T, Zhang X, Zheng B, Yang J. Formative objective structured clinical examination with immediate feedback improves surgical clerks' self- -confidence and clinical competence. *Med. Teach.* **2023,** 45, 212–218. https://doi.org/10.1080/0142159X.2022.2126755
27. Malone M, Way D, Leung C, Danforth D, Maicher K, Vakil J, Kman N, San M. Evaluation of high -fidelity and virtual reality simulation platforms for assessing fourth -year medical students' encounters with patients in need of urgent or emergent care. *Ann. Med.* **2024,** 56, 2382947. https://doi.org/10.1080/07853890.2024.2382947
28. Martinsen S, Espeland T, Berg E, Samstad E, Lillebo B, Slørdahl T. Examining the educational impact of the Mini -CEX: a randomized controlled study. *BMC Med. Educ.* **2021,** 21, 228. https://doi.org/10.1186/s12909-021-02670-3
29. Okubo Y, Nomura K, Saito H, Saito N, Yoshioka T. Reflection and feedback in ambulatory education. *Clin. Teach.* **2014,** 11, 355–360. https://doi.org/10.1111/tct.12164
30. Olupeliyawa A, Balasooriya C, Hughes C, O'Sullivan A. Educational impact of an assessment of medical students' collaboration in health care teams. *Med. Educ.* **2014,** 48, 146–156. https://doi.org/10.1111/medu.12318
31. Parikh P, Brown R, White M, Markert R, Eustace R, Tchorz K. Simulation -based end -of -life care training during surgical clerkship: assessment of skills and perceptions. *J. Surg. Beef.* **2015,** 196, 258–263. https://doi.org/10.1016/j.jss.2015.03.019
32. Patel K, Anderson R, Becker C, Taylor W, Liu A, Varshney A, Ali N, Nath B, Pelletier S, Shields H, Osman N. Dual coaching of medical clerkship students' history -taking skills by volunteer in patients at the bedside and faculty physicians on Zoom during the COVID -19 pandemic. *Adv. Med. Educ. Pract.* **2024,** 15, 923–933. https://doi.org/10.2147/AMEP.S472324
33. Perrig M, Berendonk C, Rogausch A, Beyeler C. Sustained impact of a short small -group course with systematic feedback in addition to regular clinical clerkship activities on musculoskeletal examination skills: a controlled study. *BMC Med. Educ.* **2016,** 16, 35. https://doi.org/10.1186/s12909-016-0554-z
34. Phinney L, Fluet A, O'Brien B, Seligman L, Hauer K. Beyond checking boxes: exploring tensions with use of a workplace -based assessment tool for formative assessment in clerkships. *Academic Med.* **2022,** 97, 1511–1520. https://doi.org/10.1097/ACM.0000000000004774
35. Qureshi A, Zehra T. Simulated patient's feedback to improve communication skills of clerkship students. *BMC Med. Educ.* **2020,** 20, 15. https://doi.org/10.1186/s12909-019-1914-2
36. Reid H, Branford K, Reynolds T, Baldwin M, Dotters-Katz S. It's getting hot in here: piloting a telemedicine OSCE addressing menopause concerns for obstetrics and gynecology clerkship students. *MedEdPORTAL* **2021,** 11146. https://doi.org/10.15766/mep_2374-8265.11146
37. Rogausch A, Beyeler C, Montagne S, Jucker-Kupper P, Berendonk C, Huwendiek S, Gemperli A, Himmel W. The influence of students' prior clinical skills and context characteristics on Mini -CEX scores in clerkships: a multilevel analysis. *BMC Med. Educ.* **2015,** 15, 208. https://doi.org/10.1186/s12909-015-0490-3

38. Rouse M, Newman J, Waller C, Fink JRIME and reason: multi -station OSCE enhancement to neutralize degree inflation. *Online Educ. Med.* **2024,** 29, 2339040. https://doi.org/10.1080/10872981.2024.2339040
39. Ryan M, Richards A, Perera R, Park Y, Stringer J, Waterhouse E, Dubinsky B, Khamishon R, Santen S. Generalizability of the Ottawa Surgical Competency Operating Room Evaluation (O -SCORE) scale to assess medical student performance on core EPAs in the workplace: findings from one institution. *Academic Med.* **2021,** 96, 1197–1204. https://doi.org/10.1097/ACM.0000000000003921
40. Ryan M, Gielissen K, Shin D, Perera R, Gusic M, Ferenchick G, Ownby A, Cutrer W, Obeso V, Santen S. How well do workplace -based assessments support summative entrustment decisions? A multi -institutional generalisability study. *Med. Educ.* **2024,** 58, 825–837. https://doi.org/10.1111/medu.15291
41. Shikino K, Tsukamoto T, Noda K, Ohira Y, Yokokawa D, Hirose Y, Sato E, Mito T, Ota T, Katsuyama Y, Uehara T, Ikusaka M. Do clinical interview transcripts generated by speech recognition software improve clinical reasoning performance in mock patient encounters? A prospective observational study. *BMC Med. Educ.* **2023,** 23, 272. https://doi.org/10.1186/s12909-023-04246-9
42. Sullivan S, Bingman E, O'Rourke A, Pugh C. Piloting virtual surgical patient cases with third -year medical students during the surgery rotation. *Am. J. Surg.* **2016,** 211, 689–696.e1. https://doi.org/10.1016/j.amjsurg.2015.11.021
43. Torre D, Hemmer P, Durning S, Dong T, Swygert K, Schreiber -Gregory D, Kelly W, Pangaro L. Gathering validity evidence on an internal medicine clerkship multistep exam to assess medical student analytic ability. *Teach. Learn. Med.* **2021,** 33, 28–35. https://doi.org/10.1080/10401334.2020.1749635
44. Natesan P, Batley N, Bakhti R, El -Doueihi P. Challenges in measuring ACGME competencies: considerations for milestones. *Int. J. Emerg. Med.* **2018,** 11, 39. https://doi.org/10.1186/s12245-018-0198-3
45. McCoy L, Pettit R, Lewis J, Bennett T, Carrasco N, Brysacz S, et al. Developing technology -enhanced active learning for medical education: challenges, solutions, and future directions. *J. Osteopath. Med.* **2015,** 115, 202–211. https://doi.org/10.7556/jaoa.2015.042
46. Brydges R, Hatala R, Zendejas B, Erwin P, Cook D. Linking simulation -based educational assessments and patient -related outcomes: a systematic review and meta -analysis. *Academic Med.* **2015,** 90, 246–256. https://doi.org/10.1097/ACM.0000000000000549
47. Sapci A, Sapci H. Artificial intelligence education and tools for medical and health informatics students: systematic review. *JMIR Med. Educ.* **2020,** 6, e19285. https://doi.org/10.2196/19285
48. Goh P. The vision of transformation in medical education after the COVID -19 pandemic. *Korean J. Med. Educ.* **2021,** 33, 171–174. https://doi.org/10.3946/kjme.2021.197
49. Holmboe E, Sherbino J, Englander R, Snell L, Frank J, on behalf of the ICBME Collaborators. A call to action: the controversy of and rationale for competence -based medical education. *Med. Teach.* **2017,** 39, 574–581. https://doi.org/10.1080/0142159X.2017.1315067
50. Pottle J. Virtual reality and the transformation of medical education. *Future Healthc. J.* **2019,** 6, 181–185. https://doi.org/10.7861/fhj.2019-0036
51. Scott A, Gartner A. Low fidelity simulation in a high fidelity world. *Postgrad. Med. J.* **2019,** 95, 687–688. https://doi.org/10.1136/postgradmedj-2019-FPM.9
1. Tainter C, Wong N, Bittner E. Innovative strategies in critical care education. *J. Crit. care* **2015,** 30, 550–556. https://doi.org/10.1016/j.jcrc.2015.02.001
53. Fragkos K. Reflective practice in healthcare education: an umbrella review. *Educ. Sci.* **2016,** 6, 27. https://doi.org/10.3390/educsci6030027

54. Jorwekar G. Reflective practice as a method of learning in medical education: history and review of literature. *Int. J. Res. Med. Sci.* **2017,** 5, 1188. https://doi.org/10.18203/2320-6012.ijrms20171223
55. MacAskill W, Chua W, Woodall H, Pinidiyapathirage J. Beyond the written reflection: a systematic review and qualitative synthesis of creative approaches to reflective learning among medical students. *Perspect. Med. Educ.* **2023,** 12, 361–371. https://doi.org/10.5334/pme.914
56. Leung K, Peisah C. A mixed -methods systematic review of group reflective practice in medical students. *Healthcare* **2023,** 11, 1798. https://doi.org/10.3390/healthcare11121798
57. Eljack M, Elhadi Y, Mahgoub E, Ahmed K, Mohamed M, Elnaiem W, et al. Physician experiences with teleconsultations amidst conflict in Sudan. *Sci. Rep.* **2023,** 13, 22688. https://doi.org/10.1038/s41598-023-49967-5
58. Radfar A, Chevalier C, Rouse N, Patriche D, Filip I. Telemedicine: new horizons in healthcare. *J. Hosp. Adm.* **2017,** 6, 40. https://doi.org/10.5430/jha.v6n2p40
59. Kuo R, Delvecchio F, Babayan R, Preminger G. Telemedicine: recent developments and future applications. *J. Endourol.* **2001,** 15, 63–66. https://doi.org/10.1089/08927790150500971
60. Malhotra P, Ramachandran A, Chauhan R, Soni D, Garg N. Assessment of knowledge, perception, and willingness of using telemedicine among medical and allied healthcare students studying in private institutions. *Telehealth Med. Today* **2020,** https://doi.org/10.30953/tmt.v5.228
61. Otto L, Schlieter H, Harst L, Whitehouse D, Maeder A. The telemedicine community readiness model—successful telemedicine implementation and scale -up. *Front. Digit. Health* **2023,** 5, 1057347. https://doi.org/10.3389/fdgth.2023.1057347
62. Waseh S, Dicker A. Telemedicine training in undergraduate medical education: mixed -methods review. *JMIR Med. Educ.* **2019,** 5, e12515. https://doi.org/10.2196/12515
63. Hindman D, Kochis S, Apfel A, Prudent J, Kumra T, Golden W, et al. Improving medical students' OSCE performance in telehealth: the effects of a telephone medicine curriculum. *Academic Med.* **2020,** 95, 1908–1912. https://doi.org/10.1097/ACM.0000000000003622