Revista Española
de Educación Médica

Supplementary Tables.

Mapping Clinical Competence Assessment Methods for Medical Clerkships: A Scoping Review of Current Practices and Emerging Trends.

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

[1] Facultad de Medicina, Universidad de Chile, Chile, nicolaslavados@ug.uchile.cl, https://orcid.org/0009-0005-4223-3288.
[2] Facultad de Medicina, Universidad de Chile, Chile, joaquin.salgado.g@ug.uchile.cl, https://orcid.org/0009-0005-3307-1440.
[3] Facultad de Medicina, Universidad de Chile, Chile, silvanajimenez@ug.uchile.cl, https://orcid.org/0000-0003-3730-7447.
[4] Facultad de Medicina, Universidad San Sebastián, Chile, levarito@uchile.cl, https://orcid.org/0009-0007-4861-2144.
[5] Departamento de Educación en Ciencias de la Salud, Facultad de Medicina, Universidad de Chile, Chile, ojerez@uchile.cl, https://orcid.org/0000-0003-0869-5938.
* Correspondencia: nicolaslavados@ug.uchile.cl

Table S1. Search Strategy

Database	Terms with a publication date limit October 27, 2024	Combination
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

Table S2: Data Items

Category	Details
Primary Outcomes	
Assessment Methods	Specific methods used to assess clinical competence, such as OSCEs, Mini-CEX, DOPS, WBAs, and simulation-based assessments.
Competency Areas	Areas of competence being assessed, including clinical reasoning, communication skills, professionalism, teamwork, and technical skills.
Feedback Mechanisms	Types of feedback provided, whether formative or summative, and the mechanisms through which feedback was delivered.
Technological Tools	Use of technological tools in assessments, such as digital platforms, virtual reality, high-fidelity simulators, and telemedicine.
Additional Variables	
Authors and Year	The authors of the study and the year of publication.
Setting	The clinical setting or clerkship where the study was conducted.
Sample Size and Participant Level	The number of participants and their level of medical education.
Innovative Components	Any unique or innovative aspects of the assessment methods used.
Other Variables	
Participant Characteristics	Details about the participants, such as their year in medical school and the specific clerkship they were involved in.
Student Perceptions	Qualitative and quantitative data about students' perceptions.
Implications for Practice	Whether recommendations for medical training or further research are needed.

OSCE: Objective Structured Clinical Examinations; MINI CEX: Mini Clinical Evaluation Exercise; DOBS: Direct Observation of Procedural Skills; WBAs: Workplace-Based Assessments

Tabla S3. Summary of characteristics of included studies

Author(s)	Year	Country	Study Design	Sample Size	Participant Level	Setting
Kasai et al.	2020	Japan	Prospective cohort study	75	Clerkship students	Clinical Clerkship
Shikino et al.	2023	Japan	Prospective observational	79	5th-year medical students	Clinical Clerkship
Haruta et al.	2024	Japan	Longitudinal survey	831	Medical students	Clinical Practice Medical Interview Sessions
Bord et al.	2015	USA	Cross-sectional study	80	2nd to 4th-year medical students	Emergency Medicine Clerkship
Malone et al.	2024	USA	Nonrandomised, 2 groups	216	4th-year medical students	Emergency Medicine Clerkship
Olupeliyawa et al.	2014	Australia	Cross-sectional study	25	Final-year medical students	Final Clinical Rotation
Qureshi & Zehra	2020	Pakistan	Randomized controlled trial (RCT)	80	Final-year medical students	Final-year clerkship
Gran et al.	2016	Norway	Qualitative study	30	5th-year medical students	General Practice Clerkship
Bozzo et al.	2020	Chile	Longitudinal study	462	6th-year medical students	Internal Medicine Clerkship
Rouse et al.	2024	USA	Pre- and post-intervention study	200	3rd-year medical students	Internal Medicine Clerkship
Torre et al.	2021	USA	Validation study	477	3rd-year medical students	Internal Medicine Clerkship
Perrig et al.	2016	Switzerland	Pre- and post-intervention study	48	4th-year medical students	Internal Medicine Clerkship, University of Berne
Patel et al.	2024	USA	Pilot intervention study	13	2nd-year medical students	Medical Clerkship
Reid et al.	2021	USA	Pilot study	32	3rd-year medical students	OB/GYN Clerkship
Okubo et al.	2014	Japan	Prospective cohort study	68	5th and 6th-year medical students	Out-patient care unit, Tokyo Women's Medical University
Costich et al.	2024	USA	Pilot study	131	Clerkship students	Pediatric Primary Care Setting
Klaphekee et al.	2022	USA	Pilot study	109	3rd-year medical students	Psychiatry Clerkship
Parikh et al.	2015	USA	Cross-sectional study	389	3rd-year medical students	Surgery Clerkship
Sullivan et al.	2016	USA	Mixed-methods study	98	3rd-year medical students	Surgery Clerkship
Luo et al.	2023	China	Randomized controlled	38	4th-year medical students	Surgical Clerkship

			study (RCT)			
Kim et al.	2016	USA	Pre- and post-intervention study	110-123	3rd-year medical students	Various Clerkships
Martinsen et al.	2021	Norway	Randomized controlled trial (RCT)	38	5th-year medical students	Various Clerkships
Phinney et al.	2022	USA	Qualitative study	35	3rd- and 4th-year medical students	Various Clerkships
Rogausch et al.	2015	Switzerland	Multilevel analysis	165	4th-year medical students	Various Clerkships
Ryan et al.	2021	USA	Cross-sectional study	220	3rd-year medical students	Various Clerkships
Ryan et al.	2024	USA	Cross-sectional study	1810	Clerkship-level medical students	Various Institutions

Table S4. Assessment of the quality of included articles

Study	Study design	Sampling	Type of data	Validity of evaluation instrument	Data analysis	Outcome	Total score
Bord et al. (2015)	3	3	3	3	3	3	18
Bozzo et al. (2020)	1	1	3	3	1	1,5	10,5
Costich et al. (2024)	2	3	3	3	3	3	17
Gran et al. (2016)	1	1	3	3	1	1,5	10,5
Haruta et al. (2024)	1,5	1	3	3	2	1,5	12
Kasai et al. (2020)	1,5	1	3	3	2	1,5	12
Kim et al. (2016)	1,5	2	3	3	2	1,5	13
Klapheke et al. (2022)	1	1	3	3	1	1,5	10,5
Luo et al. (2023)	1,5	1	3	3	2	1,5	12
Malone et al. (2024)	1,5	1	3	3	2	1,5	12
Martinsen et al. (2021)	2	3	3	3	3	3	17
Okubo et al. (2014)	3	3	3	3	3	3	18
Olupeliyawa et al. (2014)	1,5	1	3	3	2	1,5	12
Parikh et al. (2015)	1	1	3	3	1	1,5	10,5
Patel et al. (2024)	1,5	1	3	3	2	1,5	12
Perrig et al. (2016)	2	1	3	3	2	1,5	12,5
Phinney et al. (2022)	1	1	3	3	1	1,5	10,5
Qureshi & Zehra (2020)	2	1	3	3	2	1,5	12,5
Reid et al. (2021)	1,5	1	3	3	2	1,5	12
Rogausch et al. (2015)	1,5	1	3	3	2	1,5	12
Rouse et al. (2024)	3	3	3	3	3	3	18
Ryan et al. (2024)	2	3	3	3	3	3	17
Ryan et al. (2021)	1	1	3	3	1	1,5	10,5
Shikino et al. (2023)	1,5	1	3	3	2	1,5	12
Sullivan et al. (2016)	1,5	1	3	3	2	1,5	12
Torre et al. (2021)	1,5	1	3	3	2	1,5	12

Tabla S5. Results of Individual Studies in Terms of Outcomes and Their Respective Effect Measures

Study	Year	Outcome	Effect Measure(s)
Bord et al.	2015	Development of an OSCE for assessment of clinical skills in EM clerkship	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 during internal medicine clerkship	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 ($p \leq 0.05$)
Costich et al.	2024	Implementation and faculty perception of EPA-grounded WBAs in an outpatient primary care setting during a pediatrics clerkship	Faculty feedback: Increased delivery of specific, task-oriented feedback ($p = 0.006$), greater satisfaction with opportunities to provide feedback (not significant), more feedback within 24 hours of an encounter (not significant), Themes from interviews: Benefits of WBAs, barriers to feedback provision, suggestions for improvement
Gran et al.	2016	Feedback experiences during clerkship	Thematic analysis for identifying key themes and feedback experiences, consistency in themes across interviews.
Haruta et al.	2024	Validation of the Simulated Patient Assessment Tool (SPAT) for assessing medical students' clinical performance using simulated patients	Exploratory factor analysis: Revealed two factors, communication and physician performance, with a cumulative contribution rate of 60.47%, Internal consistency: Overall Cronbach's alpha coefficient was 0.929, indicating high internal consistency, One-way ANOVA: Significant differences in SPAT total scores among SPs ($F(34, 760) = 16.79$, $p < 0.001$) and scenarios ($F(20, 774) = 11.39$, $p < 0.001$), Convergent validity: Moderate correlation ($r = 0.212$, $p < 0.05$) between SPAT and post-CC OSCE total scores
Kasai et al.	2020	Improvement in clinical performance and professionalism of clerkship students	Mini-CEX scores: Significant improvements in medical interviewing (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 scores: Significant improvements in doctor–patient relationship skills (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 a mini-CEX requirement across all third-year clerkships and its impact on direct observation and clinical skills	Adherence: 92% of required mini-CEX forms were completed, with 78% indicating specific feedback was given, Student report: Significant increases in student report of direct observation of physical examination in all clerkships (e.g., surgery: 49% to 87%, $p < 0.0001$), AAMC GQ: Increased rates of faculty observation of history taking and physical examination in all clerkships post-intervention, OSCE performance: Decrease in summative OSCE failure rates from 12% pre-intervention to 2% post-intervention ($p = 0.0046$), p-values: $p < 0.0001$ for surgery physical examination observation, $p = 0.0046$ for OSCE failure rates
Klapheke et al.	2022	Pilot study of workplace-based assessments using EPAs and the RIME model in a psychiatry clerkship	Mean scores on EPAs and RIME: Students' mean skill profile exceeded 4.0 on EPA1, EPA6, and reporter of RIME, Student feedback: 44% found the feedback helpful, 32.2% felt evaluations were fair, Faculty feedback: Mean score for ease of completing EPA ratings was 3.89 (SD = 1.27), and for sufficient explanation on completing EPA ratings was 4.33 (SD = 0.5)
Luo et al.	2023	Improvement in surgical clerks' self-confidence and clinical competence	Self-confidence assessment (SCA): Significant improvements in all stations (history taking: pre: 2.68 ± 0.82, post: 3.47 ± 0.84, $p < 0.01$; physical examination: pre: 2.47 ± 0.84, post: 3.42 ± 0.77, $p < 0.01$; CPR, clinical reasoning, dressing change of the wound, surgical asepsis). Mini-CEX: Significant improvement (OSCE group: 6.59 ± 0.62, control group: 5.31 ± 1.09, $p < 0.01$). DOPS: No significant improvement (OSCE group: 5.15 ± 0.58, control group: 4.96 ± 0.63, $p > 0.05$)
Malone et al.	2024	Urgent and emergent care skills	Wilcoxon signed-rank test ($T = 337.5$, $p = 0.001$), McNemar test ($p \leq 0.001$), thematic analysis for qualitative data, significance determined by $p \leq 0.05$.
Martinsen et al.	2021	Clinical skills measured by mini-CEX assessments	ANOVA ($F = 3.603$, $p = 0.066$), ANCOVA ($F = 1.884$, $p = 0.179$), mean scores (3.5–3.6 out of 4), standard deviations (0.55–0.63).
Okubo et al.	2014	Evaluation of a clinical clerkship program in an outpatient care setting to improve clinical reasoning abilities of medical students	Mini-CEX scores: Higher in all areas for students in the program, SCT scores: Significantly higher post-course (pre-course: 77.4 ± 10.5; post-course: 86.8 ± 10.2), OSCE scores: Higher for students in the intervention group compared to the control group
Olupeliyawa et al.	2014	Assessment of the educational impact of the T-MEX on medical students' collaboration in health care teams	Generalisability coefficient: 0.62 with three forms, 0.80 predicted with eight forms, Content analysis of feedback and reflections, Thematic analysis of focus group and interview data
Parikh et al.	2015	Communication, empathy, and trust in end-of-life care	Descriptive statistics for mean scores (mean = 89.0%, SD = 6.7%), standard deviations, correlation with trust 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 to 0.83), p-values ($p \leq 0.05$)
Phinney et al.	2022	Feedback and self-reflection in clinical settings	Thematic analysis for identifying key themes and tensions, consistency in themes across interviews.
Qureshi & Zehra	2020	Communication skills using simulated patient feedback	Pearson correlation coefficient ($r = 0.83$) for internal consistency of LCSAS (Cronbach's $\alpha = 0.83$), pre/post-test scores improvement (mean difference = 1.5, $p \leq 0.05$).
Rogausch et al.	2015	Analysis of the influence of students' prior clinical skills and context characteristics on mini-CEX scores in clerkships	Regression analysis: Trainers' clinical position was the most influential predictor of mini-CEX scores (regression coefficient = 0.55, $p < 0.001$ for residents compared to heads of department), Task complexity and clinic size: Significant predictors of mini-CEX scores, OSCE performance: Not a significant predictor of mini-CEX scores, Correlation: Weak correlation between OSCE scores and mini-CEX scores ($r = 0.26$ for overall mini-CEX scores, $r = 0.27$ for domain mini-CEX scores), p-values: $p < 0.001$ for trainers' clinical position, $p < 0.05$ for task complexity and clinic size
Rouse et al.	2024	Enhanced OSCE to neutralize grade inflation and provide a more comprehensive assessment of clinical skills	Comparison of pre- and post-intervention OSCE scores (pre: M = 94.25%, SD = 5.65%, post: M = 81.00%, SD = 6.88%), student surveys (average ratings: 4.4 for Reporter, 4.4 for Interpreter, 4.2 for Manager, 3.5 for time allotment, 3.5 for difficulty)
Ryan et al.	2021	Core EPAs measured by O-SCORE scale	Generalizability theory for assessing reliability (Phi coefficient = 0.19 to 0.44), mean scores (3.48–3.62), variance components (student = 3.5% to 8%, rater = 29.6% to 50.3%).
Ryan et al.	2024	Evaluation of the reliability of WBAs for summative entrustment decisions in medical education	Phi coefficient: >0.7 threshold for acceptable reliability, Variance attributed to the learner: <10% for most analyses, Number of observations required for reasonable reliability: Range = 3 to >560, Median = 60
Shikino et al.	2023	Evaluation of the effectiveness of SRS feedback on clinical reasoning performance in medical students during mock patient encounters	Diagnostic accuracy: SRS feedback group showed higher diagnostic accuracy rates (pre-test: 51.3%, post-test: 89.7%) compared to the IC recorder feedback group (pre-test: 57.5%, post-test: 67.5%) ($p = 0.037$), Mini-CEX scores: Significant improvements in medical interviewing, physical examination, professionalism, organization/efficiency, and overall clinical competence in the SRS feedback group ($p < 0.001$), Checklist scores: Higher total checklist scores in the SRS feedback group (pre-test: 12.2, post-test: 16.1) compared to the IC recorder feedback group (pre-test: 13.1, post-test: 13.8) ($p < 0.001$), Feedback time: SRS feedback group had shorter feedback time (22.6 ± 2.1 min) compared to the IC recorder feedback group (27.7 ± 2.1 min) ($p = 0.04$)
Sullivan et al.	2016	Improvement in clinical decision-making skills	Repeated measures ANOVA (acute diverticulitis case: Pillai's Trace = 0.807, $F(2,36) = 75.279$, $p < 0.000$; GI bleeding case: Pillai's Trace = 0.822, $F(2,19) = 43.941$, $p < 0.000$), thematic analysis, p-values ($p \leq 0.05$)
Torre et al.	2021	Validation of the Multistep Exam (MSX) for assessing medical student analytic clinical reasoning abilities in an internal medicine clerkship	Correlation: MSX score had a significant positive correlation with Step 2 CS ICE score ($r = 0.26$, $p < 0.01$), Multiple linear regression: MSX score was a significant predictor of Step 2 CS ICE score ($\beta = 0.19$, $p < 0.001$), explaining an additional 4% of the variance beyond NBME Medicine subject score and Medicine OSCE score, Reliability: Cronbach alpha for MSX ranged from 0.70 to 0.80
Reid et al.	2021	Evaluation of a telemedicine OSCE for managing menopausal concerns in OB/GYN clerkship students	Post-encounter note scores (median score: 20 out of 45), student surveys (78% expressed discomfort with telehealth, 66% rated the educational value as excellent or above average), differential diagnosis accuracy (100% identified menopause/perimenopause, 84% identified hyperthyroidism)
Patel et al.	2024	Evaluation of dual coaching on history-taking skills of medical clerkship students	Minicard scores (no significant improvement in history-taking skills), student surveys (average score: 1.43, with 1 being Excellent and 5 being Poor), patient surveys (average score: 1.23, with 1 being Excellent and 5 being Poor), faculty surveys (average score: 1.69, with 1 being Excellent and 5 being 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 Assessmen; ANOVA: Analysis of Variance; ANCOVA: Analysis of Covariance; SRS: Student Response Systems; MSX: Multistep Exam; WBA: Workplace-Based Assessments

Table S6. Competency Areas, Assessment methods, Technology & Tools and Innovative Components Employed

Author(s)	Year	Competency Area	Assessment Method	Technology & Tools	Innovative Components
Kasai et al.	2020	Clinical performance & professionalism	Mini-Clinical Evaluation Exercise (Mini-CEX), Professionalism Mini-Evaluation Exercise (P-MEX)	Role-play, peer review	Combining role-play and peer review during clinical rounds, multiple viewpoints in patient care
Shikino et al.	2023	Clinical reasoning & decision-making	Mock patient encounters (Mini-CEX, checklists)	Speech Recognition System (SRS), IC recorder	SRS for generating interview transcripts, providing specific and precise feedback
Haruta et al.	2024	Medical interview skills	Simulated Patient Assessment Tool (SPAT)	None	Development and validation of SPAT, use of 13-item 6-point Likert scale
Bord et al.	2015	Clinical skills	Objective Structured Clinical Examination (OSCE)	Manikin-based simulations, standardized patient encounters	Use of EM Milestones, scenarios with stable and unstable phases, immediate feedback
Malone et al.	2024	Urgent and emergent care skills	Program evaluations, surveys, faculty feedback	High-fidelity (HF) simulations, Virtual Reality (VR) simulations	Comparison of HF and VR simulations for assessing urgent and emergent care competence.
Olupeliyawa et al.	2014	Teamwork and collaboration	Teamwork Mini-Clinical Evaluation Exercise (T-MEX)	None	Focus on clinical encounters, structured reflection, stage-appropriate response scale
Qureshi & Zehra	2020	Communication skills in clinical interactions	OSCE with simulated patients (SPs)	Liverpool Communication Skills Assessment Scale (LCSAS)	Structured feedback from SPs post-interaction
Gran et al.	2016	Feedback and supervision	Structured interviews	StudentPEP	Use of StudentPEP for structured feedback and reflection
Bozzo et al.	2020	Clinical reasoning, history taking, communication	Objective Structured Clinical Examinations (OSCEs)	Standardized simulated patients	Iterative OSCE design with formative feedback and structured evaluation across multiple scenarios
Rouse et al.	2024	Clinical skills and reasoning	Multi-station OSCE	None	Integration of RIME framework, three-station format, video handover for Manager station
Torre et al.	2021	Clinical reasoning	Multistep Exam (MSX)	None	Stepwise format assessing analytical clinical reasoning, structured form for each step
Perrig et al.	2016	Musculoskeletal examination, communication, and professional behavior	Objective Structured Clinical Examination (OSCE)	None specified	Small group interactive teaching with systematic feedback from patients, peers, and instructors
Patel et al.	2024	History-taking skills	Patient interviews (Minicard tool)	Zoom, audio-video recordings	Dual coaching by inpatients and faculty physicians, immediate feedback
Reid et al.	2021	Telemedicine skills, menopause management	Telemedicine OSCE	Zoom, SP checklist	Virtual telemedicine encounter, SP feedback on digital communication, post-encounter note
Okubo et al.	2014	Clinical reasoning skills	Objective Structured Clinical	SNAPPS method, mini-CEX, 1-minute	Use of reflective practice, daily feedback, multiple

			Examination (OSCE), Script Concordance Test (SCT)	preceptor	opportunities for reflection, integration of SNAPPS and mini-CEX.
Costich et al.	2024	Core Entrustable Professional Activities (EPAs)	Workplace-Based Assessments (WBAs)	Qualtrics™, modified Chen entrustment scale	Performance-driven training (PDT) and frame-of-reference training (FORT) for faculty
Klapheke et al.	2022	Clinical skills	Workplace-based assessments (WBAs)	EPA/RIME supervisory scale	Integration of EPAs and RIME model for competency-based education
Parikh et al.	2015	Communication, empathy, and trust in end-of-life care	OSCE with standardized patients (SPs)	Simulation-based end-of-life care scenarios	Integration of palliative care in a surgery setting using SP feedback
Sullivan et al.	2016	Clinical reasoning & decision-making	Virtual Surgical Patient Cases (VSPCs)	Virtual patient cases (simulated surgical scenarios)	Cognitive apprenticeship model using VSPCs with feedback at each attempt
Luo et al.	2023	Self-confidence & clinical competence	Formative OSCE with immediate feedback	None	Immediate feedback after each OSCE station, feedback sandwich method
Kim et al.	2016	Clinical skills	Mini-Clinical Evaluation Exercise (mini-CEX)	None	Implementation across all 3rd-year clerkships, prescribed organ systems for observation
Martinsen et al.	2021	Clinical skills, history-taking, patient examination	Mini-Clinical Evaluation Exercise (mini-CEX)	Observational feedback sessions	Standardized feedback model, requiring a minimum of 8 mini-CEX assessments, OSCE, written tests, and surveys post-placement
Phinney et al.	2022	Feedback and self-reflection in clinical settings	Workplace-Based Assessment (WBA) tool	QR codes, mobile access	Two iterations of WBA: desktop and mobile tools; student self-completion capability, cultural historical activity theory (CHAT) used to examine tensions
Rogausch et al.	2015	Clinical skills	Mini-Clinical Evaluation Exercises (mini-CEX)	None	Analysis of context characteristics and prior clinical skills on mini-CEX scores
Ryan et al.	2021	Core Entrustable Professional Activities (EPAs) including history, physical examination, differential diagnosis, etc.	Ottawa Surgical Competency Operating Room Evaluation (O-SCORE)	Mobile-friendly WBA system	Mobile-based WBA requests driven by students, integrating Core EPAs for formative feedback; frequent assessment to reach reliability
Ryan et al.	2024	Core Entrustable Professional Activities (EPAs)	Workplace-Based Assessments (WBAs)	None	Multi-institutional generalisability study, various implementation strategies, different scales and rater types

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

Table S7. Evaluation, Outcomes, and Practical Implications

Author(s)	Formative/ Summative	Feedback Mechanism	Key Findings	Student Perceptions	Student Perceptions	Implications for Practice
Bord et al.	Formative	Immediate feedback from observing faculty and standardized patients	OSCE effectively discriminates between high- and low-performing students, wide grade distribution	Students received brief feedback after each case, found the OSCE valuable for assessment	OSCE offers a useful tool for assessing EM knowledge and skills, provides insight into student performance	Recommended for use in EM clerkships to assess and improve clinical skills, can be adapted to other institutions
Bozzo et al.	Formative and Summative	Immediate verbal feedback from SPs and written feedback for summative assessments	Demonstrated competence improvement over sessions; high interobserver correlation	Students found value in structured, realistic patient scenarios and feedback	OSCEs effectively enhance competence in high-stakes clerkships	Regular OSCE integration with feedback improves clinical competence; supports using SPs for real-time formative feedback
Costich et al.	Formative	Specific, task-oriented feedback, open-ended response	Improved specificity and timeliness of feedback, greater satisfaction with feedback opportunities	Not directly mentioned in the paper	WBAs feasible in outpatient setting, improve feedback delivery	Further faculty development and training needed, explore student perspectives on WBA impact
Gran et al.	Formative	Timely, constructive feedback	GPs and students emphasized the importance of mutual trust and timely feedback for professional development.	Students valued structured feedback but found it time-consuming.	Structured tools like StudentPEP can enhance feedback processes.	Recommendations for more structured feedback mechanisms in clerkships.
Haruta et al.	Formative	Oral feedback from SPs, SPAT scores	Validity and reliability of SPAT confirmed, significant differences in scores across SPs and scenarios	Students found SP feedback valuable and constructive	SPAT effective for assessing clinical performance, need for standardization of SP assessments	Standardize SP assessment process, consider scenario selection for high-stakes exams
Kasai et al.	Formative	Peer feedback during role-play and peer review sessions	Significant improvement in communication skills, medical interviewing, physical examinations, and professionalism	Students recognized the importance of multiple viewpoints in patient care	Role-play and peer review improve clinical performance and professionalism	Effective educational strategy with limited resources, promotes comprehensive patient care
Kim et al.	Formative	Immediate feedback documented on mini-CEX forms	Increased direct observation of physical exams and history taking, decreased OSCE failure rates	Students reported higher rates of direct observation and feedback	Mini-CEX feasible across all clerkships, improves direct observation and clinical skills	Recommended for use in all clerkships to enhance observation and feedback, improve clinical skills
Klaphek e et al.	Formative	Individualized formative feedback via email	Students' mean skill profile suggested they no longer needed direct supervision on EPA1 and EPA6	Close to half found feedback helpful, most did not want performance shared with residency directors	EPA/RIME framework successful with little additional faculty time commitment	Further development of skills in EPAs 2-5 recommended, more assessments and faculty instruction needed
Luo et al.	Formative	Immediate feedback after each OSCE station	Improved self-confidence and clinical competence in history taking, physical examination, and clinical reasoning, but not in procedural skills	Students reported increased self-confidence both short-term and long-term	Formative OSCE with immediate feedback enhances self-confidence and clinical competence	Recommended for training before clinical clerkship to improve performance and confidence

Malone et al.	Formative	Evaluation surveys, open-ended questions, faculty feedback	HF simulations rated higher (4.6/5) than VR (4.4/5); competence achieved by 91.7% in HF vs. 65.5% in VR ($p \leq .001$).	Students found VR immersive and good preparation; noted need for more practice in VR.	VR can effectively assess urgent care skills but requires acclimation; HF remains superior for summative assessment.	Caution advised when using VR for summative assessment; need for additional practice in VR environments.
Martinsen et al.	Formative	Structured feedback after mini-CEX sessions, OSCE, written test	No significant improvement in OSCE and test scores, some improvement in self-reported skills, feedback valued but sparse	Students valued mini-CEX for direct feedback but desired more detailed improvement feedback	Mini-CEX feasible and valued, but no significant educational impact observed in scores	Continued use of mini-CEX with enhanced training for assessors on providing actionable feedback suggested for impact
Okubo et al.	Formative	Immediate feedback from faculty using 1-minute preceptor and mini-CEX	Significant improvement in clinical reasoning skills, higher scores in mini-CEX, SCT, and OSCE.	Positive feedback on the usefulness of practical and clinical reasoning training.	Reflective practice in an outpatient setting improves clinical reasoning skills.	Recommended for use in clinical clerkships to enhance clinical reasoning skills, adaptable to other educational settings.
Olupeliyawa et al.	Formative	Structured feedback from assessors, student reflections and action plans	Improved collaborative competencies, effective feedback, informed self-assessment	Students valued structured feedback and reflection opportunities	T-MEX facilitates learning through structured feedback, reflection, and situated learning	Recommended for assessing and developing teamwork skills in clinical settings
Parikh et al.	Formative	Immediate SP and faculty feedback	Positive correlation between OSCE performance and trust/communication scores	Students valued the experience, noting increased comfort in handling sensitive conversations	OSCEs are effective for training communication and empathy	End-of-life OSCEs in clerkships can improve early professional skills in challenging scenarios
Patel et al.	Formative	Immediate feedback from inpatients and faculty via Zoom	No significant improvement in history-taking skills, but positive feedback from students and patients	Students appreciated specific and timely feedback, patients enjoyed participating	Dual coaching is beneficial for feedback, though more sessions may be needed for skill improvement	Incorporate more frequent and longitudinal feedback sessions, consider video review for deeper feedback discussions
Perrig et al.	Formative	Immediate feedback from instructor, peers, and patients	Significant improvement in clinical skills (CS) and musculoskeletal skills (MSES) immediately after intervention and sustained improvement at follow-up (T2); IPS skills increased initially but declined over time	Students valued hands-on, small-group practice with real patients and detailed feedback	Additional targeted, small-group interventions have lasting impact on clinical skills	Including structured, feedback-oriented sessions in clerkships enhances skill retention, particularly for practical examinations like musculoskeletal skills
Phinney et al.	Formative	Supervisor observation, mobile WBA forms, student self-completion	Mobile-based WBA increased acceptance, real-time feedback facilitated; tensions in perceived summative feedback impact noted	Students valued real-time feedback, found self-completion easier but sometimes less authentic	Effective for formative assessment, ongoing tension in perceived feedback stakes	Need for cultural shift to view WBA as low-stakes, continuous supervisor support, autonomy in WBA completion for students
Qureshi & Zehra	Formative	Immediate feedback from SPs post-OSCE	Significant improvement in communication skills ($p \leq 0.05$)	Students found SP feedback valuable for skill enhancement	SP feedback is effective in enhancing communication skills	SP feedback should be integrated into clerkship OSCEs for communication training

Reid et al.	Formative	Immediate verbal feedback from SPs, post-encounter note scoring	Students correctly diagnosed menopause but struggled with management strategies, found telemedicine format challenging	Students appreciated practicing telemedicine, found it awkward but useful	Telemedicine OSCE effective for practicing virtual patient encounters and menopause care	Consider integrating telemedicine training earlier, provide didactics on menopause before the encounter
Rogausch et al.	Formative	Narrative comments, structured feedback	Mini-CEX scores influenced more by context characteristics than prior clinical skills	Not directly mentioned in the paper	Context characteristics significantly impact mini-CEX scores	Consider focusing on narrative feedback or improving WBA design to enhance score validity
Rouse et al.	Summative	Structured feedback from faculty, video review	Enhanced OSCE provided a more thorough assessment, reduced grade inflation, improved grade distribution	Students found the OSCE fair and reflective of clinical experiences	Multi-station OSCE with RIME framework effective in assessing clinical skills	Consider further adjustments to time allocation and station content, ongoing evaluation needed
Ryan et al.	Formative	Direct observation and O-SCORE feedback system	Modest reliability in using O-SCORE for medical student assessment; high variability attributed to raters	Students were tasked with initiating WBAs, sometimes rating sessions strategically	O-SCORE showed limited reliability in medical student assessment context	Suggests more robust rater training and possibly reducing number of raters or new scales specific to EPAs
Ryan et al.	Formative	Immediate formative feedback, formal coaching	Low reliability of WBAs for summative decisions, high scores across scales, increased direct observation	Students perceived increased frequency of direct observation	WBAs best used for formative feedback, not reliable for summative entrustment decisions	Further research needed to develop reliable instruments for summative decisions, use WBAs to enhance feedback and observation
Shikino et al.	Formative	Immediate feedback using SRS or IC recorder	Improved diagnostic accuracy, higher Mini-CEX and checklist scores with SRS	Students found SRS feedback more effective and efficient	SRS feedback leads to better diagnostic accuracy and clinical performance	SRS-based feedback is effective and efficient, recommended for improving clinical training
Sullivan et al.	Formative	Immediate feedback, instructor-led discussions	Improved clinical reasoning over attempts, persistent errors in some areas	Students appreciated safe environment for trial & error	VSPCs beneficial for clinical reasoning, decision-making	More cases and varied decision points recommended for enhanced learning
Torre et al.	Formative	Immediate feedback, structured scoring rubric	Significant positive correlation between MSX and Step 2 CS ICE scores, MSX Step 1 most predictive of ICE score	Students benefited from explicit steps in reasoning process	MSX useful for assessing and providing feedback on clinical reasoning	Further studies needed across different learners and medical schools

OSCE: Objective Structured Clinical Examinations; WBA: Workplace-Based Assessments; GPs: General Practitioners; PEP: Peer Education Programs; EPA: Entrustable Professional Activities; P-MEX: Professionalism Mini-Evaluation Exercise; T-MEX: Teamwork Mini-Clinical Evaluation Exercise; MINI-CEX: Mini Clinical Evaluation Exercise; VR: Virtual Reality SP: Simulated Patient; RIME: Reporter Interpreter Manager Educator; SRS: Student Response Systems; MSX: Multistep; Exam Virtual Surgical Patient Cases (VSCP); CHAT: Cultural-Historic Activity Theory; CS: Clinical Skill; MSES: Musculoskeletal; PDT: Performance-driven training; FORT: frame-of-reference training for faculty; HF: High Fidelity Emergency Care.