

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

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

Nicolás Lavados-Toro^{1*}, Joaquín Salgado-González², Silvana Jiménez-Vera³, Álvaro Herrera-Alcaíno², Oscar Jerez-Yañez^{4,5}

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

Recibido: 21/12/25; Aceptado: 28/1/26; Publicado: 29/1/26

Resumen.

El objetivo de esta revisión de alcance fue mapear los métodos utilizados para evaluar la competencia clínica durante los internados médicos e identificar tendencias emergentes asociadas a la incorporación de tecnologías educativas. Se realizó una búsqueda en PubMed, Web of Science y Scopus (octubre de 2024), incluyendo estudios publicados en los últimos 10 años que evaluaran la competencia clínica mediante métodos prácticos apoyados por tecnología. Se seleccionaron 26 estudios que incluyeron a 5.749 estudiantes de medicina en diversos contextos clínicos. Los métodos identificados abarcaron Exámenes Clínicos Objetivos y Estructurados (OSCE), Mini-CEX y otras Evaluaciones Basadas en el Entorno Clínico, incluyendo evaluaciones basadas en Actividades Profesionales Confiables (EPA). Las tecnologías emergentes integradas fueron simulación de alta fidelidad, realidad virtual, telemedicina y sistemas digitales de retroalimentación. En conjunto, los estudios reportaron mejoras en habilidades clínicas, razonamiento diagnóstico, comunicación y autoconfianza, especialmente cuando se utilizó retroalimentación inmediata en evaluaciones formativas. Persisten desafíos relacionados con la variabilidad entre evaluadores, la heterogeneidad metodológica y la necesidad de capacitación docente y recursos tecnológicos adecuados. En conclusión, los hallazgos evidencian una transición hacia modelos evaluativos propios de la educación médica basada en competencias, destacando la importancia de la observación directa, la retroalimentación estructurada y el uso estratégico de tecnologías. Se recomienda fortalecer la estandarización de prácticas evaluativas y desarrollar investigaciones que evalúen su impacto a largo plazo.

Palabras clave: competencia clínica, internado médico, OSCE, Mini-CEX, evaluación formativa, tecnología educativa.

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 competency-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. Introducción

La evaluación de la competencia clínica durante los internados médicos ha experimentado una evolución significativa, reflejando transformaciones más amplias en la educación médica de las últimas décadas (1). Tradicionalmente, los procesos evaluativos se sustentaron en exámenes escritos y orales, junto con observación de desempeño por médicos supervisores, enfoques útiles pero limitados para capturar la complejidad del quehacer clínico (2–3). En respuesta, la educación médica basada en competencias (CBME) ha cobrado fuerza, integrando evaluaciones situadas en el entorno clínico real agrupadas bajo el marco de Workplace-Based Assessments (WBA, Evaluaciones Basadas en el Entorno Clínico), que incluyen herramientas como Mini-Clinical Evaluation Exercise (Mini-CEX, Mini Evaluación Clínica), Direct Observation of Procedural Skills (DOPS, Observación Directa de Habilidades Procedimentales) y evaluaciones basadas en Entrustable Professional Activities (EPA, Actividades Profesionales Confiables) (4–7). Por otra parte, la simulación se complementa con métodos como el Objective Structured Clinical Examination (OSCE, Examen Clínico Objetivo Estructurado), en concordancia con marcos de competencias ampliamente utilizados (p. ej., ACGME) que ordenan dominios como atención al paciente, conocimiento médico, habilidades interpersonales y comunicación, profesionalismo, aprendizaje y mejora basados en la práctica, y práctica basada en sistemas (3,4,8). Paralelamente, la incorporación de tecnologías —plataformas digitales, simuladores de alta fidelidad, realidad virtual (VR) y telemedicina— ha ampliado las posibilidades de evaluación, generando experiencias estandarizadas y seguras y potenciando el aprendizaje reflexivo con retroalimentación oportuna (9–13).

No obstante, su adopción efectiva exige infraestructura, capacitación docente y estrategias de escalamiento adecuadas (14). Para asegurar transferibilidad internacional y adaptación contextual, este trabajo se enmarca en los estándares globales de la World Federation for Medical Education (WFME) para educación médica básica (edición 2020), de enfoque principios-based y no prescriptivo, que proponen dominios transversales (currículo, evaluación, recursos, aseguramiento de la calidad y gobernanza) como palancas de mejora aplicables a diferentes realidades institucionales y regionales (15). Ante esta rápida transformación —impulsada, en parte, por innovaciones tecnológicas— se vuelve necesario mapear sistemáticamente las metodologías actuales de evaluación de competencias clínicas en internados.

Esta revisión de alcance busca responder la siguiente pregunta: ¿Qué métodos de evaluación se utilizan para apoyar el desarrollo de la competencia clínica en estudiantes de medicina durante los internados, y qué tendencias emergen en la implementación de tecnología en estos métodos? Los objetivos específicos son: (i) explorar la literatura existente sobre los métodos de evaluación de la competencia clínica utilizados en internados médicos; (ii) identificar tendencias emergentes en herramientas tecnológicas integradas en dichas evaluaciones; y (iii) caracterizar métodos que articulan procesos formativos y sumativos, identificando fortalezas, debilidades y brechas en la literatura. Para efectos de este manuscrito, el término internado médico se refiere a la fase clínica final de la formación de pregrado, equivalente a medical clerkship en países anglosajones, caracterizada por la participación activa del estudiante en actividades clínicas supervisadas.

2. Métodos

Esta revisión fue diseñada para mapear y caracterizar los métodos de evaluación utilizados para medir la competencia clínica durante los internados médicos, siguiendo las directrices PRISMA-SCR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) (14). Se incluyeron estudios publicados en los últimos 10 años que comprendían métodos que evalúan, tanto formativa como sumativa, la competencia clínica durante el internado de medicina utilizando herramientas tecnológicas. Se excluyeron estudios teóricos o sin aplicación en la práctica clínica, y estudios en contexto preclínico o de posgrado. La búsqueda se realizó en octubre de 2024 en PubMed, Web of Science y Scopus, utilizando combinaciones de palabras claves específicas, como se aprecia en la tabla 1.

Tabla 1. Estrategia de búsqueda empleada en cada base de datos.

Bases de datos	Términos utilizados en la búsqueda	Combinaciones utilizadas
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))	
	#1 TITLE-ABS-KEY (("medical student" OR "medical students") AND ("clinical clerkship" OR "medical clerkship" OR "clinical rotation")	
Scopus	#2 ("competence assessment" OR "competency evaluation" OR "clinical competence assessment" OR "formative assessment" OR "summative evaluation" OR "feedback")	#4: #1 AND #2 AND #3
	#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")	
<p>OSCE: <i>Objective Structured Clinical Examinations</i>; MINI CEX: <i>Mini Clinical Evaluation Exercise</i>; DOBS: <i>Direct Observation of Procedural Skills</i>. La búsqueda estuvo limitada a artículos publicados hasta el 27 de Octubre de 2024.</p>		

Proceso de selección

Inicialmente, se evaluaron los títulos y resúmenes de todos los estudios identificados. Los duplicados fueron eliminados mediante software Rayyan (16). Posteriormente, los estudios considerados relevantes pasaron a revisión completa de texto. Se evaluó cada artículo de forma independiente y, ante discrepancias o dudas, se realizó una reconsideración para asegurar la aplicación coherente de los criterios.

Proceso de recopilación de datos

Los datos de los estudios seleccionados se extrajeron utilizando una tabla predefinida en Google Sheets, que incluía información como autores, año, país, tamaño de muestra, nivel de los participantes, área de competencia evaluada, método de evaluación, tecnología y herramientas utilizadas, y elementos innovadores. Además, se utilizó una segunda tabla para registrar si la evaluación era formativa o sumativa, el mecanismo de retroalimentación, los principales hallazgos, percepciones estudiantiles, conclusiones e implicaciones para la práctica.

Elementos de datos

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.

Evaluación del riesgo de sesgo en los estudios

Aunque las scoping reviews no requieren una evaluación formal del riesgo de sesgo, se incluyó MERSQI para describir la calidad metodológica de los estudios y contextualizar la interpretación de los hallazgos, aportando una comprensión más profunda de la solidez de la evidencia disponible. Se evaluó la calidad metodológica de los estudios utilizando el Medical Education Research Study Quality Instrument (MERSQI) (17). Esta evaluación permite mitigar posibles sesgos presentes en los estudios, particularmente aquellos relacionados con el sesgo de reporte.

Medidas de efecto

Se documentaron sistemáticamente las medidas de efecto para cada uno de los resultados estudiados, considerando medias, desviaciones estándar, coeficientes de correlación, índices de confiabilidad (alfa de Cronbach) y pruebas de significancia (ANOVA, pruebas de Wilcoxon). Las medidas se clasificaron según su propósito estadístico, como la evaluación de confiabilidad, correlación o consistencia temática en los hallazgos. Este enfoque facilitó una comparación de los resultados clínicos y los métodos de evaluación entre ellos.

Métodos de síntesis

Para determinar la elegibilidad de los estudios para la síntesis, se tabularon las características de las intervenciones y los resultados en Google Sheets, comparándolos con criterios predefinidos, permitiendo agrupar los estudios de manera coherente. Los estudios elegibles fueron procesados rigurosamente para estandarizar los formatos de los datos. Luego, los resultados se presentaron mediante resúmenes temáticos, permitiendo agrupaciones comparativas y un análisis transversal.

Declaración

Durante la realización de este trabajo, los autores utilizaron Microsoft Copilot 365 para la extracción de datos, y ChatGPT4o para sintetizar los resultados extraídos, y posteriormente para mejorar la redacción y traducir el artículo original al español. Después de usar esta herramienta, los autores revisaron y editaron el contenido según fuese necesario y asumen toda la responsabilidad por el contenido de la publicación. Además, se aplicaron revisiones humanas iterativas para minimizar sesgos potenciales, asegurar coherencia conceptual y mantener la fidelidad a los datos originales extraídos de los estudios incluidos.

Declaración de ética

Este estudio corresponde a una revisión basada en literatura, por lo tanto, no requirió aprobación de un comité de ética institucional ni consentimiento informado.

3. Resultados

Se identificaron 127 estudios (figura 1) mediante la búsqueda en las tres bases de datos electrónicas. Tras eliminar los duplicados, 45 artículos fueron seleccionados para una revisión completa del texto. Finalmente, 26 estudios cumplieron con los criterios de elegibilidad para la extracción y análisis de datos (18–43).

Características de los estudios

Se incluyeron estudios realizados en varios países, en su mayoría de EEUU, 14 estudios (18, 20, 24, 25, 27, 31, 32, 34, 36, 38–40, 42, 43), seguido por Japón con 4 estudios, (22, 23, 29, 41), Noruega (21, 28), Suiza (33, 37), Pakistán (35), Australia (30), Chile (19) y China (26) (figura 2). Los 26 estudios incluyeron un total de 5.749 estudiantes de medicina, con tamaños muestrales que oscilaron desde 13 participantes en estudios piloto (32) hasta 1.810 participantes en estudios transversales amplios (40). La mayoría fueron publicados entre 2020 y 2024, representando el 65,4% de la literatura revisada.

Los diseños metodológicos incluyeron ensayos controlados aleatorizados (ECA) (26, 28, 35), estudios cuasi-experimentales como estudios pre-post (24, 33, 38), ensayos clínicos no aleatorizados (27), diseños observacionales, como estudios transversales (18, 30, 31, 39, 40), longitudinales (19, 22), de cohorte prospectiva (23, 29), observacionales prospectivos (41) y estudios mixtos (42), además de estudios piloto (20, 25, 32, 36), cualitativos (21, 34), un análisis multinivel (37) y un estudio de validación (43).

Los estudios se centraron principalmente en internados clínicos específicos, como Medicina de Urgencia (18, 27), Medicina Interna (19, 33, 43) e Internados especializados como Ginecología y Obstetricia (OB/GYN) (36), Psiquiatría (25) y Cirugía (26, 31, 42). Algunos estudios multientorno (24, 37) ofrecieron distintas perspectivas al evaluar herramientas en diferentes ambientes clínicos y disciplinas (figura 3). La caracterización completa de los estudios se presenta en el Material Suplementario (Tabla S1). La marcada heterogeneidad en diseños (ECA, estudios transversales, piloto, cualitativos), tamaños muestrales y tipos de resultados evaluados impidió realizar comparaciones directas entre estudios y limita la posibilidad de sintetizar tendencias cuantitativas robustas.

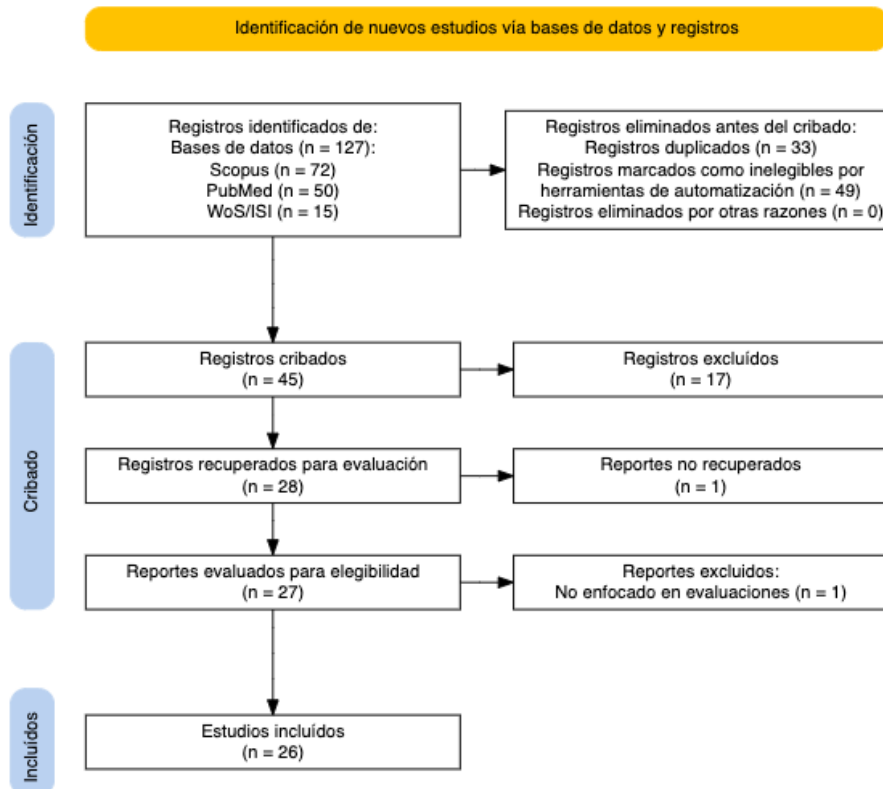


Figura 1. Diagrama de flujo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) de los estudios seleccionados.

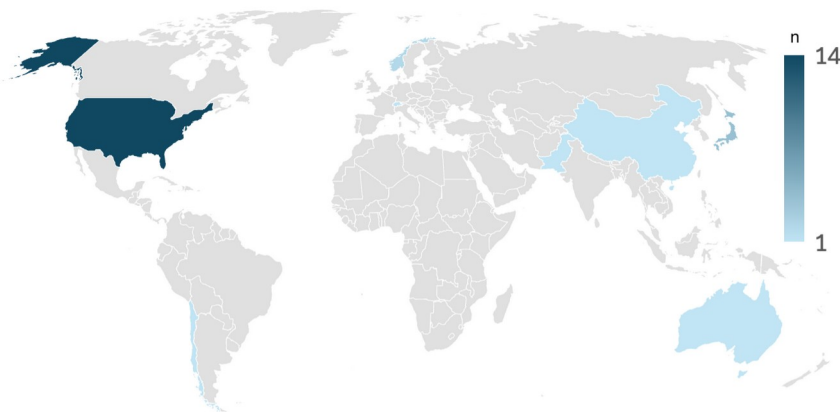


Figura 2. Mapa de distribución de estudios por país..

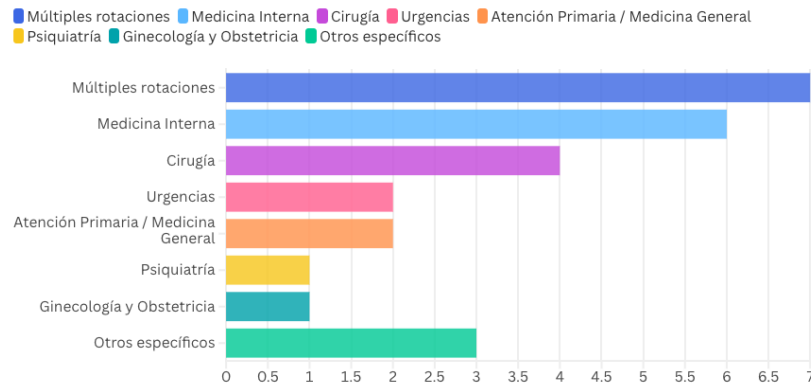


Figura 3. Escenario clínico. *Incluye escenarios como entrevistas médicas en práctica clínica, unidades ambulatorias específicas o internados no categorizables en una especialidad concreta.

Riesgo de sesgo

Las puntuaciones del instrumento MERSQI oscilaron entre 6 y 18 puntos, calculadas según los seis dominios evaluados. El puntaje promedio entre los estudios incluidos fue 12,1 (rango: 10,5 a 18). Siete estudios reclutaron participantes de múltiples instituciones, la mayoría alcanzando una tasa de respuesta superior al 75%, sugiriendo una calidad metodológica moderada a alta, especialmente en estudios con enfoques experimentales y diseños bien estructurados. No obstante, algunos estudios y pilotos presentaron limitaciones en la validez de instrumentos, tamaño muestral y alcance.

Resultados individuales: medidas de efecto

Para todos los resultados, se extrajeron estadísticas resumidas y medidas de efecto con su respectiva precisión, las cuales se presentan en la tabla 2. Por ejemplo, Bord et al. (2015) reportaron una puntuación promedio de OSCE de 70,5% (DE = 7,2%), con una correlación biserial puntual promedio de 0,24, indicando capacidad moderada de discriminación entre estudiantes (18). Bozzo et al. (2020) observaron mejoras significativas en habilidades clínicas durante el internado de medicina interna, con una correlación de Spearman = 0,739 ($p < 0,0001$) y un alfa de Cronbach = 0,8 (19). Costich et al. (2024) demostraron aumento en la entrega de retroalimentación específica y orientada a tareas ($p = 0,006$) en evaluaciones basadas en actividades profesionales confiables en un entorno de atención primaria pediátrica (20). Shikino et al. (2023) utilizaron un sistema de reconocimiento de voz para generar retroalimentación logrando un aumento en la exactitud diagnóstica de 51,3% a 89,7%, con mejoras en puntuaciones de Mini-CEX y listas de verificación (41). Luo et al. (2023): observaron mejoras en autoconfianza y competencias clínicas, aunque las habilidades procedimentales no mejoraron de forma significativa (26). Estos resultados muestran cómo diversas intervenciones educativas influyeron positivamente en el desarrollo de la competencia clínica.

Áreas de competencia

Los estudios evaluaron una gama de competencias en la formación médica, destacando no solo las habilidades clínicas tradicionales, sino también aspectos profesionales y comunicacionales (figura 4).

- Rendimiento clínico y profesionalismo: estudios como los de Kasai et al. (2020) y Shikino et al. (2023) abordaron el desempeño clínico y el profesionalismo, subrayando la toma de decisiones y el razonamiento profesional (23, 41).
- Habilidades clínicas: Investigaciones como las de Bord et al. (2015), Klapheke et al. (2022) y Kim et al. (2016) exploraron competencias clínicas generales, mientras que Malone et al.

(2024) y Martinsen et al. (2021) profundizaron en habilidades específicas como el manejo de urgencias, anamnesis y examen físico (18, 24, 25, 27, 28).

- Comunicación e interacción interpersonal: habilidades en entornos sensibles, como cuidados al final de la vida, fueron abordadas por los estudios de Qureshi & Zehra (2020) y Parikh et al. (2015). También se destacó la empatía y comunicación en los trabajos de Bozzo et al. (2020) y Patel et al. (2024) (19, 31, 32, 35).
- Trabajo en equipo y colaboración: el estudio de Olupeliyawa et al. (2014) resaltó la importancia de las habilidades colaborativas (30).
- Retroalimentación y supervisión: estudios como los de Gran et al. (2016) y Phinney et al. (2022) destacaron el valor de la retroalimentación estructurada y la autorreflexión como mecanismos fundamentales para el aprendizaje clínico (21, 34).
- Actividades Profesionales Confiables: se usaron como referencia en estudios como Costich et al. (2024) y Ryan et al. (2021, 2024), vinculando actividades clínicas concretas con el desarrollo de competencias observables y medibles (20, 39, 40).
- Competencias emergentes: por ejemplo, Reid et al. (2021) incorporó el manejo de la menopausia a través de telemedicina, ampliando el alcance de las habilidades clínicas (36).

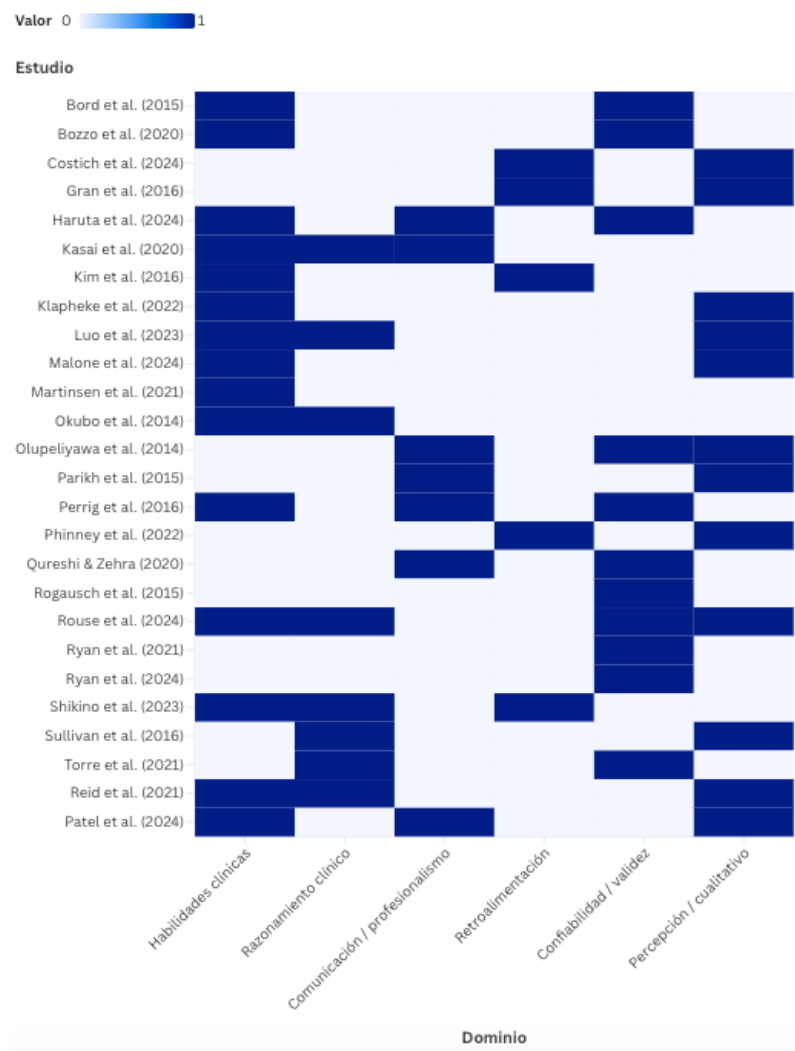


Figura 4. Heatmap de áreas de competencias evaluadas.

Tabla 2. Resultados en términos de desenlaces y medidas de efecto de los estudios incluidos.

Estudio	Resultado	Medida(s) del efecto
Bord et al. (2015)	Desarrollo de un OSCE para la evaluación de habilidades clínicas en el internado de urgencias	Puntaje promedio: 70,5% (DE = 7,2%), Dificultad de ítem: Baja (>80% correcto), Media (50-80% correcto), Alta (<50% correcto), Discriminación del ítem: Buena ($r_{pb} > 0,3$), Regular ($r_{pb} = 0,1-0,3$), Pobre ($r_{pb} < 0,1$), Correlación biserial puntual (r_{pb}): 0,24 (promedio)
Bozzo et al. (2020)	Mejoría en habilidades clínicas en medicina interna	Medidas de tendencia central, pruebas de normalidad, pruebas de Wilcoxon pareadas ($p < 0,05$), modelos de regresión lineal múltiple, correlación de Spearman ($r = 0,739$, $p < 0,0001$), alfa de Cronbach (0,8), valores p ($p \leq 0,05$)
Costich et al. (2024)	Implementación y percepción docente de las EBA basadas en EPA en atención primaria ambulatoria durante el internado de pediatría	Retroalimentación docente: Aumento en la entrega de retroalimentación específica y orientada a tareas ($p = 0,006$), mayor satisfacción con las oportunidades de retroalimentación (no significativo), más retroalimentación dentro de 24 h (no significativo). Temas de entrevistas: Beneficios de las EBA, barreras a la retroalimentación, sugerencias de mejora
Gran et al. (2016)	Experiencias de retroalimentación durante el internado	Análisis temático para identificar temas clave y experiencias de retroalimentación, consistencia de los temas entre entrevistas
Haruta et al. (2024)	Validación del Simulated Patient Assessment Tool (SPAT) para evaluar desempeño clínico con pacientes simulados	Análisis factorial exploratorio: reveló dos factores (comunicación y desempeño médico), varianza explicada 60,47%. Consistencia interna: alfa de Cronbach global 0,929. ANOVA de un factor: diferencias significativas entre pacientes simulados ($F(34,760)=16,79$, $p < 0,001$) y escenarios ($F(20,774)=11,39$, $p < 0,001$). Validez convergente: correlación moderada ($r=0,212$, $p < 0,05$) entre SPAT y ECOE post-CC
Kasai et al. (2020)	Mejoría en desempeño clínico y profesionalismo en estudiantes de internado	Mini-CEX: Mejoras significativas en anamnesis (pre: $5,52 \pm 1,05$, post: $6,57 \pm 0,75$, $p < 0,001$), examen físico (pre: $5,38 \pm 0,82$, post: $6,78 \pm 0,87$, $p < 0,001$), profesionalismo (pre: $5,87 \pm 1,00$, post: $7,13 \pm 0,81$, $p < 0,001$), juicio clínico, consejería, organización/eficiencia y competencia global. P-MEX: Mejoras en relación médico-paciente (pre: $3,00 \pm 0,34$, post: $3,38 \pm 0,24$, $p < 0,001$) y habilidades reflexivas (pre: $3,15 \pm 0,32$, post: $3,50 \pm 0,26$, $p < 0,001$). Test de rangos con signo de Wilcoxon ($p < 0,05$)
Kim et al. (2016)	Implementación de mini-CEX en todos los internados de tercer año y su impacto en observación directa y habilidades clínicas	Cumplimiento: 92% de formularios completados, 78% con retroalimentación específica. Reporte estudiantil: aumento significativo en observación directa de examen físico (ej: cirugía 49%→87%, $p < 0,0001$). AAMC GQ: más observación docente en anamnesis y examen físico. ECOE: disminución de reprobación de 12% a 2% ($p = 0,0046$). Valores p: $< 0,0001$ (observación en cirugía), 0,0046 (ECOE)
Klapheke et al. (2022)	Estudio piloto de EBA usando EPAs y modelo RIME en internado de psiquiatría	Promedios en EPAs y RIME: perfil >4.0 en EPA1, EPA6 y rol de reportero en RIME. Retroalimentación estudiantil: 44% la consideró útil, 32,2% evaluaciones justas. Retroalimentación docente: facilidad de completar EPA = 3,89 (DE=1,27), claridad en explicación = 4,33 (DE=0,5)
Luo et al. (2023)	Mejoría en autoconfianza y competencia clínica en internos de cirugía	Evaluación de autoconfianza (SCA): mejoras en todas las estaciones (ej. anamnesis pre: $2,68 \pm 0,82$, post: $3,47 \pm 0,84$, $p < 0,01$; examen físico pre: $2,47 \pm 0,84$, post: $3,42 \pm 0,77$, $p < 0,01$; RCP, razonamiento clínico, asepsia quirúrgica). Mini-CEX: mejora

		significativa (grupo ECOE: 6,59 \pm 0,62 vs. control: 5,31 \pm 1,09, $p < 0,01$). DOPS: sin diferencia significativa (grupo ECOE: 5,15 \pm 0,58 vs. control: 4,96 \pm 0,63, $p > 0,05$)
Malone et al. (2024)	Habilidades en atención urgente y emergente	Prueba de rangos con signo de Wilcoxon ($T = 337,5$, $p = 0,001$), prueba de McNemar ($p \leq 0,001$), análisis temático para datos cualitativos, significancia determinada por $p \leq 0,05$
Martinsen et al. (2021)	Habilidades clínicas medidas por evaluaciones mini-CEX	ANOVA ($F = 3,603$, $p = 0,066$), ANCOVA ($F = 1,884$, $p = 0,179$), puntajes promedio (3,5–3,6 sobre 4), desviaciones estándar (0,55–0,63)
Okubo et al. (2014)	Evaluación de un programa de internado clínico en atención ambulatoria para mejorar el razonamiento clínico de estudiantes	Mini-CEX: más alto en todas las áreas para estudiantes del programa. SCT: significativamente más alto post-curso (pre: 77,4 \pm 10,5; post: 86,8 \pm 10,2). ECOE: mayores puntajes para grupo de intervención vs control
Olupeliyawa et al. (2014)	Evaluación del impacto educativo del T-MEX en la colaboración de estudiantes en equipos de salud	Coefficiente de generalización: 0,62 con tres formularios, 0,80 predicho con ocho formularios. Análisis de contenido de retroalimentación y reflexiones. Análisis temático de grupos focales y entrevistas
Parikh et al. (2015)	Comunicación, empatía y confianza en cuidados al final de la vida	Estadísticas descriptivas de puntajes promedio (media = 89,0%, DE = 6,7%), desviaciones estándar, correlación con puntajes de confianza ($r = 0,325$, $p < 0,01$) y habilidades comunicacionales ($r = 0,383$, $p < 0,01$)
Perrig et al. (2016)	Mejoría en habilidades de examen musculoesquelético	Prueba de Friedman ($p < 0,001$ para CS, $p < 0,001$ para MSES, $p < 0,01$ para IPS), prueba de rangos de Wilcoxon (CS: $p < 0,001$, MSES: $p < 0,001$, IPS: $p < 0,001$), prueba U de Mann-Whitney (CS: $p < 0,01$, MSES: $p < 0,01$, IPS: $p < 0,01$), alfa de Cronbach (0,47–0,83), valores p ($p \leq 0,05$)
Phinney et al. (2022)	Retroalimentación y autorreflexión en entornos clínicos	Análisis temático para identificar temas clave y tensiones, consistencia de temas entre entrevistas
Qureshi & Zehra (2020)	Habilidades comunicacionales usando retroalimentación de pacientes simulados	Coefficiente de correlación de Pearson ($r = 0,83$) para consistencia interna de LCSAS (alfa de Cronbach = 0,83), mejora de puntajes pre/post-test (diferencia media = 1,5, $p \leq 0,05$)
Rogausch et al. (2015)	Influencia de habilidades clínicas previas y características del contexto sobre puntajes mini-CEX en internados	Análisis de regresión: posición clínica del docente fue el predictor más influyente (coeficiente de regresión = 0,55, $p < 0,001$ para residentes vs jefes de departamento). Complejidad de tarea y tamaño de clínica: predictores significativos. ECOE: no predictor significativo. Correlación: débil entre ECOE y mini-CEX ($r = 0,26$ global, $r = 0,27$ por dominio). Valores p : $p < 0,001$ posición docente, $p < 0,05$ complejidad y tamaño
Rouse et al. (2024)	ECOE mejorado para neutralizar inflación de notas y evaluación más completa de habilidades clínicas	Comparación ECOE pre y post-intervención (pre: $M = 94,25\%$, $DE = 5,65\%$, post: $M = 81,00\%$, $DE = 6,88\%$), encuestas estudiantiles (promedio: 4,4 Reporter, 4,4 Interpreter, 4,2 Manager, 3,5 tiempo asignado, 3,5 dificultad)
Ryan et al.	EPAs centrales medidas por escala O-	Teoría de generalización para confiabilidad (coeficiente Phi = 0,19–0,44), puntajes promedio (3,48–3,62), componentes de

(2021)	SCORE	varianza (estudiante = 3,5%–8%, evaluador = 29,6%–50,3%)
Ryan et al. (2024)	Evaluación de confiabilidad de EBA para decisiones de confianza sumativa en educación médica	Coficiente Phi: >0,7 umbral de confiabilidad aceptable. Varianza atribuida al aprendiz: <10% en la mayoría de análisis. Número de observaciones necesarias: rango 3 a >560, mediana 60
Shikino et al. (2023)	Evaluación de efectividad de retroalimentación SRS en razonamiento clínico durante encuentros simulados	Precisión diagnóstica: grupo SRS mayor (pre: 51,3%, post: 89,7%) vs IC recorder (pre: 57,5%, post: 67,5%) (p=0,037). Mini-CEX: mejoras significativas en anamnesis, examen físico, profesionalismo, organización/eficiencia y competencia global (p<0,001). Checklist: puntaje total mayor SRS (pre: 12,2, post: 16,1) vs IC recorder (pre: 13,1, post: 13,8) (p<0,001). Tiempo de retroalimentación: SRS 22,6 ±2,1 min vs IC 27,7 ±2,1 min (p=0,04)
Sullivan et al. (2016)	Mejoría en habilidades de toma de decisiones clínicas	ANOVA de medidas repetidas (diverticulitis aguda: Pillai's Trace = 0,807, F(2,36) = 75,279, p<0,000; hemorragia GI: Pillai's Trace = 0,822, F(2,19)=43,941, p<0,000), análisis temático, valores p (p ≤ 0,05)
Torre et al. (2021)	Validación del Multistep Exam (MSX) para evaluar razonamiento clínico analítico en internado de medicina interna	Correlación: MSX correlación positiva significativa con Step 2 CS ICE (r=0,26, p<0,01). Regresión lineal múltiple: MSX predictor significativo de Step 2 CS ICE (β=0,19, p<0,001), explicando 4% adicional de varianza más allá de NBME Medicina y ECOE Medicina. Confiabilidad: alfa de Cronbach 0,70–0,80
Reid et al. (2021)	Evaluación de OSCE de telemedicina para manejo de menopausia en internado de Ginecología/Obstetricia	Puntajes de nota post-encuentro (mediana: 20/45), encuestas estudiantiles (78% incomodidad con telemedicina, 66% valor educativo excelente o superior), precisión diagnóstico diferencial (100% identificaron menopausia/perimenopausia, 84% hipertiroidismo)
Patel et al. (2024)	Evaluación de doble coaching en habilidades de anamnesis de estudiantes de internado	Puntajes minicard (sin mejora significativa), encuestas estudiantiles (promedio 1,43; 1=Excelente, 5=Pobre), encuestas pacientes (promedio 1,23; 1=Excelente, 5=Pobre), encuestas docentes (promedio 1,69; 1=Excelente, 5=Pobre)

OSCE: Objective Structured Clinical Examinations (Exámenes Clínicos Estructurados y Objetivos); **EPA:** Entrustable Professional Activities (Actividades Profesionales Confiables); **SPAT:** Simulated Patient Assessment Tool (Herramienta de Evaluación con Paciente Simulado); **Mini-CEX:** Mini Clinical Evaluation Exercise (Mini Ejercicio de Evaluación Clínica); **RIME:** Reporter Interpreter Manager Educator (Reportero – Intérprete – Gestor – Educador); **SCA:** Self-Confidence Assessment (Evaluación de Autoconfianza); **ANOVA:** Analysis of Variance (Análisis de Varianza); **ANCOVA:** Analysis of Covariance (Análisis de Covarianza); **SRS:** Student Response Systems (Sistemas de Respuesta Estudiantil); **MSX:** Multistep Exam (Examen de Múltiples Etapas); **WBA:** Workplace-Based Assessments (Evaluaciones Basadas en el Entorno Clínico)

Métodos de evaluación

Los estudios revisados aplicaron una variedad de enfoques evaluativos, adaptados a los objetivos de aprendizaje y habilidades específicas (figura 5).

- Examen Clínico Estructurado Objetivo (OSCE): ampliamente adoptado para evaluar habilidades clínicas, comunicación y profesionalismo. Estudios como Bord, Bozzo y Parikh lo utilizaron con pacientes estandarizados para aumentar el realismo y la relevancia clínica (18, 19, 31). Algunos, como Rouse et al. (2024), emplearon formatos multiescenario para evaluar razonamiento y habilidades procedimentales (38).
- Dentro del marco de Workplace-Based Assessments (WBA), se identificaron herramientas como Mini-CEX, herramienta clave para observaciones breves y estructuradas del desempeño clínico en tiempo real usada por Kasai, Kim, Martinsen, entre otros, para evaluar interacción con pacientes, razonamiento diagnóstico y profesionalismo (23, 24, 28), DOPS y evaluaciones basadas en EPA.
- Los estudios que implementaron WBA en sentido amplio—incluyendo EPA-based assessments y retroalimentación en tiempo real—fueron especialmente relevantes en Costich, Ryan y Phinney (20, 34, 40).
- Herramientas especializadas: como el Simulated Patient Assessment Tool validado por Haruta et al. (2024) para entrevistas clínicas, o los casos quirúrgicos virtuales en Sullivan et al. (2016) que simulan decisiones clínicas complejas (22, 42).
- Evaluaciones de habilidades colaborativas: el instrumento T-MEX, aplicado por Olupeliyawa et al., evaluó específicamente el trabajo en equipo en contextos clínicos (30).

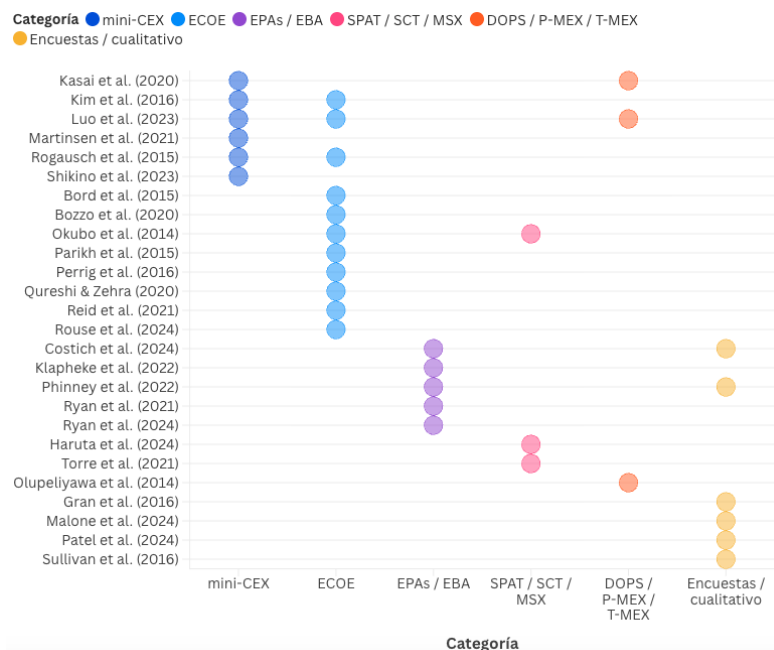


Figura 5. Dotplot de Métodos de Evaluación empleados.

Tecnologías y componentes innovadores

Los estudios emplearon tecnologías para fortalecer los procesos de evaluación, centradas en el aprendizaje activo, personalización y retroalimentación inmediata.

- Simulaciones y herramientas virtuales: destaca los simuladores de alta fidelidad y la realidad virtual. Por ejemplo, Malone et al. (2024) comparó ambos tipos de simulación para

entrenamiento en urgencias; Bord combinó maniqués y pacientes simulados con retroalimentación inmediata en urgencias (18, 27).

- Juego de roles, revisión entre pares y pacientes estandarizados (SPs): Kasai usó dramatizaciones clínicas para retroalimentación entre compañeros. Qureshi y Bozzo integraron SPs para entrenar habilidades comunicacionales (19, 23, 35).
- Reconocimiento de voz y plataformas digitales: Shikino et al. (2023) utilizó un sistema de reconocimiento automático para generar transcripciones de entrevistas y dar retroalimentación específica. Phinney y Ryan usaron plataformas móviles con códigos QR para facilitar evaluaciones WBAs en tiempo real (34, 39, 41).
- Telemedicina y aprendizaje remoto: Reid y Patel incorporaron herramientas para encuentros clínicos virtuales, evaluando habilidades en comunicación digital y atención remota (32, 36).
- Retroalimentación estructurada y autorreflexión: Gran implementó la herramienta StudentPEP para guiar la reflexión posterior a los encuentros clínicos. Luo utilizó el método “sandwich” de retroalimentación inmediata tras cada estación de OSCE (21, 26).
- Componentes innovadores: desde el modelo EPA/RIME de Klapheke, hasta el uso del entrenamiento orientado al rendimiento en Costich, estas innovaciones apuntan a una evaluación más objetiva y coherente (20, 25).

Evaluaciones formativas/sumativas, mecanismos de retroalimentación y hallazgos clave

Los estudios utilizaron una gran variedad de métodos evaluativos, tanto sumativos como formativos, cada uno incorporando métodos específicos de retroalimentación para reforzar el aprendizaje de los estudiantes.

- Evaluaciones formativas con retroalimentación inmediata. El uso de evaluaciones formativas fue una práctica adoptada para fomentar el aprendizaje en tiempo real. Bord et al. (2015) y Luo et al. (2023) ofrecieron retroalimentación inmediata tras estaciones OSCE, lo que se tradujo en mejoras significativas en habilidades clínicas (18, 26). Kasai et al. (2020) integró la retroalimentación entre pares durante dramatizaciones clínicas, observando mejoras marcadas en comunicación, entrevistas médicas y profesionalismo (23).
- Retroalimentación para el desarrollo profesional. Adoptó distintos formatos. Gran et al. (2016) subrayó la importancia de la retroalimentación oportuna y basada en la confianza mutua entre docentes y estudiantes en entornos de atención primaria (21). Kim et al. (2016) y Haruta et al. (2024) utilizaron formularios estructurados (Mini-CEX y SPAT) para proporcionar retroalimentación específica orientada a tareas, fortaleciendo el aprendizaje basado en habilidades clínicas (22, 24).
- Enfoques innovadores de retroalimentación. Phinney et al. (2022) introdujo evaluaciones WBA móviles con retroalimentación en tiempo real y autocompletado (34). Costich et al. (2024) empleó retroalimentación abierta y orientada a tareas, mejorando la especificidad y oportunidad de los comentarios, aumentando la satisfacción estudiantil (20).
- Evaluaciones sumativas para medir habilidades integrales. Aunque menos frecuentes, algunas herramientas fueron utilizadas con fines sumativos: Rouse et al. (2024) desarrolló un formato OSCE mejorado con revisión en video, que redujo la inflación de notas y aumentó la confiabilidad de la evaluación (38). Bozzo et al. (2020) combinó retroalimentación formativa

inmediata con evaluaciones escritas sumativas estructuradas por parte de SPs, resultando en mejoras sostenidas de competencia clínica (19).

- Hallazgos clave sobre competencia clínica y precisión diagnóstica. Los estudios mostraron que las evaluaciones formativas con retroalimentación inmediata conducen a mejoras consistentes en las habilidades clínicas. Shikino et al. (2023), utilizó un sistema de reconocimiento de voz para entregar retroalimentación precisa que mejoró significativamente la precisión diagnóstica de los estudiantes (41). Sullivan et al. (2016) mostró que las discusiones guiadas por instructores mejoraban el razonamiento clínico, aunque persistían algunos errores (42).
- Impacto en la autoconfianza y la retención de habilidades. Varios estudios demostraron que la retroalimentación también potencia la autoconfianza estudiantil. Luo et al. (2023) reportó un aumento en la autoconfianza tras OSCEs formativos, aunque con menos impacto en habilidades procedimentales complejas (26). Perrig et al. (2016) mostró que habilidades musculoesqueléticas mejoraron de forma sostenida tras intervenciones estructuradas, mientras que las habilidades interpersonales presentaron mejoras iniciales que decrecieron con el tiempo (33).
- Percepciones estudiantiles. Los estudiantes valoraron altamente las evaluaciones con retroalimentación estructurada e inmediata: Bozzo et al. (2020) y Luo et al. (2023) encontraron que apreciaban especialmente los OSCEs con escenarios clínicos realistas y retroalimentación directa, fortaleciendo su confianza y competencia (19,26). Kasai et al. (2020) destacó que reconocían el valor de recibir múltiples perspectivas, especialmente en sesiones de revisión entre pares (23). Martinsen et al. (2021) señaló que, aunque valoraban las sesiones Mini-CEX, expresaron el deseo de recibir sugerencias más detalladas (28).
- Conclusiones sobre el desarrollo de habilidades clínicas. Los estudios concluyen que las evaluaciones formativas con retroalimentación inmediata son claves para el desarrollo progresivo de habilidades. Bord et al. (2015) destacó que los OSCEs eran útiles para identificar fortalezas y debilidades en medicina de urgencia, ofreciendo retroalimentación (18). Kim et al. (2016) recomendó implementar Mini-CEX de forma transversal en todos los internados, para mejorar la observación directa y retroalimentación (24).
- Implicaciones para la práctica. Diversos estudios ofrecen recomendaciones prácticas para integrar estas herramientas evaluativas en la formación médica. Bozzo et al. (2020) sugirieron la incorporación de OSCEs con pacientes estandarizados (SPs) para proporcionar retroalimentación tanto formativa como sumativa, potenciando la competencia clínica en escenarios de alta exigencia (19). Además, Costich et al. (2024) recomendaron fortalecer el desarrollo docente para apoyar las evaluaciones basadas en el entorno clínico (WBAs) en el ámbito ambulatorio, destacando la necesidad de entrenar profesionales en la entrega de retroalimentación (20). Patel et al. (2024) y Reid et al. (2021) propusieron implementar sesiones de retroalimentación más frecuentes e incorporar formación en telemedicina desde etapas tempranas de la carrera médica, con el fin de adaptarse a los entornos clínicos contemporáneos (32, 36).

El resumen de los datos extraídos de tecnología y herramientas, componentes innovadores, hallazgos claves e implicaciones prácticas de los estudios incluidos se presentan en la tabla 3. La versión completa de esta tabla puede ser revisada en la Tabla S6 y S7 del suplemento.

Tabla 3. Tecnología y Herramientas, Componentes Innovadores, Hallazgos Claves e Implicaciones Prácticas de los estudios incluidos.

Autor(es)	Tecnología & Herramientas	Componentes Innovadores	Hallazgos clave	implicaciones prácticas
Kasai et al. (2020)	Juego de roles, revisión entre pares	Integración de roles múltiples	El juego de roles y la revisión entre pares mejoran el desempeño clínico en comunicación, entrevistas médicas, exámenes físicos y profesionalismo.	Estrategia educativa efectiva con recursos limitados, promueve una atención integral del paciente.
Shikino et al. (2023)	SRS	Transcripciones automáticas	La retroalimentación mediante SRS conduce a una mayor precisión diagnóstica y mejor desempeño clínico con mayores puntuaciones en Mini-CEX y SRS.	La retroalimentación basada en SRS es efectiva y eficiente, recomendada para mejorar la formación clínica.
Haruta et al. (2024)	-	Desarrollo y validación de SPAT	Se confirmó la validez y fiabilidad del SPAT, aunque se requiere estandarización en las evaluaciones de pacientes estandarizados.	Estandarizar el proceso de evaluación de pacientes estandarizados, considerar la selección de escenarios para exámenes de alta exigencia.
Bord et al. (2015)	OSCE con simulación	Escenarios con fases estables e inestables	El OSCE discrimina eficazmente entre estudiantes de alto y bajo rendimiento, con una amplia distribución de calificaciones.	Recomendado en medicina de urgencias para evaluar y mejorar habilidades clínicas; puede adaptarse a otras instituciones.
Malone et al. (2024)	Simulaciones HF y de VR	Comparación entre HF y VR	Las simulaciones en HF fueron mejor valoradas y mayormente alcanzadas que en VR	Se aconseja precaución al usar VR para evaluación sumativa; se requiere práctica adicional en entornos de VR.
Olupeliyawa et al. (2014)	T-Mex	Reflexión estructurada,	El uso de T-Mex se tradujo en una mejoría en competencias colaborativas, retroalimentación efectiva y autoevaluación.	Recomendado para evaluar y desarrollar habilidades de trabajo en equipo en entornos clínicos.
Qureshi & Zehra (2020)	LCSAS + SPs	Retroalimentación de SPs	La retroalimentación de los pacientes estandarizados es efectiva para mejorar las habilidades de comunicación.	La retroalimentación de SPs debe integrarse en los OSCE de las rotaciones clínicas para el entrenamiento en comunicación.
Gran et al. (2016)	StudentPEP	Reflexión guiada	StudentPEP puede mejorar la confianza mutua y brindar retroalimentación oportuna.	Recomendaciones para mecanismos de retroalimentación.
Bozzo et al. (2020)	SP	Diseño iterativo multiescenario	Los OSCE mejoran eficazmente la competencia en rotaciones clínicas de alta exigencia. Alta correlación entre observadores.	La integración regular del OSCE con retroalimentación mejora la competencia clínica; respalda el uso de pacientes estandarizados para retroalimentación formativa en tiempo real.
Rouse et al. (2024)	OSCE multiescena	Marco RIME, entrega en video	El OSCE multiescena con el marco RIME proporcionó una evaluación más completa, redujo la inflación de calificaciones y mejoró la distribución de notas.	Considerar ajustes adicionales en la asignación de tiempo y el contenido de las estaciones; se requiere evaluación continua.
Torre et al. (2021)	MSX	Formulario estructurado	El MSX es útil para evaluar y proporcionar retroalimentación sobre el razonamiento clínico.	Se necesitan más estudios en la materia.
Perrig et al. (2016)	Entrenamiento en grupos pequeños	Retroalimentación multifuente	Mejoría significativa en habilidades clínicas y musculoesqueléticas inmediatamente después de la intervención y sostenida en el seguimiento.	Sesiones estructuradas orientadas a la retroalimentación en las rotaciones clínicas mejora la retención de habilidades, particularmente en exámenes prácticos como las habilidades musculoesqueléticas.
Patel et al. (2024)	Zoom, grabaciones de audio y video	Doble coaching	No hubo mejoría significativa en las habilidades de anamnesis, pero se recibió retroalimentación positiva de estudiantes y pacientes.	Incorporar sesiones de retroalimentación más frecuentes y longitudinales; considerar la revisión en video para discusiones más profundas.
Reid et al. (2021)	Zoom, checklist SP	Retroalimentación digital	El OSCE en telemedicina es efectivo para practicar con pacientes y el manejo de la menopausia, pero desafiante según los estudiantes.	Considerar la integración temprana de la formación en telemedicina y proporcionar instrucción sobre la menopausia antes del encuentro.

Okubo et al. (2014)	SNAPPS, mini-CEX, preceptor de 1 minuto	Integración de SNAPPS y mini-CEX	Mejoría en razonamiento clínico; mayores puntajes en mini-CEX, SCT y OSCE.	Recomendado su uso en rotaciones clínicas; adaptable a otros entornos educativos.
Costich et al. (2024)	Qualtrics™, escala de confianza modificada de Chen	PDT y FORT para docentes	Las WBAs son factibles en el entorno ambulatorio, con mejor especificidad, oportunidad y satisfacción con la retroalimentación.	Se necesita mayor desarrollo y capacitación docente; se recomienda explorar las perspectivas de los estudiantes sobre el impacto de las WBAs.
Klapheke et al. (2022)	Escala supervisora EPA/RIME	Integración de EPAs y modelo RIME	El marco EPA/RIME tuvo éxito con un compromiso adicional mínimo de tiempo por parte de los docentes.	Se recomienda un mayor desarrollo de habilidades en las EPAs, con más evaluaciones e instrucción docente.
Parikh et al. (2015)	Simulación en cuidados de fin de vida	Integración de CP en cirugía mediante SP	Correlación positiva entre el desempeño en OSCE y las puntuaciones de confianza/comunicación y empatía.	Los OSCE de cuidados al final de la vida en las rotaciones clínicas pueden mejorar habilidades profesionales tempranas en escenarios desafiantes.
Sullivan et al. (2016)	Escenarios quirúrgicos virtuales	VSPCs secuenciales	Los VSPCs son beneficiosos para el razonamiento clínico y la toma de decisiones.	Se recomiendan más casos y puntos de decisión variados para un aprendizaje mejorado.
Luo et al. (2023)	-	“Feedback sandwich”	El OSCE formativo con retroalimentación inmediata potencia la autoconfianza y la competencia clínica, pero no en habilidades procedimentales.	Recomendado para la formación previa a la rotación clínica, con el fin de mejorar el desempeño y la confianza.
Kim et al. (2016)	-	Implementación transversal	El mini-CEX es factible en todas las rotaciones clínicas, mejora la observación directa y las habilidades clínicas, con disminución en las tasas de fracaso en OSCE.	Recomendado para su uso en todas las rotaciones clínicas para mejorar la observación, la retroalimentación y las habilidades clínicas.
Martinsen et al. (2021)	Retroalimentación observacional	Evaluaciones múltiples con retroalimentación estandarizada	El mini-CEX es factible y valorado, pero no se observó un impacto educativo significativo en los puntajes, pesar de cierta mejora en habilidades autoinformadas	Se sugiere el uso continuo del mini-CEX con mayor capacitación de los evaluadores para lograr impacto.
Phinney et al. (2022)	Códigos QR, acceso móvil	Dos iteraciones de WBA, uso de CHAT	EPA-RIME apoya evaluación formativa; requiere ajustar cultura de retroalimentación.	Necesidad de un cambio cultural para considerar las WBAs como evaluaciones de bajo riesgo; apoyo continuo del supervisor y mayor autonomía de los estudiantes en su realización.
Rogausch et al. (2015)	Ninguna	Análisis multinivel	Las puntuaciones de mini-CEX estuvieron más influenciadas por las características del contexto que por las habilidades clínicas previas.	Considerar el enfoque en la retroalimentación narrativa o mejorar el diseño de los Mini-CEX (y WBA) para aumentar la validez de las puntuaciones.
Ryan et al. (2021)	Sistema WBA compatible con móviles	WBA móvil con retroalimentación frecuente	Baja fiabilidad de las WBAs para decisiones sumativas.	Se necesita más investigación para desarrollar instrumentos fiables en decisiones sumativas; usar las WBAs para mejorar la retroalimentación.
Ryan et al. (2024)	Ninguna	WBA multi-institución	Fiabilidad modesta en el uso del O-SCORE; alta variabilidad atribuida a los evaluadores.	Se sugiere una capacitación más sólida de los evaluadores y posiblemente reducir el número de evaluadores o implementar nuevas escalas específicas

OSCE: Objective Structured Clinical Examinations (Exámenes Clínicos Estructurados y Objetivos); **SRS:** Student Response Systems (Sistemas de Respuesta Estudiantil); **EM Milestone:** Emergency Medicine Milestone (Hitos de Medicina de Emergencias). **PEP:** Peer Education Programs (Programas de Educación entre Pares); **EPA:** Entrustable Professional Activities (Actividades Profesionales Confiables); **SPAT:** Simulated Patient Assessment Tool (Herramienta de Evaluación con Paciente Simulado). **Mini-CEX:** Mini Clinical Evaluation Exercise (Mini Ejercicio de Evaluación Clínica); **RIME:** Reporter Interpreter Manager Educator (Reportero – Intérprete – Gestor – Educador); **MSX:** Multistep Exam (Examen de Múltiples Etapas). **WBA:** Workplace-Based Assessments (Evaluaciones Basadas en el Entorno Clínico); **P-MEX:** Professionalism Mini-Evaluation Exercise (Mini Evaluación de Profesionalismo); **T-MEX:** Teamwork Mini-Clinical Evaluation Exercise (Mini Evaluación Clínica de Trabajo en Equipo); **VR:** Virtual Reality (Realidad Virtual); **HF:** High Fidelity Emergency Care (Atención de Emergencias de Alta Fidelidad); **SP:** Simulated Patient (Paciente Simulado); **VSCP:** Virtual Surgical Patient Cases (Casos de Paciente Quirúrgico Virtual); **CHAT:** Cultural-Historic Activity Theory (Teoría de la Actividad Histórico-Cultural); **CS:** Clinical Skill (Habilidad Clínica); **MSES:** Musculoskeletal Examination Skills (Habilidades de Examen Musculoesquelético); **PDT:** Performance-Driven Training (Entrenamiento Orientado al Desempeño); **FORT:** Frame-of-Reference Training for Faculty (Entrenamiento de Marco de Referencia para Docentes).

4. Discusión

Interpretación

La presente revisión ofrece un panorama actualizado de los métodos de evaluación de competencias clínicas durante el internado médico, evidenciando una tendencia sostenida hacia enfoques coherentes con la educación médica basada en competencias (CBME), que priorizan la observación directa, la retroalimentación oportuna y la integración selectiva de tecnologías digitales (6–8, 15). En conjunto, los estudios incluidos muestran que la evaluación del desempeño clínico debe abarcar dimensiones técnicas y transversales (comunicación, profesionalismo, razonamiento clínico, trabajo en equipo), en línea con marcos ampliamente utilizados en educación médica (3, 6–8). Este movimiento responde, además, a transformaciones del entorno asistencial posteriores a la pandemia por COVID-19, que aceleraron ajustes en la medición de aprendizajes y el uso de modalidades virtuales (1).

Evaluación situada y retroalimentación formativa como motores de aprendizaje

Los hallazgos concuerdan en que las Workplace-Based Assessments (WBA) —marco que agrupa herramientas como Mini-CEX, DOPS y evaluaciones basadas en EPA— contribuyen al desarrollo progresivo de competencias en la práctica clínica real al facilitar observación estructurada y retroalimentación inmediata focalizada en el desempeño (9, 23-24, 28, 34, 37, 40-41). Este valor formativo es consistente con literatura que posiciona la retroalimentación como elemento central para cerrar brechas entre conocimiento y ejecución clínica, favorecer la autorreflexión y sostener trayectorias de mejora (7–9). No obstante, la evidencia sugiere que la retroalimentación, por sí sola, no garantiza mejoras uniformes en todas las áreas: ciertas habilidades procedimentales complejas requieren práctica deliberada repetida y oportunidades de desempeño auténtico, lo que acota el impacto de intervenciones exclusivamente tecnológicas o puntuales (23, 26, 33). Asimismo, la necesidad de alinear la evaluación con marcos de progresión por hitos y de medir competencias de manera válida se vincula con desafíos conceptuales y operativos que deben considerarse al diseñar y utilizar WBA en pregrado (44).

Integración tecnológica y telemedicina: oportunidades y condiciones de implementación

La incorporación de simulación, realidad virtual y telemedicina amplía las posibilidades de evaluación al crear escenarios seguros, estandarizados y replicables, y al preparar al estudiantado para contextos clínicos crecientemente digitalizados (10, 11, 27, 32, 36, 42). En términos curriculares, esta transformación se apoya en estrategias de aprendizaje activo con tecnología (45), en la evidencia que vincula la simulación con resultados educativos relevantes (46) y en el avance de herramientas digitales/IA con potencial formativo y evaluativo (47). Este impulso se vio reforzado tras la pandemia (48) y se alinea con llamados a consolidar la CBME con criterios de calidad y coherencia curricular (49). En este marco, la realidad virtual y las simulaciones de distinta fidelidad ofrecen rutas complementarias cuya adopción debe ser contextualizada según objetivos pedagógicos, recursos y carga de trabajo (50–52). En contextos latinoamericanos, la implementación progresiva y costo-consciente es clave: pueden priorizarse Mini-CEX en formato papel o apps móviles simples, pacientes simulados entrenados localmente y retroalimentación estructurada basada en guías narrativas; además, la simulación de baja fidelidad y los estudios de caso virtuales asincrónicos constituyen alternativas costo-efectivas frente a tecnologías de mayor complejidad (46, 50–52). El desarrollo del pensamiento reflexivo como competencia transversal sitúa a la retroalimentación y la reflexión guiada como piezas que convierten la información evaluativa en aprendizaje (53–56). Particular relevancia adquiere la telemedicina como ámbito de evaluación y aprendizaje: permite valorar comunicación mediada por tecnología, razonamiento con información distribuida y toma de decisiones en entornos remotos (57–59). La aceptación y disposición de estudiantes, así como los modelos de preparación institucional y escalamiento, influyen en su

implementación sostenible (60–61). En el pregrado, la incorporación planificada de entrenamiento en telemedicina y su evaluación estructurada muestran buena factibilidad y utilidad percibida (62–63).

Heterogeneidad y calidad metodológica: implicaciones para interpretar los hallazgos

La heterogeneidad de diseños, tamaños muestrales, contextos y desenlaces dificulta comparaciones directas y limita inferencias causales robustas, algo esperable en mapeos de alcance y en el campo de la evaluación clínica en pregrado (3,4,7,8). La variabilidad entre evaluadores y la fiabilidad modesta observada en algunas herramientas dentro del conjunto de WBA — particularmente en estudios multicéntricos y escalas de observación— subrayan la necesidad de entrenamiento de evaluadores y refinamiento de instrumentos (20,34,39,40). Para contextualizar la solidez de la evidencia, se consideró la calidad metodológica mediante MERSQI; si bien las puntuaciones sugieren calidad moderada a alta en parte de los estudios, persisten limitaciones en validez de instrumentos, potencia estadística y estandarización de medidas, que deben ponderarse al interpretar el conjunto (17, 18–43).

Implicaciones para el currículo y la política institucional

Los resultados refuerzan la conveniencia de ecosistemas de evaluación longitudinal, de bajo riesgo y centrados en el desarrollo, combinando observación directa, retroalimentación narrativa y uso estratégico de tecnología con propósitos educativos claros (7–9, 24, 34). A nivel institucional, se recomienda la estandarización de prácticas de WBA (definiciones, rúbricas, expectativas, frecuencia mínima), con formación docente sistemática —p. ej., frame-of-reference training y performance-driven training— para mejorar consistencia, especificidad y oportunidad del feedback (20, 39, 40, 49).

Por otra parte, la integración de nuevas herramientas evaluativas debe alinearse con prioridades de salud pública (envejecimiento, salud mental) y la creciente alfabetización digital del equipo de salud. La incorporación de marcos como los de AAMC y ACGME puede orientar competencias en alfabetización digital, colaboración interprofesional y atención virtual centrada en el paciente (48, 50–52). Dado que gran parte de la literatura se fundamenta en marcos normativos de Estados Unidos, es pertinente vincular estos hallazgos con estándares globales (p. ej., WFME) para favorecer transferibilidad y adaptación en escenarios latinoamericanos, donde las condiciones institucionales, culturales y de recursos difieren (15).

Brechas de conocimiento y prioridades de investigación

A partir del mapeo realizado, se identifican brechas prioritarias: 1) Impacto a largo plazo de evaluaciones formativas sobre desempeño clínico y resultados de pacientes; 2) Validez y confiabilidad de herramientas emergentes (incluidas escalas y plataformas digitales) en diferentes contextos e idiomas; 3) Estandarización de taxonomías y marcos de WBA (exigencias mínimas de observaciones y entrenamiento de evaluadores) para decisiones educativas informadas; 4) Equidad y escalabilidad de tecnologías (simulación y telemedicina) en entornos con recursos limitados; y 5) Evidencia contextual en Latinoamérica e hispanohablantes, donde la adopción tecnológica y la organización de internados difieren de los marcos normativos predominantes (9, 20, 34, 39, 40, 45–52, 57–63). Para atender estas brechas se requieren estudios multicéntricos con diseños más homogéneos y mayor potencia, estrategias explícitas de calibración de evaluadores, y evaluaciones económicas y de implementación que informen decisiones de política educativa (7–9, 39, 40, 49–52).

Limitaciones

Esta revisión presenta varias limitaciones que condicionan la interpretación y el alcance de los hallazgos. En primer lugar, la heterogeneidad de los estudios incluidos—en diseños (ECA, cuasi-experimentales, observacionales, cualitativos), tamaños muestrales, contextos y desenlaces—

dificulta comparaciones directas y limita la posibilidad de establecer relaciones causales robustas, algo esperable en el campo de la evaluación clínica en pregrado y en mapeos de alcance (3, 4, 7, 8). Esta diversidad se refleja también en la variabilidad entre evaluadores y la fiabilidad modesta observada en algunas herramientas dentro del conjunto de Workplace-Based Assessments (WBA), lo que subraya la necesidad de capacitación docente sistemática y de refinamiento de instrumentos para mejorar consistencia y uso educativo (20, 34, 39, 40). Asimismo, la mezcla de medidas objetivas y autorreportadas (p. ej., percepciones, autoconfianza) y de resultados de corto plazo añade incertidumbre sobre la sostenibilidad del impacto educativo a mediano y largo plazo (23, 26, 33).

Segundo, aunque se aplicó MERSQI para describir la calidad metodológica de los estudios, su uso en una scoping review no busca excluir evidencia, sino contextualizar la solidez del conjunto. Persisten limitaciones en validez de instrumentos, potencia estadística y estandarización de medidas que deben considerarse al interpretar los resultados (17, 18–43). Además, la heterogeneidad de marcos conceptuales y de prácticas evaluativas reportadas—en ocasiones sin criterios uniformes de implementación—limita la síntesis cuantitativa y refuerza el carácter descriptivo de esta revisión (3, 6–8).

Tercero, puede existir sesgo de publicación, dado que estudios con resultados positivos o significativos podrían estar sobrerrepresentados, mientras que intervenciones con efectos nulos reciben menor difusión, fenómeno descrito de forma recurrente en investigación educativa (3, 6–8). A esto se suma un posible sesgo de idioma, pues la búsqueda priorizó literatura en inglés y español; trabajos en otros idiomas podrían no haber sido captados, afectando la exhaustividad del mapeo (16).

Cuarto, el predominio geográfico de la evidencia en países desarrollados (principalmente Estados Unidos, Japón y Suiza) limita la transferibilidad de ciertas conclusiones. En particular, la aplicabilidad de tecnologías de alta fidelidad o realidad virtual puede verse reducida en contextos con menores recursos, como ocurre en múltiples facultades de medicina latinoamericanas, donde la infraestructura, el soporte técnico y la disponibilidad de tiempo docente son más restringidos (46, 50–52). Esta situación refuerza la necesidad de promover investigación en contextos hispanohablantes y latinoamericanos, considerando realidades institucionales, culturales y de recursos distintas, para favorecer una adaptación contextual de los modelos de evaluación basados en competencias (15).

Quinto, respecto de la integración tecnológica, los estudios muestran beneficios de simulación y VR en el aprendizaje y la evaluación; sin embargo, persisten barreras de implementación, brechas de acceso y costos que condicionan su escalamiento, especialmente fuera de centros de alta complejidad (46, 50–52). Del mismo modo, la telemedicina se perfila como un ámbito indispensable para la formación y la evaluación clínica, pero su adopción efectiva depende de la preparación institucional, la aceptación del estudiantado y la planificación curricular, factores que varían ampliamente entre entornos y pueden afectar la equidad en su implementación (57–63). Estas asimetrías implican que parte de los efectos positivos observados en entornos con mayor preparación tecnológica no sean directamente extrapolables a instituciones latinoamericanas sin estrategias graduales y diseños costo-efectivos (50–52, 60–61).

Sexto, aunque la adopción de marcos como AAMC/ACGME puede guiar la alineación curricular hacia competencias en alfabetización digital, colaboración interprofesional y atención virtual centrada en el paciente, gran parte de la literatura analizada se basa en normativas de EE.UU., lo que requiere vincular estos hallazgos con estándares globales (p. ej., WFME) para mejorar su transferibilidad y pertinencia en América Latina (15, 48–52). En ausencia de estudios locales comparables y de validaciones transculturales, los procesos de adopción deberían

acompañarse de evaluaciones de implementación y análisis económicos que consideren la realidad de cada sistema académico-asistencial (49–52, 60–61).

Finalmente, decisiones metodológicas propias del diseño—ventana temporal (últimos 10 años), conjunto de bases consultadas y criterios de inclusión/exclusión—pueden haber dejado fuera literatura relevante anterior o en fuentes no indexadas, lo que constituye otra limitación inherente a este tipo de mapeos (16). En conjunto, estas limitaciones indican que los resultados deben interpretarse principalmente como una caracterización de prácticas y tendencias emergentes, y no como evidencia definitiva de efectividad comparativa; su principal valor reside en orientar decisiones curriculares y prioridades de investigación bajo una lectura contextualizada de la calidad metodológica y la factibilidad de adopción en distintos escenarios (7–9,17,46,50–52,57–63).

5. Conclusiones

- La revisión de 26 estudios y más de 5.700 estudiantes confirma una transición global hacia la educación médica basada en competencias, donde herramientas como los métodos evaluativos estandarizados como OSCE y las herramientas agrupadas bajo el marco de WBA—incluyendo Mini-CEX, DOPS y evaluaciones basadas en EPA—se consolidan como el estándar actual para evaluar desempeño clínico real. Se observa, además, una integración creciente de tecnologías como la realidad virtual, la telemedicina y sistemas de reconocimiento de voz para estandarizar y enriquecer estos procesos.
- El valor fundamental de estas metodologías reside en su capacidad para ofrecer retroalimentación inmediata y estructurada, lo cual impacta positivamente en la autoconfianza del interno y en su precisión diagnóstica. Sin embargo, la implementación efectiva de estas innovaciones no es automática; requiere superar brechas significativas en la capacitación docente y asegurar la sostenibilidad de los recursos tecnológicos institucionales.
- En síntesis, para que estas innovaciones sean transformadoras, las facultades deben equilibrar adecuadamente procesos formativos y sumativos, e incorporar tecnología como facilitadora del aprendizaje clínico reflexivo.

Material suplementario: Anexo 1. Estrategia de búsqueda para cada base de datos; Items extraídos; Características de los estudios; Evaluación de Calidad; Resultados y medición de efecto; Área de Competencia, Método de Evaluación, Tecnología y Herramientas, Componentes Innovadores, Hallazgos Clave, Conclusiones, Implicaciones prácticas; Tipo de retroalimentación, Mecanismo de Retroalimentación, Percepción de los Estudiantes.

Financiación: No ha habido financiación.

Declaración de conflicto de interés: Los autores declaran no tener ningún conflicto de interés.

Contribuciones de los autores: Conceptualización: NLT, SJV, AHA, Procuramiento de datos: NLT, Metodología/análisis formal/validación: NLT, Administración del proyecto: NLT, AHA, Escritura - borrador original: NLT, JSG, SJV, Escritura - revisión y edición: NLT, JSG, AHA, OJY.

6. Referencias.

1. Hauer K, Lockspeiser T, Chen H. The COVID-19 pandemic as an imperative to advance medical student assessment: three areas for change. *Acad. Med.* **2021**, *96*, 182–185. <https://doi.org/10.1097/ACM.00000000000003764>
2. Ten C, Regehr G. The power of subjectivity in the assessment of medical trainees. *Acad. 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 competency. *Acad. 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 competency-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 competency-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. *Acad. 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. *Acad. Pediatr.* **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 WFME. Basic medical education: WFME global standards for quality improvement. The 2020 revision. *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. Descripción y análisis de ECOE con pacientes simulados en internado de Medicina Interna 2016–2017 en Facultad de Medicina Universidad de Chile. *Rev. Méd. Chile* **2020**, 148, 810–817. <https://doi.org/10.4067/S0034-98872020000600810>

20. Costich M, Friedman S, Robinson V, Catalozzi 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. *Acad. 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 randomised 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. Res.* **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 inpatients 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, Fluett 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. *Acad. 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 menopausal 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 J. R.I.M.E. and reason: multi-station OSCE enhancement to neutralize grade inflation. *Med. Educ. Online* **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. *Acad. 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. *Acad. 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 competency-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 amongst 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. *Acad. Med.* **2020**, 95, 1908–1912. <https://doi.org/10.1097/ACM.0000000000003622>

