

Material suplementario

Más allá de la inmersión tecnológica: realidad aumentada y realidad mixta en la educación médica de grado. Revisión crítica de estudios comparativos

Este material presenta la estrategia bibliográfica conservada, la caracterización crítica de los informes examinados y los límites de las inferencias por desenlace. La valoración es cualitativa y no equivale a una evaluación formal mediante RoB 2, ROBINS-I o GRADE.

S1. Estrategia de búsqueda y trazabilidad documental

Período de búsqueda y cobertura

La localización bibliográfica se desarrolló progresivamente entre enero y septiembre de 2026, con fecha de cierre declarada el 29 de septiembre de 2026. Los informes examinados se publicaron entre 2017 y 2026; la actualización de publicaciones recientes se centró en 2024-2026. Las fechas de cada consulta individual no se conservaron.

Fuentes de información

PubMed/MEDLINE fue la fuente de búsqueda bibliográfica documentada. PubMed Central y los portales editoriales se emplearon para recuperar informes y comprobar sus datos. Las revisiones previas de Baashar et al. (2022) (3), Williams et al. (2025) (5) y Zhang et al. (2025) (4) se utilizaron como vías de identificación de estudios primarios y como antecedentes; no se contabilizaron como unidades de análisis.

Tabla S1. Fuentes consultadas y documentación disponible

Fuente	Finalidad	Alcance del registro
PubMed/MEDLINE (PubMed)	Identificación y actualización bibliográfica de estudios primarios.	Ecuación principal conservada; período enero-septiembre de 2026. Sin historial íntegro de filtros, fechas por consulta ni exportaciones de resultados.
PubMed Central	Recuperación de textos completos y cotejo bibliográfico.	No se documenta una búsqueda independiente con ecuación, fecha y recuento propios.
Sitios editoriales indexados	Acceso a publicaciones originales y verificación de información.	Sin estrategias independientes ni recuentos históricos recuperables.
Revisiones de Baashar (3), Williams (5) y Zhang (4)	Identificación retrospectiva de referencias de trabajos primarios.	Sin inventario original que individualice el origen, los duplicados y la selección de cada registro.

Ecuación conservada de PubMed/MEDLINE

("augmented reality"[Title/Abstract] OR "mixed reality"[Title/Abstract] OR "extended reality"[Title/Abstract] OR hologram*[Title/Abstract] OR HoloLens[Title/Abstract]) AND ("medical students"[Title/Abstract] OR "medical student"[Title/Abstract] OR "undergraduate medical education"[Title/Abstract] OR "medical education"[Title/Abstract]) AND (anatomy[Title/Abstract] OR histology[Title/Abstract] OR "clinical skills"[Title/Abstract] OR "procedural training"[Title/Abstract] OR surgery[Title/Abstract] OR simulation[Title/Abstract] OR learning[Title/Abstract] OR training[Title/Abstract])

La ecuación emplea OR entre sinónimos y AND entre los bloques de tecnología, población y ámbito educativo. El campo [Title/Abstract] restringe la búsqueda al título o resumen; hologram* incorpora truncamiento. La sintaxis documentada no contiene límites explícitos de publicación, idioma o diseño. No puede establecerse si se aplicaron filtros adicionales en la interfaz.

Localización histórica y límites de reproducibilidad

Los estudios históricos se identificaron a partir de referencias contenidas en revisiones previas; la actualización se orientó a publicaciones de 2024-2026. No se conserva el inventario completo de resultados recuperados, duplicados, decisiones de exclusión ni fechas exactas por búsqueda. Por ello, la ecuación puede volver a ejecutarse, pero no permite reproducir retrospectivamente el mismo flujo de identificación y cribado. Esta limitación fundamenta la clasificación del manuscrito como revisión crítica y no como revisión sistemática exhaustiva.

S1A. Comprobación aritmética de recuentos históricos no auditables

Una versión anterior del manuscrito incorporó un diagrama PRISMA con los recuentos que se transcriben a continuación. Se verifica únicamente su coherencia aritmética: no se conservan los registros originales que permitan acreditar identificación, deduplicación, cribado o exclusiones. Por esta razón el diagrama no se incluye en la versión final del artículo y estos números no se presentan como resultados reproducibles de selección.

Tabla S1A. Coherencia interna de los recuentos declarados

Etapas	n	Comprobación o alcance
Identificación en bases y editoriales	146	$146 + 22 = 168$
Identificación mediante revisiones y referencias	22	Vía adicional sin listado histórico de registros
Registros identificados (total)	168	Suma de ambas vías
Duplicados eliminados	31	$168 - 31 = 137$
Registros cribados por título/resumen	137	$137 - 95 = 42$
Registros excluidos por título/resumen	95	Etapas declaradas del cribado
Informes buscados a texto completo	42	$42 - 3 = 39$
Informes no recuperados	3	Sin lista histórica de informes no recuperados
Informes evaluados a texto completo	39	$39 - 21 = 18$
Informes excluidos a texto completo	21	Motivos declarados: $6 + 5 + 6 + 4 = 21$
Informes incluidos en la figura original	18	16 de pertinencia directa y dos contextuales en el corpus analizado

Referencia metodológica: Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. J Med Libr Assoc. 2021;109(2):174-200. doi:10.5195/jmla.2021.962.

Participación de los autores en los procedimientos

Las declaraciones de contribución atribuyen a tres autores la búsqueda y selección, a dos la extracción y curación, y a tres la interpretación. Esa atribución no demuestra que el cribado o la evaluación crítica se realizaran de forma independiente o por duplicado. No se conservan registros individuales de discrepancias o arbitraje.

Tabla S1B. Participación documentada por fase

Proceso	Participación declarada	Independencia y discrepancias
Búsqueda y selección	Tres autores	No consta la distribución individual por fase ni un procedimiento documentado de resolución de discrepancias.
Extracción y curación	Dos autoras	No puede acreditarse extracción independiente por duplicado ni consenso registrado.
Evaluación crítica e interpretación	Tres autores en la interpretación	No consta evaluación separada del sesgo o arbitraje formal.

S2. Fichas de caracterización y valoración crítica de los informes examinados

Los informes se presentan cronológicamente según diseño, tamaño muestral y limitaciones para la inferencia. La ausencia de información no equivale a riesgo bajo, y un tamaño muestral reducido afecta principalmente a la precisión. Se distinguen los resultados obtenidos en tareas simuladas de la transferencia clínica y los dos informes de carácter contextual.

1. Noll et al., 2017 (29)

Diseño y muestra. Ensayo aleatorizado; n = 44.

Valoración crítica. RA móvil integrada en aprendizaje autónomo combinado. Los resultados sobre conocimiento y aceptación corresponden a una intervención y tarea específicas; no acreditan transferencia clínica.

Fuente: <https://doi.org/10.2196/mhealth.7943>

2. Stojanovska et al., 2020 (15)

Diseño y muestra. Diseño por cohortes con componente cruzado y control adicional; n = 64.

Valoración crítica. La comparación con disección cadavérica combina modalidades y grupos. La ausencia de diferencias significativas no demuestra equivalencia formal ni eficiencia económica.

Fuente: <https://doi.org/10.1007/s40670-019-00834-x>

3. Schoeb et al., 2020 (21)

Diseño y muestra. Ensayo aleatorizado, simple ciego; n = 164.

Valoración crítica. La puntuación OSCE evalúa el cateterismo vesical en una tarea concreta. Un mejor resultado en esa estación no demuestra desempeño sostenido con pacientes ni competencia clínica global.

Fuente: <https://doi.org/10.1186/s12909-020-02450-5>

4. Weeks et al., 2021 (14)

Diseño y muestra. Ensayo aleatorizado; n = 30.

Valoración crítica. La medición de conocimiento se basó en una prueba inmediata de cinco ítems. El resultado favorable no permite establecer retención, transferencia ni una mejora anatómica general.

Fuente: <https://doi.org/10.1016/j.acra.2020.07.008>

5. Pickering et al., 2022 (16)

Diseño y muestra. Estudio cuasialeatorizado; 200 participantes declarados y denominadores analíticos variables.

Valoración crítica. El informe presenta 127 + 75 observaciones en la prueba de elección múltiple y 84 + 62 = 146 en respuestas cortas y puntuación global. Los denominadores no son intercambiables; la asignación tampoco acredita aleatorización plena.

Fuente: <https://doi.org/10.1002/ase.2113>

6. Farshad-Amacker et al., 2023 (25)

Diseño y muestra. Diseño cruzado con orden aleatorio; n = 44.

Valoración crítica. Cada participante realizó ambas modalidades ecográficas. La interpretación debe considerar el orden, la práctica previa y el aprendizaje entre procedimientos; no corresponde tratarlo como un ensayo paralelo.

Fuente: <https://doi.org/10.1007/s00330-022-09220-5>

7. Guha et al., 2023 (22)

Diseño y muestra. Ensayo aleatorizado de factibilidad; n = 36.

Valoración crítica. Se comunicó evaluación objetiva de habilidades quirúrgicas con evaluadores cegados. La muestra y el carácter de factibilidad limitan la precisión y la extrapolación a desempeño con pacientes.

Fuente: <https://doi.org/10.1186/s12909-023-04122-6>

8. Hayasaka et al., 2023 (24)

Diseño y muestra. Tres grupos aleatorizados; n = 30 (10 participantes por grupo).

Valoración crítica. La precisión se evaluó en un modelo de práctica de anestesia epidural. Los resultados describen desempeño en simulación, no transferencia a la práctica clínica.

Fuente: <https://doi.org/10.1186/s12871-023-02133-w>

9. Malik et al., 2023 (30)

Diseño y muestra. Estudio educativo aleatorizado; 73 participantes inicialmente y 56 analizados.

Valoración crítica. Se comunicaron cuatro exclusiones por experiencia previa y 13 por asistencia o evaluaciones incompletas; quedaron grupos de 35 y 21 estudiantes. La ganancia fue de 12 frente a 9 puntos en una prueba de 20. El visor fue usado por los docentes, no por los estudiantes; las exclusiones condicionan la inferencia.

Fuente: <https://doi.org/10.1002/lary.30364>

10. d'Aiello et al., 2023 (31)

Diseño y muestra. Estudio exploratorio con tres grupos asignados aleatoriamente; n = 59 (20, 20 y 19).

Valoración crítica. Se compararon RM, vídeos holográficos y diapositivas. Las diferencias variaron según el contenido evaluado y no sustentan un efecto uniforme de la tecnología.

Fuente: <https://doi.org/10.1007/s10916-023-01959-8>

11. Zeedzen-Scheffers et al., 2024 (17) — evidencia contextual

Diseño y muestra. Pretest-posttest; n = 28; muestra mixta de Medicina y Ciencias Biomédicas.

Valoración crítica. El informe describe la asignación de manera ambigua (aleatorización frente a orden de inscripción/sexo). Al no presentar resultados separables para Medicina, se utiliza como evidencia contextual.

Fuente: <https://doi.org/10.1038/s41598-024-76379-w>

12. Sarkar et al., 2024 (26)

Diseño y muestra. Ensayo piloto aleatorizado; n = 34.

Valoración crítica. No se observaron diferencias significativas en los principales resultados técnicos de punción lumbar, aunque aumentó la confianza percibida. El tamaño piloto limita la precisión y no autoriza inferir equivalencia.

Fuente: <https://doi.org/10.1089/jmxr.2024.0001>

13. Hadida Barzilai et al., 2025 (23)

Diseño y muestra. Ensayo aleatorizado; n = 21.

Valoración crítica. En mastoidectomía simulada se informaron puntuaciones de 19,5/25 frente a 12/25, valoradas por dos otólogos cegados. El cegamiento favorece la medición; el escenario simulado y la muestra pequeña limitan la generalización.

Fuente: <https://doi.org/10.1002/lary.31791>

14. Zingg et al., 2025 (18)

Diseño y muestra. Ensayo aleatorizado; n = 48.

Valoración crítica. La intervención integra holografía, adaptación y gamificación. Aunque favoreció las pruebas teórica y práctica, no puede aislarse el efecto causal del visor respecto de los demás componentes.

Fuente: <https://doi.org/10.1097/ACM.0000000000006012>

15. Felten et al., 2025 (27)

Diseño y muestra. Ensayo aleatorizado; 188 estudiantes y 53 procedimientos analizados por protocolo.

Valoración crítica. El éxito clínico fue de 13/28 frente a 10/25 (p = 0,85). La unidad de análisis del desenlace primario son los procedimientos, y la interrupción por futilidad limita la precisión; la ausencia de superioridad no demuestra equivalencia.

Fuente: <https://doi.org/10.1186/s12909-025-07536-6>

16. Mulder et al., 2026 (28)

Diseño y muestra. Ensayo aleatorizado de no inferioridad; n = 56.

Valoración crítica. Se emplearon márgenes de 3 mm para EPSS y 4 mm para TAPSE, con exclusión de dos mediciones por desenlace. La inclusión de valores extremos impedía confirmar la no inferioridad de EPSS; no se infiere coste-efectividad ni competencia ecográfica general.

Fuente: <https://doi.org/10.1186/s12909-026-08897-2>

17. Xiao et al., 2026 (20) — evidencia contextual

Diseño y muestra. Estudio comparativo; n = 152, según el resumen primario.

Valoración crítica. El resumen describe resultados en histología y seguimiento a diez semanas, pero no permite verificar completamente asignación, pérdidas, instrumentos o decisiones analíticas. Se considera evidencia contextual y no sustento directo de eficacia.

Fuente: <https://doi.org/10.1002/ca.70139>

18. Starszak et al., 2026 (19)

Diseño y muestra. Ensayo aleatorizado de tres grupos; n = 236.

Valoración crítica. Se compararon RM, impresión 3D y control. La ventaja de adquisición inmediata y la ausencia de diferencias significativas en retención deben interpretarse por separado; no se demuestra transferencia clínica.

Fuente: <https://doi.org/10.1186/s12909-026-08959-5>

S3. Criterios de síntesis por desenlace

La síntesis separa conocimiento y comprensión espacial, desempeño procedimental simulado, retención y transferencia clínica. Los resultados de satisfacción, confianza, motivación y usabilidad se consideran complementarios y no sustituyen medidas objetivas de competencia. Las comparaciones sin significación estadística no se interpretan como pruebas de equivalencia.

Las estimaciones procedentes de metaanálisis anteriores se utilizan solo para contextualizar los hallazgos y no se combinan de nuevo con los estudios primarios. Tampoco se estima un efecto global propio ni se emplea el número de resultados estadísticamente significativos como indicador de eficacia. Las divergencias se examinan atendiendo al diseño, comparador, instrumento, duración del seguimiento y pertinencia para Medicina de grado.

La síntesis interpretativa no sustituye una evaluación formal de riesgo de sesgo por desenlace ni una calificación GRADE de certeza. La información documental no permite acreditar evaluación por duplicado o arbitraje independiente.

S4. Matriz de observaciones metodológicas por estudio

La matriz describe características relevantes para la interpretación de cada informe; no constituye una aplicación formal de RoB 2 o ROBINS-I ni asigna categorías globales de riesgo de sesgo. Se diferencia la falta de información de la presencia demostrada de una amenaza metodológica.

Tabla S4. Diseño, limitaciones observables y alcance de los hallazgos

Estudio y diseño	Asignación y datos	Medición y seguimiento	Alcance de la inferencia
Noll 2017 (29) — Aleatorizado	Secuencia y ocultación de la asignación no verificadas con la documentación disponible	Conocimiento y aceptación en RA móvil; transferibilidad no medida	Apoyo a aprendizaje autónomo concreto
Stojanovska 2020 (15) — Cohortes y componente cruzado	Asignación entre cohortes y posible efecto de orden	Pruebas prácticas; no equivalencia formal	Apoyo directo con reservas
Schoeb 2020 (21) — Aleatorizado; simple ciego	Sin valoración formal completa del sesgo por resultado	OSCE de cateterismo; cegamiento limitado a componentes	Apoyo a OSCE específico
Weeks 2021 (14) — Aleatorizado	Muestra de 30; precisión limitada	Prueba inmediata de cinco ítems; sin retención	Apoyo de alcance inmediato

Estudio y diseño	Asignación y datos	Medición y seguimiento	Alcance de la inferencia
Pickering 2022 (16) — Cuasialeatorizado	Asignación no plenamente aleatoria; n variable según prueba	MCQ y respuesta corta no coincidentes; datos completos distintos	Apoyo directo condicionado al desenlace
Farshad-Amacker 2023 (25) — Cruzado con orden aleatorio	Posible efecto de orden/práctica previa	Tiempo/tarea de ecografía intervencionista	Apoyo directo condicionado al diseño cruzado
Guha 2023 (22) — RCT de factibilidad	n=36; precisión limitada	Destreza simulada con evaluadores cegados	Apoyo a destreza simulada, no clínica
Hayasaka 2023 (24) — Tres grupos aleatorizados	10 participantes por brazo	Precisión en simulador de epidural	Apoyo limitado a modelo de práctica
Malik 2023 (30) — Estudio educativo aleatorizado	Exclusiones tras asignación, 73 inicialmente, 56 analizados	Prueba de 20 puntos; visor usado por docentes	Apoyo al formato didáctico ensayado
d'Aiello 2023 (31) — Tres grupos aleatorizados	Muestras 20/20/19; análisis exploratorio	Conocimiento según contenido y modalidad	Apoyo exploratorio
Zeedzen-Scheffers 2024 (17) — Asignación descrita de forma ambigua	Medicina y Ciencias Biomédicas no separadas	Ganancia neuroanatómica; población mixta	Contexto, no inferencia específica a Medicina
Sarkar 2024 (26) — Piloto aleatorizado	n=34; potencia y precisión limitadas	Indicadores técnicos no significativos; confianza mayor	Apoyo limitado para punción simulada
Hadida Barzilai 2025 (23) — RCT prospectivo	n=21; precisión externa limitada	Fresado valorado por dos evaluadores cegados	Apoyo a simulación de mastoidectomía
Zingg 2025 (18) — RCT	Intervención multicomponente sin aislamiento de RA	Pruebas teórica/práctica inmediatas	Apoyo al conjunto de RA adaptativa y gamificación
Felten 2025 (27) — RCT con análisis por procedimiento	53 procedimientos por protocolo; interrupción por futilidad	Éxito clínico 13/28 vs 10/25; no se demuestra equivalencia	Apoyo condicionado a transferencias observadas
Mulder 2026 (28) — RCT de no inferioridad	EPSS sensible a valores excluidos	EPSS/TAPSE, márgenes específicos y análisis sensibles	Apoyo restringido al análisis de POCUS

Estudio y diseño	Asignación y datos	Medición y seguimiento	Alcance de la inferencia
Xiao 2026 (20) — Comparativo; solo resumen	No verificables asignación, pérdidas y resultados completos	Histología, seguimiento 10 semanas según resumen	Contexto por acceso documental incompleto
Starszak 2026 (19) — RCT de tres grupos	Sin valoración formal completa del sesgo por resultado	Ventaja inmediata; sin significación en seguimiento	Apoyo diferenciado por momento

Nota. Las observaciones se vinculan con los desenlaces descritos y no equivalen a juicios de riesgo de sesgo determinados mediante preguntas de señalización. Contexto identifica los estudios de población no separable o acceso documental restringido.

S5. Confianza interpretativa por desenlace

La confianza interpretativa se consideró según pertinencia de la población, acceso al informe primario, objetividad del desenlace, comparador, precisión, seguimiento y coherencia de los resultados. No se asignaron categorías formales de certeza GRADE.

Tabla S5. Evidencia disponible y límites de interpretación

Desenlace	Informes orientadores	Aspectos favorables y limitaciones
Conocimiento/anatomía	Weeks (14), Pickering (16), Zingg (18), Starszak (19)	Comparación experimental en varias tareas. Medidas inmediatas, intervenciones compuestas y resultados discrepantes: no superioridad general
Habilidad procedimental simulada	Schoeb (21), Guha (22), Hayasaka (24), Hadida Barzilai (23)	OSCE/pruebas técnicas, cegamiento en algunos estudios. Tareas heterogéneas, muestras pequeñas y escenarios simulados: no generalizar a clínica
Retención	Starszak (19); Xiao (20) contextual	Evaluación diferida en algunos diseños. Resultados no uniformes; Xiao (20) solo disponible en resumen: persistencia no establecida
Transferencia clínica	Felten (27)	Resultado realizado en pacientes. 53 procedimientos y parada por futilidad: evidencia insuficiente de superioridad
No inferioridad POCUS	Mulder (28)	Diseño y márgenes concretos para EPSS/TAPSE. Sensibilidad de EPSS a exclusiones: conclusión depende del análisis y del desenlace
Motivación/usabilidad/carga	Noll (29), Malik (30), d'Aiello (31); Xiao (20) contextual	Medidas de experiencia de usuario. No sustituye medidas objetivas; susceptibilidad a efectos de novedad

Desenlace	Informes orientadores	Aspectos favorables y limitaciones
Costes/ escalabilidad	Ningún estudio proporciona evaluación económica integrada comparable	Criterios de implementación planteados. No hay prueba de coste-efectividad o sostenibilidad institucional

Abreviaturas: RA, realidad aumentada; RM, realidad mixta; XR, realidad extendida; OSCE, examen clínico objetivo estructurado; EPSS, separación del punto E al septo; TAPSE, excursión sistólica del plano del anillo tricuspídeo; POCUS, ecografía en el punto de atención.