Revista Española de Educación Médica

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

Beyond technological immersion: augmented and mixed reality in undergraduate medical education. A critical review of comparative studies.

Lauricia Jacqueline Macías Mendoza[1], Yoiler Batista Garcet[1], Lissette Beatriz Villavicencio Cedeño[2]

[1] Universidad San Gregorio de Portoviejo, Carrera de Medicina, Portoviejo, Manabí, Ecuador. ljmacias@sangregorio.edu.ec, https://orcid.org/0000-0002-8779-9091; ybatista@sangregorio.edu.ec, https://orcid.org/0000-0002-7851-5763. [2] Universidad Nacional de Tumbes, Escuela de Posgrado, Doctorado en Estadística Matemática Aplicada, Tumbes, Perú. iso_villa@hotmail.com, https://orcid.org/0000-0001-8475-3947

* Autora para correspondencia: ljmacias@sangregorio.edu.ec

Recibido: 30/9/26; Aceptado: 27/10/26; Publicado: 29/10/26

Resumen

Introducción: La realidad aumentada (RA) y la realidad mixta (RM) permiten integrar información digital con el entorno físico, pero su utilidad educativa depende de los resultados de aprendizaje que alcanzan. **Objetivo:** Analizar críticamente estudios comparativos sobre RA/RM en Medicina de grado, atendiendo a la competencia evaluada y al diseño instruccional. **Métodos:** Revisión crítica con síntesis narrativa de un corpus identificado mediante revisiones previas y búsquedas progresivas en PubMed entre enero y septiembre de 2026, con fecha de corte el 29 de septiembre de 2026. Se examinaron conocimiento, comprensión espacial, desempeño procedimental, retención y transferencia; la experiencia del estudiante se trató como resultado secundario. Se diferenciaron informes directamente pertinentes y evidencia complementaria por población o acceso documental. **Resultados:** Se examinaron 18 informes publicados entre 2017 y 2026 (16 directamente pertinentes y dos de carácter contextual). Los beneficios en determinadas tareas anatómicas y procedimentales coexistieron con comparaciones sin diferencias significativas. Las mejoras inmediatas no se mantuvieron en todos los seguimientos. En POCUS, la no inferioridad de mediciones específicas fue sensible a exclusiones analíticas; en punción lumbar no se demostró mayor éxito clínico. **Conclusiones:** La evidencia examinada sugiere beneficios específicos, sin establecer superioridad general de la RA/RM. La correspondencia entre recurso, actividad y competencia constituye una explicación pedagógica plausible. La cobertura no exhaustiva, las limitaciones de los estudios y la ausencia de una evaluación formal completa de sesgo y certeza restringen las recomendaciones curriculares.

Palabras clave: realidad aumentada; realidad mixta; educación médica; estudiantes de Medicina; anatomía; competencia clínica; revisión crítica.

Abstract

Introduction: Augmented reality (AR) and mixed reality (MR) integrate digital information with the physical environment, but their educational value depends on learning outcomes. **Objective:** To critically examine comparative studies of AR/MR in undergraduate medical education, considering the assessed competency and instructional design. **Methods:** This critical review used narrative

synthesis of a corpus located through prior reviews and progressive PubMed searches carried out from January to September 2026, with a bibliographic cutoff of 29 September 2026. Outcomes included knowledge, spatial understanding, procedural performance, retention and transfer; learner experience was considered secondary. Directly relevant reports were distinguished from supplementary evidence limited by population or document access. **Results:** Eighteen reports published between 2017 and 2026 were examined (16 directly relevant and two contextual). Benefits in selected anatomical and procedural tasks coexisted with comparisons showing no statistically significant differences. Immediate gains were not consistently sustained. In POCUS, non-inferiority for specific measurements was sensitive to analytical exclusions; lumbar puncture training did not demonstrate higher clinical success. **Conclusions:** The evidence examined suggests task-specific benefits without establishing general superiority of AR/MR. Alignment between technology, learning activity and competency is a plausible educational explanation. Non-exhaustive coverage, study limitations and the absence of a complete formal risk-of-bias and certainty assessment restrict curricular recommendations.

Keywords: augmented reality; mixed reality; medical education; medical students; anatomy; clinical competence; critical review.

1. Introducción

La realidad aumentada (RA) superpone información digital sobre el entorno real, mientras que la realidad mixta (RM) permite una integración más interactiva entre objetos virtuales y el espacio físico. En educación médica, ambas tecnologías se han utilizado para representar anatomía tridimensional, guiar procedimientos, entrenar habilidades quirúrgicas y apoyar la toma de decisiones. Su atractivo pedagógico deriva de una posibilidad difícil de reproducir con materiales bidimensionales: mantener simultáneamente el contexto real y añadir capas de información visual, espacial o procedimental. La realidad extendida (XR) se emplea como término paraguas; su denominación no basta para considerar elegible una intervención, que debe incorporar efectivamente RA o RM (1, 2). El interés por estas tecnologías ha crecido con rapidez, pero la evidencia no es uniforme. Un metaanálisis de 2022, basado en 13 estudios y 654 participantes, encontró mejoría en tiempo de desempeño, confianza y satisfacción, pero no demostró ventajas consistentes en conocimiento o habilidades. Una síntesis más reciente, publicada en 2025 e integrada por 29 publicaciones, informó mayor puntuación en habilidades, menor tasa de fracaso y menor tiempo de ejecución, sin una ventaja estadísticamente significativa en conocimiento. La heterogeneidad y el posible sesgo de publicación obligan, no obstante, a interpretar estos efectos con cautela (3, 4).

En anatomía, Williams et al. identificaron 17 estudios comparativos de RA/RM; en 12 predominaron resultados sin diferencias significativas frente a otros métodos. Las síntesis de García-Robles et al. y Moro et al. abarcan tecnologías y poblaciones más amplias, por lo que sus efectos no deben trasladarse automáticamente a la RA/RM en Medicina de grado. Mahmud et al. examinaron aplicaciones para habilidades clínicas desde un diseño de revisión de alcance (5-8). En la Revista Española de Educación Médica, González Salgado et al. abordaron criterios de diseño de entornos virtuales para anatomía y Ampuero Valenzuela et al. analizaron herramientas para la enseñanza de la morfología. Estos trabajos aportan contexto pedagógico, pero no sustituyen la evidencia comparativa específica sobre RA/RM (9, 10).

Esta revisión analiza críticamente los resultados de RA/RM en Medicina de grado y sus condiciones de aplicación. La pregunta central es para qué competencia, frente a qué comparador y con qué diseño instruccional se observan beneficios. La aportación consiste en distinguir adquisición inmediata, retención y transferencia, y en examinar la contribución del conjunto

didáctico que acompaña al recurso. Las relaciones identificadas constituyen explicaciones interpretativas, no pruebas causales de moderadores pedagógicos.

2. Métodos

2.1. Diseño y alcance

Se realizó una revisión crítica de estudios comparativos sobre RA/RM en educación médica de grado, con síntesis narrativa organizada por competencias y momentos de evaluación. La búsqueda bibliográfica se desarrolló progresivamente entre enero y septiembre de 2026 y los informes examinados se publicaron entre 2017 y 2026. Se contrastaron resultados, comparadores y limitaciones, sin estimar un efecto global ni afirmar exhaustividad. Al no conservarse historiales íntegros de búsqueda, filtros y selección independiente, no se acredita una revisión sistemática actualizada; por ello, el trabajo se clasifica como revisión crítica de estudios comparativos. La presentación de la síntesis se orientó por SWiM, sin declarar su aplicación completa (11). Los materiales suplementarios S1, S1A y S2-S5 constituyen una organización retrospectiva de la información conservada, no el registro contemporáneo de un procedimiento sistemático.

2.2. Pregunta de revisión

La pregunta se formuló mediante PICOS: en estudiantes de Medicina de grado (P), ¿las intervenciones educativas basadas en RA o RM (I), comparadas con enseñanza convencional, recursos 2D/3D, simulación estándar u otra metodología activa (C), mejoran conocimiento, comprensión espacial, desempeño procedimental, retención, transferencia, carga cognitiva o resultados educativos relacionados (O), en estudios experimentales o cuasiexperimentales comparativos (S).

2.3. Fuentes y estrategia de búsqueda

Se diferenciaron tres períodos: la ejecución progresiva de la búsqueda bibliográfica (enero-septiembre de 2026), el intervalo de publicación de los informes examinados (2017-2026) y la fase de actualización de publicaciones recientes (2024-2026). La fecha de cierre bibliográfico fue el 29 de septiembre de 2026; no debe confundirse con la fecha de ejecución individual de cada consulta.

La identificación histórica se apoyó en las listas de estudios primarios de las revisiones de Baashar et al. (3), Zhang et al. (4) y Williams et al. (5). Estos trabajos secundarios se utilizaron como vías de localización y contextualización, no como unidades de análisis añadidas a los estudios primarios. La actualización bibliográfica se apoyó en PubMed/MEDLINE, consultado mediante PubMed; PubMed Central y sitios editoriales indexados se emplearon principalmente para recuperar textos completos y verificar información. También se examinaron referencias de las publicaciones pertinentes. El proceso de identificación histórica y la actualización se sometieron a los criterios de pertinencia descritos en el apartado 2.4.

Se conserva una ecuación de PubMed formada por tres bloques unidos mediante AND (tecnología, población y ámbito educativo), con términos alternativos enlazados por OR, búsquedas en campos Title/Abstract y truncamiento en hologram*. El material suplementario S1 reproduce la ecuación de PubMed conservada y distingue las reglas de pertinencia declaradas de los filtros automáticos de búsqueda, cuya aplicación histórica no puede acreditarse. No se conservan los días de ejecución por fuente, los historiales completos de filtros, los archivos de exportación ni los recuentos originales de resultados y duplicados. Por ello, la ecuación puede reutilizarse, pero no cabe reconstruir retrospectivamente una búsqueda exhaustiva ni validar mediante registros primarios los recuentos del diagrama de flujo presentado en una versión anterior. Estas limitaciones sustentan la denominación de revisión crítica adoptada; los recuentos históricos se documentan, sin carácter probatorio, únicamente en S1A.

2.4. Criterios de pertinencia y desenlaces

- Población de interés: estudiantes de Medicina de grado. Los informes con población mixta sin resultados separables se presentan únicamente como contexto educativo; no sustentan conclusiones específicas de eficacia en Medicina.
- Intervención: RA o RM con superposición o integración contextual de contenido digital con el entorno físico. Se distinguió el uso directo por el estudiante de la docencia transmitida mediante un visor utilizado por el profesor. La denominación XR, por sí sola, no establece que exista un componente de RA/RM.
- Comparador: enseñanza convencional, atlas, recursos bidimensionales o tridimensionales, simulación estándar, instrucción docente u otra estrategia educativa activa.
- Desenlaces principales de la síntesis: conocimiento, comprensión espacial y desempeño procedimental medidos objetivamente; retención en una evaluación diferida y transferencia a una tarea o contexto distinto del entrenamiento. Desenlaces secundarios: satisfacción, motivación, confianza, usabilidad y carga cognitiva. Esta jerarquía organiza el análisis y no se presenta como preespecificada en un protocolo. OSCE y OSPE corresponden a evaluaciones clínicas y prácticas objetivas estructuradas, respectivamente.
- Diseño: ensayo aleatorizado, estudio comparativo prospectivo o cuasiexperimental con grupo comparador.
- Delimitación: la realidad virtual sin componente RA/RM, la formación exclusivamente de residentes o profesionales y los trabajos sin comparación no aportaron evidencia directa para responder la pregunta. Las revisiones anteriores y los informes con limitaciones de población o acceso se utilizaron como contexto, diferenciados de las comparaciones directamente pertinentes.

2.5. Selección, extracción y evaluación crítica

La búsqueda y selección de publicaciones se atribuyen a las tres autores, según la declaración de contribuciones. La extracción y curación de datos se atribuyen a MM y VC (dos autoras), y el análisis y la interpretación a los tres autores. Estas cifras describen su participación general, no el número de evaluadores independientes en cada fase. La documentación conservada no permite acreditar cribado por duplicado, extracción independiente, evaluación separada del sesgo, procedimientos de consenso ni arbitraje por un tercer evaluador. No se atribuyen retrospectivamente esos procedimientos al estudio.

Se examinaron títulos y resúmenes de los trabajos potencialmente pertinentes y se contrastó, cuando estuvo disponible, la publicación primaria para identificar población, intervención, comparador, diseño, desenlaces y resultados. La expresión original «o mediante información bibliográfica suficiente» se eliminó por su ambigüedad: cuando no se dispuso del informe completo (por ejemplo, Xiao et al.), la información se restringió a lo comunicado en el resumen original y se presentó como contexto, sin considerarla verificación íntegra de elegibilidad ni soporte suficiente para una inferencia de eficacia. Las muestras mixtas sin análisis separables para estudiantes de Medicina (17) también se distinguen del cuerpo directamente aplicable. Las decisiones individuales de inclusión/exclusión y la resolución de discrepancias no cuentan con una bitácora histórica recuperable; la falta de trazabilidad se declara como limitación del trabajo. El suplemento S1 y su apartado S1A detallan las limitaciones de trazabilidad; S2 presenta fichas críticas individuales y S3 describe los criterios de síntesis. Ninguno acredita procedimientos no registrados.

La evaluación crítica narrativa se centró en la asignación de participantes y la confusión, la adecuación del comparador, las desviaciones y exclusiones analíticas, los datos faltantes, las propiedades de los instrumentos y la selección o temporalidad de los resultados. La matriz individualizada del suplemento S4 presenta información y aspectos no verificables de cada informe, considerando las particularidades del diseño descrito. Estos aspectos guardan relación conceptual

con RoB 2 y ROBINS-I (12, 13), pero no constituyen juicios formales por resultado obtenidos con sus preguntas de señalización y algoritmos. Por esa razón no se atribuyen categorías bajo, moderado ni alto riesgo de sesgo, ni se presenta como completado un procedimiento independiente que no puede acreditarse.

2.6. Estrategia de síntesis

No se efectuó un metaanálisis propio debido a las diferencias en intervenciones, comparadores, instrumentos y momentos de evaluación. La síntesis se organizó por competencia y tiempo de evaluación (adquisición inmediata, retención y transferencia), y se contrastaron explícitamente los resultados favorables, los no significativos y los de no inferioridad. Los efectos publicados en metaanálisis anteriores se aportan como contexto, sin incorporarlos como nuevas observaciones ni sumar nuevamente estudios primarios. No se empleó el recuento de estudios con $p < 0,05$ como estimador de eficacia. La lectura narrativa atribuyó mayor apoyo a resultados objetivos de población pertinente y a informes completos, y menor peso interpretativo a resúmenes, muestras no separables, diseños piloto o hallazgos dependientes de exclusiones. Esta clasificación no es una ponderación cuantitativa ni una evaluación formal de certeza. Se retiró el diagrama de flujo PRISMA de la versión principal porque sus recuentos históricos no son auditables; el anexo S1A conserva exclusivamente su examen retrospectivo.

2.7. Confianza en los hallazgos

La confianza interpretativa se examinó por desenlace, atendiendo a pertinencia de la población, acceso al informe primario, comparador, medición objetiva, coherencia de resultados, precisión, duración del seguimiento y limitaciones de diseño. Los informes completos con desenlaces objetivos y comparadores pertinentes aportaron apoyo directo para las tareas específicas estudiadas; los estudios piloto, con exclusiones analíticas o con evaluaciones exclusivamente inmediatas se interpretaron con mayores reservas. Los informes sustentados solo en resúmenes o en muestras mixtas no separables tuvieron función contextual. La Tabla 2 y el suplemento S5 muestran esta jerarquía cualitativa, que no equivale a GRADE ni permite asignar niveles formales de certeza por resultado o formular recomendaciones graduadas.

3. Resultados

3.1. Informes examinados y características generales

El corpus examinado comprende 18 informes publicados entre 2017 y 2026: 16 comparaciones directamente pertinentes para Medicina de grado y dos informes contextuales (17, 20). Diez de ellos tenían fecha de publicación hasta 2023 y ocho fueron publicados entre 2024 y 2026. Esta partición temporal no equivale a una prueba de la vía individual de localización en las búsquedas históricas o de actualización, que carecen de una exportación íntegra verificable. Las investigaciones se concentraron en anatomía, neuroanatomía, histología y entrenamiento procedimental. La Tabla 1 describe el corpus documental y sus limitaciones, no una selección exhaustiva reconstruida mediante un flujo PRISMA. Zeedzen-Scheffers et al. (17) se considera contextual por incluir Medicina y Ciencias Biomédicas sin resultados separables, y Xiao et al. (20) se identifica como evidencia disponible únicamente en el resumen primario. El suplemento S1-S1A documenta la estrategia y los límites de trazabilidad; S2 contiene fichas críticas de los informes, S3 explica los criterios de síntesis, S4 documenta observaciones metodológicas y S5 presenta la síntesis cualitativa por desenlace.

Tabla 1. Características y resultados de los 18 informes examinados (16 pertinentes y 2 contextuales).

Estudio (Ref)	n	Área	Comparación	Hallazgo principal
Noll (29)	44	Dermatología	RA móvil vs. enseñanza convencional	RA móvil favoreció el aprendizaje autónomo; muestra pequeña, sin evidencia de transferencia clínica.
Stojanovska (15)	64	Anatomía	HoloLens vs. disección cadavérica	Sin diferencias significativas frente a disección; no se demuestra equivalencia.
Schoeb (21)	164	Cateterismo vesical	RM/HoloLens vs. instructor	Mayor puntuación OSCE con RM (21,49 frente a 19,96); evaluación de tarea específica.
Weeks (14)	30	Anatomía cabeza-cuello	RA holográfica vs. pantalla 2D	Mejor postest inmediato con RA (tamaño del efecto: 0,73); retención no evaluada.
Pickering (16)	200*	Neuroana-tomía	RM vs. screencast activo	Ganancias similares; algunas medidas favorecieron al screencast. Denominadores variables.
Farshad-Amacker (25)	44	Ecografía	RA-US vs. US estándar	No hubo ventaja uniforme de RA; posible influencia del orden en el diseño cruzado.
Guha (22)	36	Habilidades quirúrgicas	HoloLens vs. vídeo/enseñanza tradicional	Señales de mejora técnica con RM; estudio piloto con evaluadores cegados.
Hayasaka (24)	30	Anestesia epidural	RA/RM vs. técnica convencional	Mayor precisión en simulador epidural; tres grupos de diez participantes.
Malik (30)	56	Otorrinola-ringología	Transmisión con HoloLens 2 del docente vs. caso remoto convencional	Mayor ganancia en conocimiento (12 frente a 9 puntos); exclusiones posteriores.
D'Aiello (31)	59	Anatomía cardíaca	RM vs. vídeos holográficos vs. diapositivas	Efectos dependientes del contenido anatómico; comparación exploratoria.
Zeedzen-Scheffers (17)	28*	Neuroana-tomía	RA vs. atlas anatómico	Sin diferencia significativa en conocimiento; muestra mixta no separable (contextual).
Sarkar (26)	34	Punción lumbar	RA + simulación vs. simulación estándar	Sin diferencias técnicas significativas; mayor confianza autoinformada.
Hadida Barzilai (23)	21	Mastoidecto-mía	RA guiada vs. entrenamiento tradicional	Mayor puntuación de fresado (19,5/25 frente a 12/25); tarea simulada.
Zingg (18)	48	Anatomía	RA holográfica adaptativa vs. recursos tradicionales	Mejores resultados teóricos y prácticos; intervención multicomponente.
Felten (27)	188*	Punción lumbar clínica	RA + háptica vs. enseñanza a pie de cama	Sin superioridad en éxito clínico; 53 procedimientos analizados por protocolo.

Mulder (28)	56	POCUS cardíaco	XR/RA vs. instructor	No inferioridad en mediciones específicas; EPSS sensible a exclusiones.
Xiao (20)	152	Histología	RA vs. microscopía tradicional	Mejor retención y OSPE según el resumen; información contextual no verificada íntegramente.
Starszak (19)	236	Anatomía	RM vs. impresión 3D vs. control	Mejora inmediata con RM, sin diferencia significativa en retención diferida.

Notas. Pickering: el texto declara 200 estudiantes con pretest y postest, pero la tabla de resultados presenta 127 + 75 para preguntas de elección múltiple y 84 + 62 = 146 para respuestas cortas y puntuación global; se conserva esta discrepancia del informe primario sin inferir un denominador único. Hayasaka: 30 participantes, tres grupos de 10. d'Aiello: 59 participantes, grupos de 20, 20 y 19. Felten: 188 estudiantes participantes y 53 procedimientos analizados por protocolo, unidades que no son intercambiables. No se calcula una muestra total agregada. Zeedzen-Scheffers (17) y Xiao (20) son únicamente contextuales y no se utilizan para sostener inferencias específicas de eficacia en Medicina de grado.

3.2. Anatomía, histología y comprensión espacial

Los resultados de anatomía no fueron uniformes. Weeks et al. comunicaron mejores resultados inmediatos con representación holográfica que con pantalla 2D. Stojanovska et al. no observaron diferencias significativas en las evaluaciones prácticas comparadas; su diseño combinó cohortes y modalidades de evaluación. Pickering et al. emplearon un diseño cuasialeatorizado y encontraron ventajas del screencast en preguntas de elección múltiple, sin diferencias significativas en otras comparaciones (14-16). Como contexto, Zeedzen-Scheffers et al. estudiaron 28 participantes de Medicina y Ciencias Biomédicas: no observaron diferencias significativas en la ganancia de conocimiento. La muestra no permite una inferencia separada para Medicina (17). En TEACHANATOMY, 48 estudiantes fueron asignados a RA holográfica adaptativa con gamificación o aprendizaje tradicional. Las medianas fueron 18,8 frente a 9,4 en la prueba teórica y 14,0 frente a 5,0 en la práctica ($p < 0,001$). El contraste evalúa un conjunto de componentes educativos, no el efecto aislado del dispositivo (18) Starszak et al. incluyeron 236 estudiantes: la RM obtuvo una ventaja inmediata frente al control, pero no se observaron diferencias significativas en retención diferida (19) El resumen de Xiao et al. describe resultados favorables de RA en histología en 152 estudiantes, con seguimiento a 10 semanas; al no disponer del texto completo, se considera información complementaria y no se asume asignación aleatoria (20)

3.3. Habilidades procedimentales y quirúrgicas

Schoeb et al. comunicaron una mayor puntuación OSCE para cateterismo vesical con RM (21). Guha et al., en 36 estudiantes y con evaluadores cegados, observaron una mayor ganancia en destreza para arteriotomía y cierre frente a un tutorial en vídeo (22). En mastoidectomía, Hadida Barzilai et al. encontraron puntuaciones de fresado de 19,5/25 frente a 12/25 ($p < 0,01$), evaluadas por dos otólogos cegados (23). Hayasaka et al. estudiaron 30 participantes y observaron mejoras de precisión en un simulador de anestesia epidural (24). Estas tareas evalúan ejecución en condiciones específicas y no demuestran por sí mismas transferencia a pacientes. En el estudio cruzado de Farshad-Amacker et al., el tiempo de punción con ecografía aumentada no difirió significativamente del convencional (25). En el estudio piloto de Sarkar et al. no se identificaron diferencias significativas en los desenlaces técnicos principales de punción lumbar, pese a una experiencia subjetiva favorable (26). Felten et al. examinaron transferencia clínica: de 188 estudiantes participantes, 53 procedimientos fueron analizados por protocolo. El éxito fue 13/28 (46,4%) frente a 10/25 (40,0%); el apartado de resultados informa $p = 0,85$. El ensayo se detuvo anticipadamente por futilidad. Estos hallazgos no establecen equivalencia ni descartan diferencias clínicamente relevantes con precisión suficiente (27). Mulder et al. incluyeron 56 estudiantes y

compararon entrenamiento con HoloLens 2 frente a instrucción convencional para ecografía cardíaca a pie de cama (POCUS). Informaron no inferioridad para la separación del punto E al septo (EPSS) y la excursión sistólica del plano del anillo tricuspídeo (TAPSE), con márgenes de 3 y 4 mm, respectivamente. Sin embargo, excluyeron dos mediciones por desenlace; para EPSS, la inclusión de los valores extremos del grupo XR impedía confirmar estadísticamente la no inferioridad. La interpretación debe limitarse a estos desenlaces y análisis. El diseño no constituye una evaluación de coste-efectividad ni demuestra escalabilidad institucional (28).

3.4. Conocimiento, retención y transferencia

Los resultados de conocimiento declarativo fueron heterogéneos. Como contexto externo, Zhang et al. comunicaron una diferencia media ponderada de 2,92 puntos (IC95% −1,73 a 7,57), sin evidencia estadística de superioridad. El intervalo incluye ausencia de efecto y posibles beneficios; no demuestra equivalencia. Su resultado agregado de habilidades presentó heterogeneidad elevada (I^2 = 95,8%), lo que limita la interpretación de una magnitud única entre tareas distintas (4). La evaluación longitudinal fue menos abundante que la inmediata. Starszak et al. no encontraron diferencias significativas en el seguimiento (19). El resultado favorable a las 10 semanas comunicado en el resumen de Xiao et al. constituye un indicio complementario, insuficiente para establecer la persistencia general del beneficio (20). Felten et al. evaluaron procedimientos en pacientes, pero el número de procedimientos y la interrupción por futilidad limitan la precisión (27). Las inferencias sobre retención y transferencia deben distinguir desenlace, seguimiento y unidad analizada; no se extrapolan desde un postest inmediato.

3.5. Carga cognitiva, motivación y experiencia del estudiante

La experiencia del estudiante fue frecuentemente favorable, aunque no de forma uniforme. Noll et al. evaluaron RA móvil en aprendizaje autónomo combinado (29). Malik et al. analizaron 56 estudiantes: la ganancia media en un cuestionario de 20 puntos fue de 12 frente a 9 puntos. Los visores los usaban los docentes y los alumnos recibían la transmisión por Microsoft Teams; se formularon más preguntas, pero no aumentó significativamente la participación global medida. Las exclusiones posteriores a la asignación y el desequilibrio entre grupos limitan la interpretación (30). d'Aiello et al. evaluaron 59 estudiantes en tres modalidades y encontraron efectos dependientes del contenido (31). La aceptación no sustituye a la demostración de competencia clínica.

3.6. Limitaciones metodológicas de los estudios

Las limitaciones se analizaron estudio por estudio y se vincularon al desenlace interpretado. La asignación cuasialeatoria de Pickering (16), la variación de denominadores y las exclusiones posteriores a la asignación en Malik (30) limitan la comparabilidad interna; el diseño cruzado de Farshad-Amacker (25) requiere atender al orden y al aprendizaje previo. El cegamiento de evaluadores comunicado por Guha (22) y Hadida Barzilai (23) refuerza la medición de determinadas tareas simuladas, aunque no su generalización clínica. En Mulder (28), la inferencia sobre EPSS es sensible a la exclusión de valores; en Felten (27), la interrupción por futilidad y los 53 procedimientos analizados restringen la precisión sobre transferencia. Los estudios basados solo en resúmenes no permiten comprobar todos los dominios pertinentes. La matriz S4 diferencia limitaciones observables de información no verificada, sin convertirlas en clasificaciones formales RoB 2/ROBINS-I. El tamaño reducido incide sobre todo en precisión y la ausencia de cegamiento del estudiante no demuestra automáticamente un riesgo alto en todos los desenlaces.

Tabla 2. Alcance interpretativo de la evidencia por desenlace.

Desenlace	Patrón observado	Restricción principal	Inferencia admisible
Conocimiento y comprensión espacial	Mejoras en tareas concretas y comparaciones sin diferencias.	Comparadores e instrumentos heterogéneos; asignación y datos incompletos en algunos estudios.	No se demuestra superioridad general.
Desempeño procedimental	Beneficios en cateterismo, arteriotomía, epidural o mastoidectomía; resultados variables en otras tareas.	Muestras pequeñas y tareas simuladas; distinta asistencia durante la prueba.	No extrapolar a competencia clínica global.
Retención	Seguimiento anatómico sin diferencias; señal favorable en histología comunicada solo en resumen.	Pocos seguimientos, tiempos distintos y acceso documental limitado en histología.	Persistencia del efecto incierta.
Transferencia y POCUS	Punción lumbar sin superioridad; POCUS con no inferioridad sensible a exclusiones.	Unidades analíticas distintas; futilidad y sensibilidad a valores excluidos.	Conclusiones restringidas a cada resultado.
Experiencia y costes	Aceptación frecuentemente favorable; sin síntesis económica propia.	Autoinforme y posible novedad; costes no evaluados de forma comparable.	No inferir eficacia, ahorro o sostenibilidad.

4. Discusión

La síntesis muestra resultados favorables en determinadas tareas de anatomía y procedimientos, junto con comparaciones sin ventajas significativas. Su interpretación más prudente es que el valor de RA/RM depende del problema educativo y de las condiciones de aplicación. La alineación entre visualización, actividad y competencia es una explicación plausible; no fue sometida aquí a una prueba causal ni a una metarregresión propia. Las diferencias entre los resultados favorables y los que no muestran ventajas estadísticamente significativas pueden depender del comparador (instrucción convencional frente a recursos activos), la intensidad y duración de la actividad, la experiencia previa del estudiante, la estructura de la retroalimentación y la proximidad entre la tarea practicada y el instrumento de evaluación. TEACHANATOMY (18) integra adaptación y gamificación en la intervención; Guha (22) incorpora práctica procedimental guiada y evaluación cegada; Pickering (16) utiliza un comparador activo integrado en el currículo. No se dispone de un análisis de moderadores propio que separe el efecto del dispositivo de estos componentes. La mejora durante una tarea guiada tampoco equivale a la competencia mantenida sin ayuda o transferida al ámbito clínico. Estas explicaciones son hipótesis didácticas, no mecanismos causales demostrados por la síntesis (16, 18, 22).

El efecto sobre la carga cognitiva constituye otra explicación posible. La información integrada en el campo de acción podría facilitar orientación y coordinación; una interfaz compleja también podría añadir exigencias ajenas al objetivo de aprendizaje. Estas posibilidades requieren medidas válidas y diseños que separen contenido, tiempo de exposición, retroalimentación y dispositivo. El resumen de Xiao et al. aporta un indicio en histología, pero su acceso documental limitado impide verificar los mecanismos y no permite generalizarlos a otras intervenciones (20).

Tang et al. ya señalaron en 2020 limitaciones de calidad, validación y factibilidad en la investigación sobre RA (32). La presente síntesis añade un contraste entre adquisición inmediata, retención y transferencia en un corpus más reciente de educación de grado. Su aportación es interpretativa y debe situarse dentro de sus límites de búsqueda y verificación. Las evaluaciones futuras deberían seleccionar instrumentos coherentes con la competencia, justificar sus propiedades de medición, informar intervalos de confianza y programar tareas de transferencia y seguimiento.

Las implicaciones institucionales requieren distinguir resultados educativos de resultados de implementación. La no inferioridad en dos mediciones de POCUS no prueba ahorro, sostenibilidad ni sustitución general del profesorado (28). Las revisiones de Betancourt Masri et al. y Zelada Espejo et al., publicadas en esta revista, aportan contexto sobre enseñanza tecnológica y modelos tridimensionales, pero no establecen coste-efectividad de la RA/RM (33-34). La decisión de adopción requiere costes de adquisición, mantenimiento, licencias, formación, accesibilidad y tiempo docente, junto con resultados educativos mantenidos.

La incorporación de recursos digitales debe responder a objetivos educativos y contextos concretos. Un estudio cuasiexperimental en Bioquímica Médica comunicó mejoras académicas con aprendizaje inteligente y rutinas del pensamiento (35), mientras que una reflexión sobre virtualidad en América Latina destacó oportunidades de acceso y flexibilidad (36). Ninguno de estos trabajos evaluó RA/RM, por lo que se citan como antecedentes pedagógicos y no como pruebas de su eficacia.

La cobertura bibliográfica no es exhaustiva: se concentró en PubMed/MEDLINE y las referencias de revisiones anteriores, sin ejecutar de forma documentada búsquedas equivalentes en bases educativas y bibliográficas comerciales. No se conservan las fechas diarias, exportaciones y decisiones completas de cribado, ni puede acreditarse revisión independiente por duplicado o consenso formal. Tampoco se completaron evaluaciones RoB 2/ROBINS-I por resultado ni una valoración GRADE de certeza. Estos límites impiden tratar la síntesis como revisión sistemática exhaustiva y restringen cualquier recomendación curricular. La evidencia procedente de resúmenes o poblaciones mixtas se identifica como complementaria. La heterogeneidad, el seguimiento limitado, las muestras pequeñas y los denominadores variables restringen además la precisión y la transferencia de las inferencias.

4.1. Implicaciones para la decisión curricular

La tabla 3 presenta una lista de criterios que se pueden usar para la decisión curricular, para planificar y evaluar implementaciones. Es una elaboración interpretativa, no un conjunto de efectos demostrados ni una recomendación graduada.

Tabla 3. Criterios para la decisión curricular.

Pregunta curricular	Criterio recomendado	Indicador de calidad	Riesgo que se evita
¿Qué competencia se pretende desarrollar?	Definir conocimiento, habilidad o desempeño observable.	Objetivo y estándar explícitos.	Confundir exposición tecnológica con aprendizaje.
¿Qué aporta específicamente la RA/RM?	Usarla cuando añade visualización, guía contextual o interacción relevante.	Justificación funcional del recurso.	Tecnocentrismo.
¿Cómo se integrará?	Combinar con práctica activa, feedback y guía docente cuando sea necesaria.	Secuencia reproducible.	Uso aislado o demostrativo.
¿Cómo se evaluará?	Priorizar pruebas objetivas, OSCE/OSPE, precisión, errores y	Instrumentos con evidencia de validez.	Confundir satisfacción con competencia.

	tiempo.		
¿Se mantiene y transfiere?	Incluir evaluación diferida y tareas nuevas o clínicas.	Retención y transferencia.	Sobreestimar el postest inmediato.
¿Es sostenible?	Analizar dispositivos, soporte técnico, tiempo docente y escalabilidad.	Costo total y frecuencia de uso.	Adquisición costosa e infrautilizada.

5. Conclusiones

- Los estudios examinados sugieren que la RA/RM puede beneficiar tareas educativas concretas, especialmente cuando incorpora información espacial o guía procedimental relevante. Los resultados de conocimiento y seguimiento son heterogéneos; la ausencia de diferencias significativas no demuestra equivalencia. La evidencia de transferencia clínica es limitada y la no inferioridad observada en POCUS debe interpretarse conforme a sus desenlaces, márgenes y exclusiones analíticas.
- La incorporación curricular puede explorarse mediante evaluaciones locales con objetivos explícitos, comparadores adecuados y medición de desempeño, retención, transferencia y costes. Estas propuestas orientan investigación e implementación; no representan beneficios institucionales demostrados. La satisfacción y la confianza son resultados complementarios. Las conclusiones deben circunscribirse a las tareas y condiciones estudiadas, evitando que la novedad tecnológica se interprete como evidencia suficiente de eficacia educativa.

Contribuciones de autoría: Conceptualización: LJMM y YBG. Metodología: YBG y LBVC. Búsqueda y selección de estudios: LJMM, YBG y LBVC. Extracción y curación de datos: LJMM y LBVC. Análisis, síntesis e interpretación de la evidencia: LJMM, YBG y LBVC. Redacción del borrador original: LJMM y YBG. Revisión crítica y edición del manuscrito: LJMM, YBG y LBVC. Supervisión: YBG.

Financiación: La investigación no recibió financiación.

Conflictos de interés: Los autores declaran no tener conflictos de interés relacionados con este trabajo.

Consideraciones éticas: La revisión utilizó exclusivamente información publicada y no incluyó datos individuales identificables; por ello no requirió consentimiento informado ni aprobación de un comité de ética.

Declaración sobre el uso de inteligencia artificial: En la fase bibliográfica se utilizó Perplexity AI como apoyo para identificar y revisar publicaciones potencialmente pertinentes. Esta herramienta no sustituye el juicio científico de los autores ni generaró datos primarios. La selección de los estudios, la comprobación final de fuentes y resultados, la interpretación de la evidencia y la aprobación del manuscrito son responsabilidad de los autores.

Material suplementario: estrategia bibliográfica conservada, caracterización crítica de los informes examinados y límites de las inferencias por desenlace.

6. Referencias

1. Barteit S, Lanfermann L, Bärnighausen T, et al. Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review. *JMIR Serious Games*. **2021**, 9, e29080. https://doi.org/10.2196/29080
2. Dhar P, Rocks T, Samarasinghe RM, et al. Augmented reality in medical education: students' experiences and learning outcomes. *Med Educ Online.* **2021**, 26, 1953953. https://doi.org/10.1080/10872981.2021.1953953
3. Baashar Y, Alkawsi G, Wan Ahmad WNW, et al. Effectiveness of Using Augmented Reality for Training in the Medical Professions: Meta-analysis. *JMIR Serious Games.* **2022**, 10, e32715. https://doi.org/10.2196/32715
4. Zhang R, Jin X, Liu M, et al. The effectiveness of augmented reality/mixed reality in medical education: a meta-analysis. *BMC Med Educ.* **2025**, 25, 1586. https://doi.org/10.1186/s12909-025-08166-8

5. Williams A, Sun Z, Vaccarezza M. Comparison of augmented reality with other teaching methods in learning anatomy: A systematic review. *Clin Anat.* **2025**, 38, 168-185. https://doi.org/10.1002/ca.24234
6. García-Robles P, Cortés-Pérez I, Nieto-Escámez FA, et al. Immersive virtual reality and augmented reality in anatomy education: A systematic review and meta-analysis. *Anat Sci Educ.* **2024**, 17, 514-528. https://doi.org/10.1002/ase.2397
7. Moro C, Birt J, Stromberga Z, et al. Virtual and Augmented Reality Enhancements to Medical and Science Student Physiology and Anatomy Test Performance: A Systematic Review and Meta-Analysis. *Anat Sci Educ.* **2021**, 14, 368-376. https://doi.org/10.1002/ase.2049
8. Mahmud M, Sari DCR, Sari D, et al. The application of augmented reality for improving clinical skills: a scoping review. *Korean J Med Educ.* **2024**, 36, 65-79. https://doi.org/10.3946/kjme.2024.285
9. González ST, Ibeas C, Gravert I, et al. Características de un escenario de Realidad Virtual para el aprendizaje de anatomía: Una revisión bibliográfica. *Rev Esp Edu Med.* **2022**, 3, 81-92. https://doi.org/10.6018/edumed.542861
10. Ampuero IP, Riquelme-Vilches V, Vargas E, et al. Influencia de las Herramientas Tecnológicas en la Enseñanza de la Morfología en la Educación Médica: Una Revisión Sistemática. *Rev Esp Edu Med.* **2024**, 5, 631071. https://doi.org/10.6018/edumed.631071
11. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ.* **2020**, 368, l6890. https://doi.org/10.1136/bmj.l6890
12. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ.* **2019**, 366, l4898. https://eprints.whiterose.ac.uk/id/eprint/150579/
13. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ.* **2016**, 355, i4919. https://doi.org/10.1136/bmj.i4919
14. Weeks JK, Pakpoor J, Park BJ, et al. Harnessing Augmented Reality and CT to Teach First-Year Medical Students Head and Neck Anatomy. *Acad Radiol.* **2021**, 28, 871-876. https://doi.org/10.1016/j.acra.2020.07.008
15. Stojanovska M, Tingle G, Tan L, et al. Mixed Reality Anatomy Using Microsoft HoloLens and Cadaveric Dissection: A Comparative Effectiveness Study. *Med Sci Educ.* **2020**, 30, 173-178. https://doi.org/10.1007/s40670-019-00834-x
16. Pickering JD, Panagiotis A, Ntakakis G, et al. Assessing the difference in learning gain between a mixed reality application and drawing screencasts in neuroanatomy. *Anat Sci Educ.* **2022**, 15, 628-635. https://doi.org/10.1002/ase.2113
17. Zeedzen I, Karstens J, van den Hurk M, et al. Comparing the effectiveness of augmented reality and anatomical atlases in student preparation for neuroanatomy dissection. *Sci Rep.* **2024**, 14, 24939. https://doi.org/10.1038/s41598-024-76379-w
18. Zingg L, Häusler M, Hein J, et al. Using Hologram-Based Augmented Reality in Anatomy Learning: The TEACHANATOMY Randomized Trial. *Acad Med.* **2025**, 100, 695-704. https://doi.org/10.1097/ACM.0000000000006012
19. Starszak K, Karaś R, Bajor G, et al. Advancing anatomical education: comparative effectiveness of mixed reality and 3D printing technologies. *BMC Med Educ.* **2026**, 26, 734. https://doi.org/10.1186/s12909-026-08959-5
20. Xiao R, Chen X, Yang X, et al. Enhancing Histological Learning Through Augmented Reality: A Comparative Study of Traditional Microscopy and AR-Based Instruction. *Clin Anat.* **2026**, publicación anticipada en línea. https://doi.org/10.1002/ca.70139
21. Schoeb DS, Schwarz J, Hein S, et al. Mixed reality for teaching catheter placement to medical students: a randomized single-blinded, prospective trial. *BMC Med Educ.* **2020**, 20, 510. https://doi.org/10.1186/s12909-020-02450-5

22. Guha P, Lawson J, Minty I, et al. Can mixed reality technologies teach surgical skills better than traditional methods? A prospective randomised feasibility study. *BMC Med Educ.* **2023**, 23, 144. https://doi.org/10.1186/s12909-023-04122-6
23. Hadida Barzilai D, Tejman-Yarden S, Yogev D, et al. Augmented Reality-Guided Mastoidectomy Simulation: A Randomized Controlled Trial Assessing Surgical Proficiency. *Laryngoscope.* **2025**, 135, 894-900. https://doi.org/10.1002/lary.31791
24. Hayasaka T, Kawano K, Onodera Y, et al. Comparison of accuracy between augmented reality/mixed reality techniques and conventional techniques for epidural anesthesia using a practice phantom model kit. *BMC Anesthesiol.* **2023**, 23, 171. https://doi.org/10.1186/s12871-023-02133-w
25. Farshad NA, Kubik RA, Kolling C, et al. Learning how to perform ultrasound-guided interventions with and without augmented reality visualization: a randomized study. *Eur Radiol.* **2023**, 33, 2927-2934. https://doi.org/10.1007/s00330-022-09220-5
26. Sarkar K, Coote J, Abdul R, et al. Feasibility, Adoption, and Utility of Augmented Reality in Lumbar Puncture Simulation by Medical Students: A Pilot Study. *J Med Ext Real.* **2024**, 1, 191-202. https://doi.org/10.1089/jmxr.2024.0001
27. Felten R, Bigaut K, Wirth T, et al. Advancing medical training with augmented reality and haptic feedback simulator: outcomes of a randomized controlled trial on lumbar puncture. *BMC Med Educ.* **2025**, 25, 1231. https://doi.org/10.1186/s12909-025-07536-6
28. Mulder TA, Jacobs RC, Zuurmond MS, et al. Augmedicine: Ultrasound: Training in ultrasound using XR – a randomized controlled trial. *BMC Med Educ.* **2026**, 26, 778. https://doi.org/10.1186/s12909-026-08897-2
29. Noll C, von Jan U, Raap U, et al. Mobile Augmented Reality as a Feature for Self-Oriented, Blended Learning in Medicine: Randomized Controlled Trial. *JMIR Mhealth Uhealth.* **2017**, 5, e139. https://doi.org/10.2196/mhealth.7943
30. Malik R, Abbas JR, Jayarajah C, et al. Mixed Reality Enhanced Otolaryngology Case-Based Learning: A Randomized Educational Study. *Laryngoscope.* **2023**, 133, 1606-1613. https://doi.org/10.1002/lary.30364
31. d'Aiello AF, Cabitza F, Natali C, et al. The Effect of Holographic Heart Models and Mixed Reality for Anatomy Learning in Congenital Heart Disease: An Exploratory Study. *J Med Syst.* **2023**, 47, 64. https://doi.org/10.1007/s10916-023-01959-8
32. Tang KS, Cheng DL, Mi E, et al. Augmented reality in medical education: a systematic review. *Can Med Educ J.* **2020**, 11, e81-e96. https://doi.org/10.36834/cmej.61705
33. Betancourt Masri L, Bello Astorga W, Osses Alvarado I, et al. Influencia del uso de nuevas tecnologías en la educación de semiología médica. Revisión sistemática. *Rev Esp Edu Med.* **2024**, 5, 631521. https://doi.org/10.6018/edumed.631521
34. Zelada Espejo MI, Segura Beltrán F, Margas Cavieres F, et al. Impacto de la impresión 3D en el aprendizaje de estudiantes de medicina: una revisión sistemática. *Rev Esp Edu Med.* **2024**, 5, 626811. https://doi.org/10.6018/edumed.626811
35. Batista GY, Villavicencio Cedeño LB. Uso de aprendizaje inteligente y rutinas del pensamiento en la enseñanza de la Bioquímica Médica. Educ Med Super. **2025**, 39, e4667. https://ems.sld.cu/index.php/ems/article/view/4667.
36. Batista GY, Villavicencio Cedeño LB. Virtualidad en educación médica: oportunidades para América Latina. *Investig Educ Med.* **2026**, 15(57), 132-133. https://doi.org/10.22201/fm.20075057e.2026.57.25747

© 2026 Universidad de Murcia. Enviado para su publicación en acceso abierto bajo los términos y condiciones de la licencia Creative Commons Reconocimiento-NoComercial-Sin Obra Derivada 4.0 España (CC BY-NC-ND) (http://creativecommons.org/licenses/by/4.0/).