JOURNAL OF EDUCATION, RESEARCH AND INNOVATION

TIPO: Revisión Sistemática
PUBLICADO: 14 de Octubre, 2026
VOLUMEN: 1
NÚMERO: 1
DOI: https://doi.org/10.5281/zenodo.23377939

Tutoría generativa y aprendizaje en Educación Superior: Una revisión sistemática

Generative tutoring and learning in Higher Education: A systematic review

Tutoria generativa e aprendizagem em Educação Superior: Uma revisão sistemática

Olvic Lucia Tabares Jaramillo[1]
Mariana Noemí Medina Suarez[2]
Shirley Paola Catuto Solano[3]
David Eduardo González Naranjo[4]

[1]Doctora en Educación, graduada de la Universidad de Caldas, (Colombia). Doctora en Pensamiento Complejo, graduada de la Multiversidad Mundo Real Edgar Morin, (México). Especialista en Gerencia Educativa con énfasis en Gestión de Proyectos, graduada de la Universidad Católica de Manizales, (Colombia). Magíster en Educación, graduada de la Universidad Católica de Manizales, (Colombia). Licenciada en Educación Ambiental, graduada de la Universidad de Caldas, (Colombia).
[2]Magíster en Diseño y Evaluación Educativa graduada de la Universidad Tecnológica Empresarial de Guayaquil, (Ecuador). Doctora en Educación graduada de la Universidad Andrés Bello, (Venezuela).
[3]Profesora de Educación Básica graduada de la Universidad Estatal Península de Santa Elena, (Ecuador). Licenciada en Educación Básica graduada de la Universidad Estatal Península de Santa Elena, (Ecuador). Maestra en Educación Superior graduada de la Universidad San Ignacio de Loyola, (Perú). Doctora en Educación graduada de la Universidad Católica Andrés Bello, (Venezuela).
[4]Licenciatura en Enfermería. Máster Universitario en Dirección y Gestión de Unidades de Enfermería por la UNIR, (España). PhD. en Enfermería por la Universidad Nacional Mayor de San Marcos en Lima, (Perú).

Resumen

El objetivo de la investigación se centró en sintetizar los resultados de la tutoría, el andamiaje y la retroalimentación generativos en educación superior, considerando la mediación docente y la distinción entre desempeño inmediato, retención y transferencia. Se realizó una revisión sistemática con síntesis narrativa. La búsqueda en Scopus, combinó términos de tecnologías generativas, educación superior y apoyo al aprendizaje. Se incluyeron artículos publicados después de 2021, de acceso abierto, en inglés, español o portugués. El proceso de selección siguió PRISMA 2020 y la evaluación metodológica utilizó MMAT 2018. Resultados. De 1519 registros se evaluaron 274 informes completos y se incluyeron 134 estudios; todos fueron valorados con MMAT y conservados en la síntesis. Los artículos abarcaron 78 fuentes y se concentraron en 2025 y 2026. La integración temática comprendió escritura, otras competencias lingüísticas, ciencias y programación, salud, aprendizaje transversal y educación física. Se observaron beneficios específicos de intervenciones con práctica guiada, retroalimentación utilizable y participación reflexiva. También se identificaron resultados nulos, ventajas de la enseñanza humana y efectos inmediatos que no persistieron. Las mejoras dependieron del comparador, la tarea y el conocimiento previo. La tutoría generativa puede complementar la enseñanza universitaria cuando preserva el juicio del estudiante y se integra en una secuencia pedagógica explícita. Su valor debe comprobarse mediante actuaciones independientes y evaluaciones diferidas, sin equiparar la calidad del producto asistido con aprendizaje duradero.

Abstract

The objective of the investigation is centered on synthesizing the results of tutoring, and support and generative feedback in higher education, considering teaching mediation and the distinction between immediate performance, retention and transfer.. A systematic review with narrative synthesis was carried out. The search in Scopus, combined terms of generative technologies, higher education and learning support. Articles published after 2021 are included, with open access, in English, Spanish or Portuguese. The selection process followed PRISMA 2020 and the methodological evaluation used MMAT 2018. Results. Of 1519 records, 274 complete reports were evaluated and 134 studies were included; all were valued with MMAT and preserved in synthesis. The articles encompassed 78 sources and were concentrated in 2025 and 2026. Thematic integration comprised writing, other linguistic skills, science and programming, health, transversal learning and physical education. Specific benefits were observed from interventions with guided practice, usable feedback and reflective participation. Null results, advantages of human learning and immediate effects that did not persist were also identified. The best depends on the comparator, the task and previous knowledge. Generative tutoring can complement university teaching when it preserves the student's judgment and is integrated into an explicit pedagogical sequence. Its value must be proven through independent actions and deferred evaluations, without equating the quality of the assisted product with long-term learning.

Resumo

O objetivo da investigação se concentrou em sintetizar os resultados da tutoria, o andamismo e a retroalimentação generativos na educação superior, considerando a mediação docente e a distinção entre desenvolvimento imediato, retenção e transferência.. Foi realizada uma revisão sistemática com síntese narrativa. A pesquisa na Scopus, combinou termos de tecnologias generativas, educação superior e apoio à aprendizagem. Foram incluídos artigos publicados após 2021, de acesso aberto, em inglês, espanhol ou português. O processo de seleção seguiu PRISMA 2020 e a avaliação metodológica utilizou MMAT 2018. Resultados. Dos 1519 registros foram avaliados 274 relatórios completos e incluídos 134 estudos; todos os fortes valorados com MMAT e conservados na síntese. Os artigos abrangeram 78 fontes e se concentraram em 2025 e 2026. A integração temática compreendia escritura, outras competências linguísticas, ciências e programação, saúde, aprendizagem transversal e educação física. Observe os benefícios específicos de intervenções com prática guiada, retroalimentação utilizável e participação reflexiva. Também se identificam resultados nulos, vendas de conhecimento humano e efeitos imediatos que não persistem. As melhores coisas dependiam da comparação, da taxa e do conhecimento anterior. A tutoria generativa pode complementar o ensino universitário, preservando o suco do estudante e se integrando a uma sequência pedagógica explícita. Seu valor deve ser comparado por meio de atuações independentes e avaliações de feridas, sem equiparar a qualidade do produto assistido com aprendizagem duradoura.

PALABRAS CLAVES: *Evaluación formativa; Autorregulación; Retroalimentación; Transferencia del aprendizaje; Aprendizaje autónomo.*
KEYWORDS: *Formative assessment; Self-regulation; Feedback; Transfer of learning; Autonomous learning.*
PALAVRAS-CHAVE: *Avaliação formativa; Autorregulação; Retroalimentação; Transferência de aprendizagem; Aprendizagem autônoma.*

OPEN ACCESS

CORRESPONDENCIA
Olvic Lucia Tabares Jaramillo
ienormalsagradocorazonriosucio@sedcaldas.gov.co

[1]Secretaría de Educación de Caldas, (Colombia).
[1]ienormalsagradocorazonriosucio@sedcaldas.gov.co
[2]Universidad Estatal Península de Santa Elena, (Ecuador).
[2]mmedina2614@upse.edu.ec
[3]Universidad Estatal Península de Santa Elena, (Ecuador).
[3]scatuto0593@upse.edu.ec
[4]Universidad Iberoamericana del Ecuador, (Ecuador).
[4]dgonzalez@unibe.edu.ec

RECIBIDO: 12 de Agosto, 2026
REVISADO: 25 de Agosto, 2026
APROBADO: 15 de Septiembre, 2026
PUBLICADO: 14 de Octubre, 2026

CITACIÓN SUGERIDA

Tabares, O., Medina, M., Catuto, S. & González, D. (2026). Tutoría generativa y aprendizaje en Educación Superior: Una revisión sistemática. *CENIT Internacional Journal of Education, Research and Innovation.* 1(1). 54 - 77. https://doi.org/10.5281/zenodo.23377939

COPYRIGHT

Este artículo se publica en acceso abierto bajo los términos de la licencia CC BY-NC 4.0 Atribución/ Reconocimiento-NoComercial 4.0 Internacional. Los contenidos publicados en esta revista pueden copiarse, compartirse y adaptarse con fines no comerciales, siempre que se reconozcan debidamente la autoría y la fuente original, se proporcione un enlace a la licencia y se indiquen las modificaciones realizadas. Su utilización no debe sugerir el respaldo de los autores o de la revista a las adaptaciones o usos derivados. Estas libertades son irrevocables mientras se cumplan los términos de la licencia. Asimismo, no podrán imponerse restricciones legales ni medidas tecnológicas que limiten los derechos de uso concedidos.

Introducción

La tutoría universitaria exige retroalimentación oportuna, oportunidades de práctica y apoyos ajustados al conocimiento previo. Los sistemas de inteligencia artificial generativa amplían la disponibilidad de estas funciones mediante conversaciones, explicaciones, preguntas y comentarios sobre producciones estudiantiles. Sin embargo, la disponibilidad de una respuesta no asegura que el estudiante comprenda el procedimiento que la produce ni que pueda reproducirlo en una situación nueva. La pregunta educativa central es, por tanto, qué condiciones convierten esa asistencia en aprendizaje y cuáles favorecen únicamente la resolución inmediata de una tarea.

La discusión sobre los grandes modelos de lenguaje comprende posibilidades de personalización y dificultades relacionadas con la exactitud, los sesgos y la dependencia de las respuestas generadas. Kasneci et al. (2023) sitúan estas posibilidades junto a la necesidad de desarrollar competencias para utilizar y evaluar críticamente sus productos. Esta perspectiva desplaza la atención desde la herramienta hacia la relación entre estudiante, tarea y acompañamiento docente. Una conversación puede estimular la explicación de un razonamiento, pero también permitir que se eludan las operaciones necesarias para aprenderlo.

La revisión de revisiones de Bond et al. (2024) identifica una expansión de las aplicaciones de inteligencia artificial en educación superior y reclama mayor rigor metodológico, consideración contextual y atención a sus implicaciones éticas. Para examinar la tutoría generativa, estas preocupaciones requieren observar intervenciones concretas: quién recibe la ayuda, qué actividad realiza, cuál es el comparador y cómo se mide el aprendizaje. Comparar un tutor con una clase expositiva responde a una pregunta distinta de compararlo con enseñanza activa, retroalimentación experta o un sistema generativo con otras reglas pedagógicas.

El marco de evaluación formativa de Nicol y Macfarlane-Dick (2006) permite interpretar la retroalimentación como un proceso que fortalece la capacidad del estudiante para juzgar y regular su propio trabajo. Desde esta perspectiva, recibir comentarios resulta insuficiente: es necesario comprenderlos, contrastarlos y decidir cómo aplicarlos. En tutoría generativa, la calidad de esa participación puede depender de consignas que exijan una respuesta inicial, explicaciones de los cambios y comprobación de las sugerencias. La autonomía se expresa en decisiones justificadas y en la capacidad de reconocer cuándo una ayuda es pertinente.

Otra distinción necesaria es la que separa desempeño y aprendizaje. Soderstrom y Bjork (2015) muestran que las mejoras observadas durante la práctica no constituyen por sí solas evidencia de cambios duraderos. En el contexto generativo, una mayor calidad del producto asistido puede coexistir con dificultades posteriores para resolver una tarea equivalente. Por ello, una revisión debe distinguir resultados inmediatos, retención y transferencia, además de precisar si la evaluación se realiza con o sin acceso a la herramienta. Las medidas de satisfacción, confianza o utilidad percibida complementan esta información, pero no la sustituyen.

La evidencia reciente ofrece resultados contrastantes. El ensayo de Kestin et al. (2025) mostró ventajas de una tutoría diseñada con principios pedagógicos frente a enseñanza activa presencial. En cambio, Bauer et al. (2025) no encontraron una ventaja de la retroalimentación adaptativa en el desempeño estadístico y observaron menor interés que con comentarios estáticos. Deriba et al. (2026) identificaron un beneficio de la retroalimentación docente en una prueba diferida de accesibilidad. Estas diferencias aconsejan evitar una interpretación uniforme de la eficacia y examinar la función concreta que desempeña la asistencia.

En la enseñanza de lenguas, la tutoría comprende revisión escrita, conversación, lectura y práctica de vocabulario; en las ciencias de la salud, incorpora pacientes virtuales y razonamiento clínico; en programación, ofrece depuración y orientación sobre procedimientos. También aparecen intervenciones en formación docente, diseño, música y actividad física. Tal diversidad amplía las posibilidades de uso, pero dificulta la comparación directa de resultados. Una mejora en precisión gramatical, una calificación de examen y una actuación clínica observada representan desenlaces educativos diferentes, aun cuando todas las intervenciones empleen un chatbot.

El objetivo de esta revisión fue sintetizar

los resultados de 134 estudios sobre tutoría, andamiaje, retroalimentación y aprendizaje adaptativo generativos en educación superior, con especial atención a las condiciones que favorecen el aprendizaje autónomo. Se plantearon tres preguntas: qué resultados educativos se asocian con estas intervenciones; cómo varían según la tarea, el comparador y la mediación docente; y qué evidencia aportan sobre retención y transferencia después de la práctica. La contribución consiste en relacionar los hallazgos con decisiones pedagógicas concretas, incluyendo resultados favorables, nulos y favorables al comparador.

Métodos

Diseño y delimitación

Se realizó una revisión sistemática con síntesis narrativa estructurada. El reporte siguió PRISMA 2020 para describir la identificación, selección e inclusión de estudios (Page et al., 2021). La búsqueda se documentó conforme a los principios de PRISMA-S (Rethlefsen et al., 2021). La pregunta se delimitó mediante población, intervención, comparación, desenlaces y diseño. La población comprendió estudiantes de grado y posgrado, incluidos programas universitarios de formación profesional y especialización. La intervención consistió en apoyo generativo integrado en una actividad de aprendizaje; la mera opinión sobre una herramienta no constituyó una intervención elegible.

Se incluyeron estudios empíricos con comparación entre condiciones, cohortes o momentos de evaluación, y resultados educativos observados mediante pruebas, productos, actuaciones o indicadores de proceso. Las percepciones y escalas de autorregulación se consideraron complementarias. La evaluación sin asistencia se analizó como una dimensión del desenlace y no como requisito universal de inclusión; de este modo, se distinguió el desempeño asistido de la evidencia de aprendizaje independiente sin confundirlos. Se admitieron diseños aleatorizados, no aleatorizados y mixtos. No se estableció un tamaño muestral mínimo.

Se excluyeron revisiones sin componente empírico, ensayos de opinión, validaciones exclusivamente técnicas sin resultados educativos, investigaciones fuera de educación superior y aplicaciones sin tutoría, andamiaje, retroalimentación o personalización generativos. Los artículos que combinaron una revisión conceptual con una intervención elegible se conservaron por su componente empírico. Los comparadores incluyeron enseñanza habitual, retroalimentación humana, recursos digitales convencionales, otras configuraciones generativas y mediciones previas. Esta amplitud permitió examinar tanto el valor añadido de la tecnología como las diferencias entre sus formas de integración.

Fuente y estrategia de búsqueda

Se utilizó Scopus y se fijó como fecha de búsqueda el 14 de Julio de 2026. La estrategia combinó tres bloques en título, resumen y palabras clave: tecnologías generativas, educación superior y funciones de apoyo al aprendizaje. Se aplicaron los filtros de año de publicación posterior a 2021, tipo documental artículo, idiomas inglés, español o portugués y acceso abierto. La búsqueda produjo 1519 registros.

La búsqueda no incorporó un límite superior de año en la ecuación; la elegibilidad temporal se vinculó a la disponibilidad del artículo en la fecha de consulta. La selección final comprendió publicaciones fechadas entre 2023 y 2026. No se realizaron búsquedas complementarias en otras bases ni se incluyeron registros recuperados por rastreo de citas. Las referencias conceptuales y metodológicas empleadas en la introducción y la discusión se identificaron por separado y no se contabilizaron como estudios de la síntesis. A cotninuación se presenta la estrategia de búsqueda:

TITLE-ABS-KEY (“generative AI” OR “generative artificial intelligence” OR GenAI OR ChatGPT OR “large language model*” OR “GPT-3*” OR “GPT-4*” OR “GPT-5*” OR “Google Gemini” OR “Claude AI” OR DeepSeek) AND TITLE-ABS-KEY (“higher education” OR universit* OR universidad* OR college* OR undergraduate* OR postgraduate* OR “educación superior” OR “ensino superior”) AND TITLE-ABS-KEY (tutor* OR scaffold* OR feedback OR “adaptive learning” OR “personalized learning” OR “personalised learning” OR retroaliment* OR andamiaje OR “aprendizaje personalizado”) AND PUBYEAR > 2021 AND (LIMIT-TO (DOCTYPE , “ar”)) AND (LIMIT-TO (LANGUAGE , “English”) OR LIMIT-TO (LANGUAGE , “Spanish”) OR LIMIT-TO (LANGUAGE , “Portuguese”)) AND (LIMIT-TO (OA , “all”))

Selección y extracción

Los registros se exportaron con sus campos

bibliográficos, resúmenes e identificadores. Se comprobaron coincidencias exactas de DOI, EID y título normalizado con año; no se identificaron duplicados exactos. El cribado se desarrolló en dos etapas: examen de título y resumen y valoración de elegibilidad del informe completo. Dos revisores contrastaron las decisiones y resolvieron las discrepancias mediante consenso, utilizando una razón principal por exclusión. Se mantuvo una relación entre cada informe, su identificador de la base y su referencia bibliográfica.

La extracción comprendió autoría, año, contexto educativo, diseño, características de la intervención, comparador, instrumentos, momento de evaluación y principales resultados. Se diferenciaron las comparaciones entre grupos de las mejoras internas de un grupo. También se registraron las intervenciones compuestas, en las que la asistencia generativa se combinó con formación, cambios curriculares o retroalimentación docente. El informe fue la unidad de síntesis; no se sumaron participantes entre publicaciones, para evitar una precisión artificial y posibles dobles recuentos entre cohortes relacionadas.

Evaluación metodológica con MMAT

Se aplicó el Mixed Methods Appraisal Tool, versión 2018, a los 134 estudios (Hong et al., 2018). Primero se examinaron la claridad de la pregunta y la pertinencia de los datos para responderla. Después se seleccionó la familia de criterios correspondiente al diseño. En los ensayos se valoraron la asignación, comparabilidad inicial, integridad de los desenlaces, cegamiento de evaluadores y adherencia. En los estudios no aleatorizados se examinaron representatividad, medición, integridad de los datos, control de confusión y ejecución de la intervención.

En los estudios mixtos se valoraron la justificación de la combinación metodológica, integración de componentes, interpretación conjunta, tratamiento de divergencias y calidad de los componentes. La presencia de un análisis de varianza mixto no se consideró suficiente para clasificar una investigación como de métodos mixtos. Los juicios se expresaron por criterio, sin calcular porcentajes de calidad ni una puntuación global. para los diseños mixtos incluye los bloques cualitativo, cuantitativo e integrador.

Los 134 estudios superaron las preguntas iniciales, fueron evaluados y permanecieron incluidos. La conservación no se interpretó como ausencia de limitaciones: los criterios no satisfechos se incorporaron a la lectura crítica de los resultados. La decisión de inclusión se basó en la elegibilidad y la contribución a la pregunta, sin utilizar un umbral numérico de aprobación. De este modo, la valoración metodológica orientó el peso interpretativo de la evidencia y permitió mantener visibles los hallazgos discordantes.

Síntesis de los resultados

Se desarrolló una síntesis sin metaanálisis, con atención a los principios de organización y transparencia de SWiM (Campbell et al., 2020). Cada estudio se asignó a un eje principal: escritura; otras competencias lingüísticas; programación, ingeniería y ciencias; salud; aprendizaje transversal y formación docente; o educación física. Las funciones de retroalimentación, mediación, autorregulación y evaluación diferida se analizaron transversalmente. La asignación principal evitó contar dos veces un estudio en la descripción del corpus, aunque una publicación pudiera informar varios temas.

La heterogeneidad de intervenciones, comparadores, escalas y tiempos de evaluación desaconsejó un efecto combinado. No se efectuó un recuento de significación estadística como prueba de eficacia. Se examinaron la dirección del contraste, el tipo de resultado y las limitaciones del diseño; los tamaños de efecto se mencionaron únicamente como resultados particulares. La ausencia de una diferencia estadísticamente significativa no se interpretó como equivalencia. Se priorizó la coherencia entre el resultado medido y la conclusión educativa, especialmente cuando se formularon afirmaciones sobre autonomía, retención o transferencia.

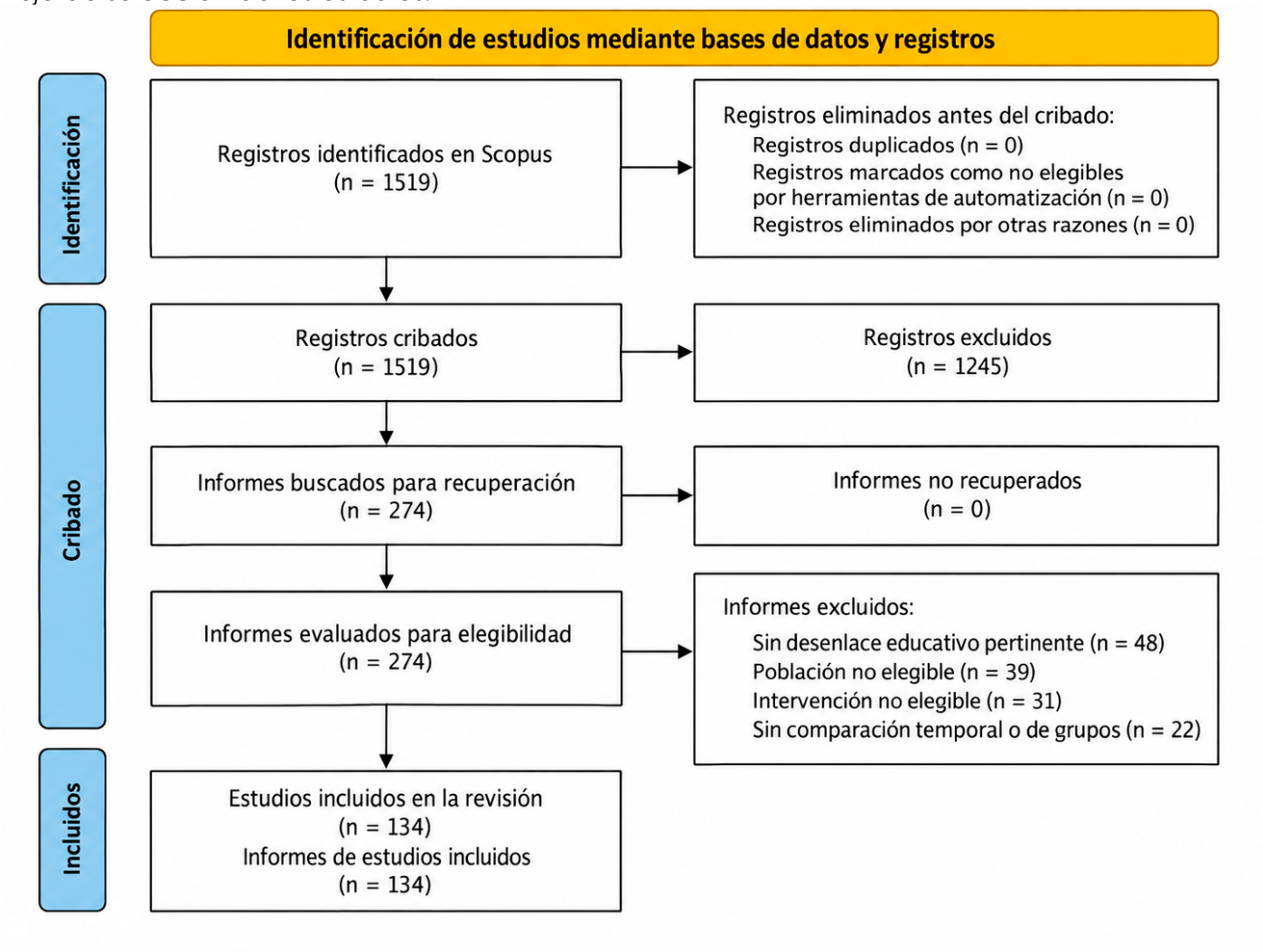
Resultados

Selección y características del corpus

Se identificaron 1519 registros y se cribaron los 1519 tras la comprobación de duplicados. Se excluyeron 1245 por título y resumen. Se recuperaron y evaluaron 274 informes completos; 140 se excluyeron por ausencia de resultados educativos pertinentes (48), población fuera de educación superior (39), intervención sin apoyo generativo elegible (31) o ausencia de contraste entre condiciones o momentos (22). La síntesis incluyó 134 informes correspondientes a 134 estudios. La Figura 1 presenta la secuencia y sus cantidades.

Figura 1.

Flujo de selección de los estudios.



Fuente: elaboración propia, con la estructura de reporte de PRISMA 2020 (Page et al., 2021).

Los 134 estudios se distribuyeron en 78 fuentes y estuvieron publicados en inglés. Por año, se incluyeron uno de 2023, nueve de 2024, 40 de 2025 y 84 de 2026. La concentración en los dos últimos años fue del 92,54 %. La selección comprendió experiencias de grado, formación docente, posgrado y especialización. Los modelos de intervención abarcaron asistentes generales, tutores personalizados, pacientes virtuales, agentes con funciones pedagógicas delimitadas y sistemas integrados con recursos curriculares.

Tabla 1.

Distribución de los estudios por eje principal

Eje de síntesis	n	%
Escritura académica y retroalimentación	37	27,61
Otras competencias lingüísticas	27	20,15
Programación, ingeniería y ciencias	22	16,42
Formación en ciencias de la salud	25	18,66
Aprendizaje transversal y formación docente	18	13,43
Educación física y fisiología del ejercicio	5	3,73
Total	134	100,00

Nota. Cada estudio se asignó a un solo eje principal; los mecanismos pedagógicos se analizaron de manera transversal.

Resultados de la evaluación MMAT

La evaluación se completó para los 134 estudios y ninguno se excluyó después de su valoración metodológica. Todos presentaron una pregunta identificable y datos pertinentes para examinarla. La distribución por familias

de diseño y los resultados por criterio se presentan en la Tabla 2. Las principales reservas interpretativas se relacionaron con el cegamiento de evaluadores, el control de confusión en comparaciones no aleatorizadas y la solidez de los componentes integrados en estudios mixtos.

Tabla 2.

Juicios por familia de diseño y criterio MMAT

Familia	n	C1	C2	C3	C4	C5
Aleatorizados	30	30/0	30/0	29/1	4/26	30/0
No aleatorizados	54	54/0	54/0	54/0	15/39	54/0
Mixtos: Integración	50	50/0	50/0	50/0	50/0	3/47

Nota. Cada celda muestra Sí/No; los criterios C1–C5 corresponden a la familia indicada.

Estas reservas condicionaron la lectura de los efectos: una comparación entre clases intactas no recibió la misma interpretación causal que una asignación aleatoria, y una asociación entre frecuencia de uso y calificaciones no se consideró evidencia suficiente de que el uso causara la mejora. La retención de todos los estudios permitió conservar la diversidad de resultados, incluidos los favorables a la enseñanza humana y aquellos en los que la asistencia generativa no añadió beneficios detectables.

Escritura académica y retroalimentación

Las intervenciones de escritura mostraron que la mejora depende de la dimensión evaluada y de cómo se utiliza el comentario. Alalwi et al. (2026) observaron ventajas en vocabulario, precisión y organización mediante retroalimentación inmediata. Allehyani et al. (2026) encontraron una reducción de errores morfológicos, mientras que Vo y Ho (2026) registraron mayor precisión gramatical con una modalidad apoyada en ChatGPT. Estos resultados se concentran en tareas susceptibles de revisión reiterada y no implican por sí mismos desarrollo equivalente de todas las competencias discursivas.

La precisión tampoco coincidió siempre con la complejidad. Mahmud (2025) encontró mejores resultados de corrección escrita con ChatGPT, aunque la identificación de errores fue ligeramente más exacta mediante retroalimentación humana. Alkhalifah y Almutlaq (2026) no hallaron una ventaja sólida en complejidad sintáctica global o clausal; las diferencias exploratorias a nivel de sintagma perdieron significación tras corregir comparaciones múltiples. Hu et al. (2026) mostraron una distribución complementaria: la retroalimentación generativa favoreció precisión, coherencia y organización, mientras la docente apoyó mejor la complejidad gramatical y léxica.

En tareas escritas más amplias, Mujeeb et al. (2026) informaron mejoras en gramática, vocabulario, cohesión y calidad global mediante Grammarly y ChatGPT, dentro de una intervención combinada. Jamshed et al. (2025) registraron reducción de errores con retroalimentación distribuida por WhatsApp; el uso de ChatGPT para puntuar las producciones exige considerar la independencia de la medición. Polakova y Ivenz (2024) encontraron avances en concisión, gramática e incorporación de información, y Smithsarakarn (2026) destacó la organización del texto, con un desarrollo léxico menos inmediato. La recurrencia de resultados favorables debe interpretarse según el instrumento y la duración del curso.

Los beneficios fueron más específicos en escritura para fines profesionales. Mohamed et al. (2025) observaron mejoras en contenido y estructura de correos, con diferencias por nivel de dominio y sin ganancias claras en cohesión. Y. Pan (2025) combinó comentarios generativos vinculados a rúbricas con revisión entre pares en inglés médico; los resultados favorecieron contenido, estructura y uso lingüístico, junto a reservas sobre dependencia y distancia emocional. En chino como segunda lengua, G. Huang y Chen (2026) relacionaron la retroalimentación con mejores escritos y autoeficacia, aunque los estudiantes señalaron el riesgo de expresiones formularias.

La alfabetización en retroalimentación aparece como un mecanismo relevante. Oktarin et al. (2024) encontraron mejoras conjuntas en escritura, compromiso y comprensión del uso de comentarios. Yang, Hussin y Hashim (2025) observaron más revisiones de contenido y mayor autoeficacia cuando la interacción permitía discutir las sugerencias. Alshehri et al. (2026) integraron planificación, verificación de afirmaciones y revisión iterativa, con resultados

favorables en escritura y pensamiento crítico digital. Andewi et al. (2025) documentaron una evolución desde consultas simples hacia intercambios de varios turnos, pero también dificultades para verificar la exactitud y evitar la delegación excesiva.

La asignación de responsabilidades dentro del proceso de escritura produjo diferencias importantes. C. Lai et al. (2026) compararon quién elaboraba el borrador y quién evaluaba: la configuración con producción inicial generativa y evaluación y revisión estudiantiles mostró mayores ganancias. Este resultado apoya la centralidad del trabajo evaluativo, sin justificar la delegación indiscriminada. Ren et al. (2025) añadieron ejercicios que reproducían errores recurrentes del estudiante y observaron una ventaja intermedia respecto de recibir solo comentarios generativos. Zargaran y Aldhahli (2026) identificaron beneficios distintos según la etapa: la lluvia de ideas favoreció respuesta y organización, mientras la corrección posterior benefició gramática y vocabulario.

La oportunidad temporal y la interacción social también importaron. Wiboolyasarin et al. (2026) hallaron ventajas de la retroalimentación inmediata y combinada frente a la diferida, sin beneficio adicional claro de combinar ambas respecto de la inmediata. Wiboolyasarin et al. (2024) integraron comentarios generativos en escritura colaborativa mediante wiki, con mejores resultados posteriores. Khusniyah et al. (2026) no identificaron diferencias significativas entre comentarios generativos y revisión entre pares para redactar introducciones con el modelo CaRS. En conjunto, estos estudios sugieren ajustar la secuencia y conservar espacios de deliberación entre estudiantes.

Los resultados relativos al pensamiento de orden superior fueron heterogéneos. X. Chen et al. (2026) informaron más revisión y comportamientos analíticos con retroalimentación docente apoyada en ChatGPT. Hao et al. (2026) combinaron evaluación automática y comentarios generativos, asociándolos con mayores niveles de análisis, evaluación y creación. He, Niu y Ashari (2026) encontraron ventajas de una modalidad multimodal sobre una textual en disposiciones críticas y profundidad argumentativa. H. Hong et al. (2025) describieron mejoras cuando las tareas de menor complejidad se delegaban dentro de ciclos que preservaban crítica individual, revisión compartida y reflexión. La evidencia se refiere a esas configuraciones completas, no a cualquier acceso al sistema.

Los modelos de regulación permitieron distinguir logros cercanos de aprendizajes persistentes. Selvi et al. (2026) observaron mejoras en escritura y regulación metacognitiva, moduladas por el nivel lingüístico. Yang, Huang, et al. (2025) registraron ventajas de un tutor virtual basado en corregulación frente a apoyo generativo estándar. Sin embargo, Yuan (2026) encontró que la ventaja inmediata de las indicaciones metacognitivas adaptativas no se mantuvo en la evaluación diferida. Dai (2026) mostró que los guiones de alfabetización en retroalimentación y la calibración de la autoevaluación incidían en resultados diferentes; la combinación favoreció la calidad escrita al retirarse la ayuda, sin superar a la calibración aislada en precisión del juicio propio.

Las comparaciones directas con docentes evitaron una lectura exclusivamente favorable. Escalante et al. (2023) no encontraron diferencias de aprendizaje entre retroalimentación generativa y tutoría humana, y las preferencias se dividieron. Ekşi et al. (2026) observaron mejoras en ambos grupos de escritura en turco, sin diferencia global significativa y con ventajas distintas por dimensión. Puteri et al. (2025) registraron percepciones positivas, pero sin superioridad significativa del postest frente al método convencional. Estos hallazgos no prueban equivalencia; indican que el valor añadido depende del contraste y del desenlace.

La elección de plataforma tampoco resolvió por sí sola el problema pedagógico. Kawinkoonlasate (2025) encontró mejoras con Gemini y ChatGPT y ventajas particulares de Gemini en retroalimentación multimodal e integración de fuentes, dentro de dos contextos universitarios. Nelson (2026) relacionó un flujo de trabajo estructurado con mejores resultados en escritura investigativa, pero la intervención incorporó simultáneamente tutor, videos, cuestionarios supervisados y cambios de tarea. Shi et al. (2026) identificaron una tensión decisiva: mejores puntuaciones escritas con ChatGPT coexistieron con un ideal de escritor menos favorable que con evaluación automática convencional y con relatos de pérdida de creatividad y agencia.

Competencias orales, lectura, vocabulario y aprendizaje de lenguas

En expresión oral, Abdelrady et al. (2026) observaron avances en fluidez, gestión de

turnos y autonomía mediante conversación por voz con Copilot. Al-khresheh y Alruwaili (2026) encontraron mejores puntuaciones ajustadas con andamiaje escrito de ChatGPT combinado con producción oral presencial; la diferencia inicial entre grupos obliga a atender al ajuste estadístico. Dinçer et al. (2026) documentaron cambios acústicos favorables en francés, aunque con pocos conglomerados y una intervención compuesta. Estos estudios muestran que conversación por voz, preparación escrita y práctica docente son mecanismos distintos y no deberían describirse como una única modalidad.

Las configuraciones explícitamente pedagógicas aportaron resultados más precisos. Z. C.-C. Lai (2025) encontró mayor desarrollo oral con ChatGPT dentro de BOPPPS que con práctica generativa no estructurada. Lin y Lin (2026) observaron ventajas de un agente inspirado en evaluación dinámica en complejidad y cumplimiento de la tarea, pero no en fluidez o exactitud. Sok y Shin (2025) mostraron que una consigna específica de resumen añadió beneficios en esa habilidad a los obtenidos en desempeño oral general. La especificidad de la orientación parece tan relevante como la disponibilidad de interacción.

Las experiencias sociales y afectivas complementaron los resultados lingüísticos. Laksana et al. (2026) vincularon portafolio electrónico y ChatGPT con mejores actuaciones orales y alfabetización en retroalimentación. Z. Li et al. (2026) encontraron avances en compromiso emocional y expresión oral al incorporar un chatbot en aula invertida. Qu y Chen (2026) observaron mejoras interculturales y menor ansiedad con interlocutores generativos, aunque persistieron limitaciones en señales paralingüísticas. Jia et al. (2026) evaluaron práctica informal guiada con seguimiento a diez semanas; su resumen destaca ventajas inmediatas en capacidad y disfrute, por lo que no corresponde extender automáticamente ese resultado al seguimiento.

La lectura presentó un patrón de apoyo estratégico. Aldamen et al. (2026) compararon tres clases y observaron mejores resultados descriptivos con enseñanza explícita de estrategias, especialmente cuando incorporaba ChatGPT; una sola clase por condición impidió aislar el efecto de la intervención. Ironsi y Bostanci (2026) organizaron la asistencia antes, durante y después de la lectura, con mejores comprensión y autonomía percibida. C. Huang et al. (2025) encontraron mayores resultados lectores con un aprendizaje basado en problemas personalizado, sobre todo entre estudiantes más comprometidos. Chiang et al. (2026) informaron mejoras en lectura TOEIC y mostraron la importancia de autoeficacia, condiciones facilitadoras y adecuación entre tarea y tecnología.

En vocabulario y gramática, S. Albaqami y Alahmadi (2026) encontraron mayores puntuaciones inmediatas y después de cuatro semanas en la condición apoyada por ChatGPT. Hussein y Kurt (2026) combinaron un modelo de lenguaje con reconocimiento de voz y práctica de recuperación oral, con ganancias de vocabulario productivo. Lei et al. (2026) observaron una mejora semántico-sintáctica mayor en una clase apoyada por ChatGPT, pero no una diferencia fonética semejante; la condición confundida con la clase y la muestra de 16 estudiantes restringieron la inferencia causal. Ali et al. (2025) comunicaron mejoras gramaticales en una intervención breve, mientras Lukešová y Jennings (2026) encontraron los mayores avances cuando el uso estuvo acompañado por el docente.

Las intervenciones multicomponente aportaron evidencia sobre alcance y equidad. Abdelhalim y Almaneea (2026) relacionaron proyectos colaborativos con mayores logros lingüísticos y autonomía, sin una ventaja diferencial clara en gramática. S. E. Albaqami et al. (2026) observaron mejoras en expresión oral y escritura de estudiantes de ingeniería. J. Chen y Alibakhshi (2026) combinaron formación inicial, práctica guiada y valoración humana, con ganancias en estrategias y rendimiento lingüístico. Y.-H. Chen y Keng (2026) integraron ChatGPT en aprendizaje cooperativo STAD y describieron una reducción de brechas por dominio lingüístico, aunque algunos resultados afectivos exploratorios no resistieron la corrección estadística.

La traducción, la formación docente y la comprensión auditiva ampliaron el panorama. Guo y Lin (2025) encontraron mejoras en terminología y adecuación cultural, junto a desigualdades previas que disminuyeron con andamiaje. Sarzhanova y Gulnur (2026) evaluaron una combinación de ChatGPT, traducción y herramientas adaptativas, con mejores resultados y alfabetización digital. Q. Zhou et al. (2025) registraron ganancias

auditivas mediante tareas informales guiadas. En contraste, Zheldibayeva (2025) observó ventajas inmediatas en escucha y escritura que no persistieron en el seguimiento. Por ello, el acceso a práctica individualizada debe acompañarse de estrategias para sostener los aprendizajes.

Programación, ingeniería, ciencias y estadística

La programación reunió intervenciones de apoyo a la resolución y depuración. Abdulla et al. (2024) encontraron mejores resultados con ChatGPT como recurso complementario; Atta et al. (2025) describieron beneficios de una aplicación móvil para C#; y Alshammari (2025) observó mejores resultados y satisfacción al integrar ChatGPT en un sistema adaptativo para Java. Bojorque et al. (2026) evaluaron una arquitectura móvil de generación aumentada por recuperación en minería de datos, con resultados favorables durante varios semestres. La diversidad de dispositivos y sistemas refuerza la necesidad de describir la actividad del estudiante y las funciones efectivamente disponibles.

La integración curricular proporcionó un segundo conjunto de hallazgos. Gasaymeh y Almohtadi (2024) encontraron ventajas de un aula invertida interactiva con ChatGPT frente al aula invertida convencional. Huesca et al. (2024) observaron mayores ganancias normalizadas en cursos de programación de ingeniería, y Huesca et al. (2026) identificaron la estrategia de enseñanza como un predictor relevante de esas ganancias en múltiples campus. Yang, Zhang y Jiang (2025) registraron mejoras en arquitectura de computadores mediante un modelo de docencia acompañado por asistente. Estas experiencias sugieren un valor del diseño de enseñanza completo, sin permitir atribuir cada diferencia exclusivamente al modelo generativo.

Los beneficios no se extendieron a todas las capacidades. Omeh et al. (2025) encontraron una ventaja en programación con Gemini y aprendizaje basado en problemas, pero no en pensamiento crítico o resolución de problemas; los estudiantes de mayor capacidad inicial obtuvieron más beneficio. Hou et al. (2026) mostraron que indicaciones orientadas al proceso favorecieron planificación y depuración deliberada en Python. Mellado y Cubillos (2025) encontraron mayores ganancias con videos que con Copilot en una etapa inicial de PHP. La secuencia de explicaciones y el conocimiento previo pueden, por tanto, determinar cuándo conviene introducir ayuda adaptativa.

En ciencias, Kestin et al. (2025) mostraron resultados favorables de un tutor configurado con principios de enseñanza frente a aprendizaje activo presencial. Lee et al. (2024) encontraron ventajas de solicitar intentos independientes y ofrecer pistas frente al uso convencional de ChatGPT en química. Slijepcevic y Yaylali (2025), en cambio, no observaron diferencias significativas entre Khanmigo, búsqueda web y materiales impresos en conceptos de física. Luan et al. (2025) hallaron resultados menos favorables de la retroalimentación generativa frente a la humana en un curso de neurociencia, incluido un indicador indirecto de retención. La calidad de la alternativa disponible resulta decisiva para interpretar el valor añadido.

Los estudios de estadística ilustraron la dependencia de la tarea. Bauer et al. (2025) no encontraron mayor desempeño con comentarios adaptativos y observaron menor interés. Kabba y Ejbari (2026) tampoco identificaron una ventaja general durante siete sesiones, aunque sí un contraste favorable en una sesión de cálculo multietapa. Estas diferencias aconsejan introducir la herramienta donde pueda responder a una dificultad concreta, en lugar de extender su uso por igual a todo el curso.

La transferencia aportó resultados especialmente contrastantes. Elshall et al. (2026) observaron mejores proyectos de ciencia de datos ambiental sin deterioro de exámenes conceptuales sin asistencia, al comparar dos cohortes pequeñas y dos versiones del sistema. Syromiatnykov et al. (2026) encontraron ventajas inmediatas y diferidas en tareas independientes de ingeniería agrícola, con una asignación determinista que limitó la interpretación causal. Zhang et al. (2026) registraron retención semejante, pero menor transferencia con ChatGPT que con un instructor humano. Por su parte, Deriba et al. (2026) observaron mejor retención diferida de accesibilidad con retroalimentación docente, aunque las tareas inmediatas no mostraron diferencias significativas.

Finalmente, Wang et al. (2026) encontraron mejores resultados cognitivos y prácticos al añadir retroalimentación generativa a realidad virtual para hardware y programación, con beneficios importantes entre estudiantes de menor desempeño. Este resultado contrasta

con las ventajas concentradas en estudiantes de mayor capacidad en otros trabajos. La equidad no emerge automáticamente de la personalización: depende del tipo de ayuda, las exigencias de la tarea y las condiciones iniciales del alumnado.

Formación en Ciencias de la Salud

Las aplicaciones en salud combinaron adquisición conceptual y práctica de habilidades. Gu et al. (2026) encontraron mejores resultados teóricos y clínicos con DeepSeek dentro de aprendizaje basado en problemas sobre glaucoma. Ilyas et al. (2026) comunicaron mejoras de rendimiento con itinerarios personalizados en abdomen agudo. Xu et al. (2026) observaron ventajas en conocimientos y dimensiones de una evaluación clínica breve de diagnóstico. Niu et al. (2026) integraron andamiaje y ciclos de revisión en inmunología, con resultados favorables en conocimiento y toma de decisiones. Estos hallazgos respaldan configuraciones educativas supervisadas, sin constituir evidencia directa de mejora de la atención a pacientes.

En bioquímica, L. Li et al. (2025) informaron mayor eficiencia y puntuaciones con apoyo generativo, mientras los estudiantes señalaron limitaciones en razonamiento complejo e innovación. Y. Huang et al. (2024) describieron mejoras de cumplimiento de actividades y satisfacción en prácticas de cirugía. Randombage et al. (2026) no encontraron una ventaja significativa del acompañamiento del tutor sobre uso autodirigido en conocimiento inmediato, pese a percepciones favorables. Chun et al. (2025) tampoco identificaron diferencias significativas en rendimiento, tiempo o satisfacción con un libro generativo de uso local. La aceptación y la facilidad de acceso no deben confundirse con un efecto demostrado sobre el aprendizaje.

La retención produjo resultados diversos. Fan et al. (2026) encontraron una ventaja inmediata de DeepSeek en anestesiología, que no se mantuvo como diferencia significativa al mes. Daniel et al. (2026) observaron mejoras sostenidas de preguntas de orden superior con recuperación supervisada a uno y tres meses, en un diseño de intervención diferida no aleatorizado. Zhao et al. (2026) informaron mejores resultados de anatomía y retención al integrar tutor y grafo de conocimiento; completaron el estudio 301 de 362 estudiantes. La interpretación conjunta requiere considerar tanto el mecanismo pedagógico como las pérdidas y la definición temporal del desenlace.

En razonamiento clínico, Nartthanarung et al. (2026) observaron mejoras con ChatGPT y con buscadores convencionales, sin atribución exclusiva a una modalidad. Çiçek et al. (2025) no encontraron diferencias globales entre retroalimentación generativa y experta ni inmediatamente ni a diez días, aunque los expertos obtuvieron mejores resultados en casos de mayor complejidad. Rother et al. (2026) tampoco identificaron una diferencia en examen final entre comentarios personalizados y generales; la intensidad de práctica se asoció con mejores resultados. La práctica de recuperación y la resolución de casos pueden explicar una parte importante de las ganancias que acompañan al uso del sistema.

Los pacientes virtuales ampliaron las oportunidades de ensayo. Z. Wang et al. (2025) encontraron mejores puntuaciones de entrevista clínica con pacientes basados en GPT que con juego de roles. McCarrick et al. (2025) añadieron sesiones estructuradas durante cirugía y observaron mejor evolución de la historia clínica evaluada con paciente humano y evaluador cegado. He, Chen, et al. (2026) encontraron una ventaja en entrevista clínica tras emparejamiento por propensión, con mayores beneficios entre estudiantes de mejor desempeño inicial; al ser voluntario el uso, persiste la posibilidad de confusión residual. Estos diseños aportan mayor proximidad a una actuación profesional que una encuesta de satisfacción.

Los resultados comunicativos no fueron uniformes. Lee et al. (2025) no identificaron diferencias significativas entre chatbot y juego de roles en un ensayo piloto de 19 estudiantes, con fortalezas complementarias. Jay-Jímenez et al. (2026) observaron mejoras en ambos grupos de comunicación diagnóstica, pero sin una interacción ajustada concluyente y con 51 casos completos entre 80 asignados. Lim et al. (2026) relacionaron el uso repetido de avatares ATLAS con mejor comunicación en farmacia; la autoselección y las diferencias de participación aconsejan interpretar el resultado como asociación. La simulación conversacional puede complementar la práctica interpersonal, pero no reproduce todas sus exigencias.

La supervisión fue relevante en tareas visuales y técnicas. Hong et al. (2026) encontraron mejores resultados de ultrasonografía cuando ChatGPT

se combinó con tutorías docentes; el sistema aislado mostró dificultades en preguntas con imágenes. Gokkurt Yilmaz et al. (2025) observaron mayor precisión radiográfica al mes mediante guías personalizadas que con comentarios de acierto o error. Veerabhadrappa et al. (2026) encontraron tendencias favorables en técnicas intraorales, con una diferencia significativa entre grupos restringida a un componente. Wen et al. (2026) mostraron mejores habilidades operativas inmediatas en realidad virtual dental con apoyo generativo, mientras sus resultados neurofisiológicos secundarios tuvieron carácter exploratorio.

La atención a estudiantes con dificultades aportó una última perspectiva. Sami et al. (2026) observaron convergencia de calificaciones entre usuarios inicialmente rezagados y sus compañeros tras incorporar asistentes docentes. El uso voluntario y el carácter observacional impiden asumir que toda convergencia se deba al asistente. Aun así, el estudio identifica una función práctica relevante: ofrecer un espacio accesible de consulta y práctica que puede complementar los apoyos institucionales existentes.

Aprendizaje transversal, formación docente y sistemas adaptativos

Las intervenciones transversales mostraron efectos dependientes del nivel cognitivo. Abdelmagid et al. (2025) encontraron mejoras en profundidad de conocimiento de estudiantes de posgrado. Rahman et al. (2025) observaron avances en recordar, comprender y aplicar, pero no en analizar, evaluar y crear. Kinder et al. (2025) encontraron mejor justificación escrita de decisiones diagnósticas en formación docente, sin mayor exactitud de la decisión. Así, una explicación más extensa o mejor argumentada no garantiza que la conclusión sea correcta; ambas dimensiones deben evaluarse separadamente.

Los resultados también variaron según la organización de la interacción. Chu et al. (2026) observaron mejor desempeño de proyectos con interacción colaborativa en un sistema de agentes, mientras la interacción individual favoreció la conciencia crítica. Taveros et al. (2026) encontraron mejores problemas formulados, diálogo y revisiones sustantivas cuando los agentes tenían restringida la entrega de respuestas finales. M. Pan et al. (2026) relacionaron anotación social asistida con mejor lectura y participación cognitiva, con aprovechamiento desigual según rendimiento. Estas experiencias destacan la necesidad de diseñar funciones que sostengan la negociación de significado entre estudiantes.

Los modelos de apoyo reflexivo aportaron evidencia sobre independencia. Ates (2026) comparó cuatro formas de retroalimentación en 1176 estudiantes de 48 secciones: las configuraciones reflexiva e híbrida superaron a la directa en transferencia diferida sin asistencia, mientras la híbrida mostró el patrón más favorable en aprendizaje conceptual. X. Wang et al. (2026) desarrollaron el marco IDEA y observaron ventajas en algunas tareas sin ayuda realizadas cinco días después. El segundo resultado es preliminar y de corto intervalo; ambos trabajos convergen en preservar planificación, evaluación y revisión como responsabilidades estudiantiles.

Las experiencias de formación docente y aprendizaje aplicado ampliaron las condiciones de uso. Dahri et al. (2025) encontraron beneficios de revisión móvil después de clase en un curso de evaluación educativa. Yusof (2025) relacionó práctica de recuperación apoyada por ChatGPT con mejores exámenes finales al comparar cohortes. de la Puente et al. (2026) observaron mejoras en inteligencia de mercados mediante aprendizaje experiencial en Colombia. Fayid y Kiraz (2026) hallaron una ventaja ajustada en conocimiento ambiental, pero no en actitudes o conducta. La transferencia desde una puntuación cognitiva hacia cambios de actuación debe demostrarse y no deducirse automáticamente.

Los sistemas adaptativos produjeron resultados favorables y nulos. Samat et al. (2026) registraron mejores puntuaciones y resolución de problemas mediante un currículo dinámico. X. Zhou et al. (2026) encontraron menor esfuerzo mental y mejores resultados con un sistema de recuperación y representación audiovisual en diseño. Zhu et al. (2026) observaron una trayectoria de mejora en música mediante retroalimentación adaptativa, aunque parte de la evidencia correspondió a indicadores técnicos indirectos. Estos resultados requieren distinguir rendimiento del sistema, experiencia del usuario y competencia educativa adquirida.

La disponibilidad permanente de ayuda no aseguró beneficios. Thoeni y Fryer (2026) no identificaron efectos significativos de un chatbot con recuperación sobre logro, interés, autoeficacia o compromiso durante un semestre.

Melanou et al. (2026) observaron aprendizaje en todas las condiciones, sin beneficio adicional de la incorporación generativa. Filatova et al. (2025) describieron mayor proximidad del discurso estudiantil al lenguaje disciplinar en un modelo híbrido, junto a problemas de carga cognitiva y autoridad epistémica. En consecuencia, la accesibilidad del tutor debe acompañarse de objetivos, orientación y criterios claros para valorar la información.

Educación física y habilidades motrices

La evidencia motriz puso de relieve la necesidad de comprobar la exactitud del comentario. Qiu (2024) encontró mejores resultados docentes en golpeos de bádminton, sin diferencias en desplazamientos, y detectó instrucciones generativas inexactas. En contraste, Rakha (2026) observó ventajas en habilidades de balonmano al integrar comentarios previamente verificados en enseñanza recíproca. La diferencia entre estas configuraciones sugiere que la supervisión y los criterios observables de ejecución son componentes relevantes del apoyo.

Xiao y Guo (2026) informaron mejoras de desempeño en atletismo y baloncesto con planes y retroalimentación personalizados, acompañadas de preocupaciones por dependencia y capacitación. Zhifang y Wentao (2026) encontraron beneficios de una estrategia de diagnóstico, generación, retroalimentación y ajuste en acondicionamiento físico. Ribeiro Da Silva Junior et al. (2026) observaron mayores ganancias de conocimiento en fisiología del ejercicio con un tutor específico. Este último desenlace corresponde a aprendizaje conceptual, no a rendimiento corporal; mantener esa distinción evita trasladar conclusiones entre resultados que requieren demostraciones diferentes.

Integración de los resultados

La síntesis de los 134 estudios muestra tres regularidades. Primero, las ganancias se vinculan con mayor consistencia a configuraciones delimitadas que a la mera disponibilidad de la herramienta. Segundo, el beneficio cambia según el resultado: exactitud, complejidad, motivación, autorregulación y transferencia no evolucionan necesariamente en paralelo. Tercero, los comparadores activos, las pruebas diferidas y las tareas independientes revelan límites que una evaluación exclusivamente inmediata puede ocultar.

Discusión

La evidencia reunida permite sostener una conclusión condicionada: la tutoría generativa puede ampliar la práctica y mejorar resultados específicos cuando forma parte de una secuencia pedagógica que conserva la participación intelectual del estudiante. No permite afirmar que cualquier chatbot produzca un beneficio general, que sustituya al docente o que todo producto de mayor calidad represente aprendizaje duradero. Los estudios favorables conviven con resultados nulos, ventajas del comparador y efectos que desaparecen en el seguimiento. Estas diferencias constituyen información para decidir cómo enseñar, no excepciones que deban descartarse.

Del comentario disponible al trabajo evaluativo

La primera contribución de la revisión es situar la actividad del estudiante en el centro de la explicación. Las intervenciones con pistas, revisión razonada, preguntas específicas y crítica de propuestas distribuyen de manera distinta el trabajo cognitivo. Mientras unas permiten delegar la producción, otras devuelven al estudiante la obligación de justificar, verificar y revisar. Varios estudios respaldan esta lectura desde tareas diferentes (Ates, 2026; Dai, 2026; C. Lai et al., 2026; Lee et al., 2024; Taveros et al., 2026). La implicación es que el diseño de la ayuda debe especificar qué operaciones permanecen bajo responsabilidad del estudiante.

Esta interpretación es compatible con Nicol y Macfarlane-Dick (2006), quienes destacan el papel activo del aprendiz en la generación y el uso de retroalimentación. Sin embargo, la revisión no demuestra un mecanismo universal: algunos estudios miden autorregulación mediante cuestionarios y otros observan acciones o productos. Conviene distinguir la confianza en autorregularse de la regulación efectivamente ejecutada. Los registros de revisión y depuración pueden complementar las escalas, siempre que se vinculen con una actuación educativa relevante y no solo con la cantidad de interacciones.

La mediación docente opera mediante selección de contenidos, organización de tareas, validación de respuestas y atención a dificultades. No exige intervenir en cada turno de conversación. Se documentan distintas formas de acompañamiento (Hong et al., 2026; Lukešová & Jennings, 2026; Rakha, 2026), mientras otras comparaciones muestran que

la retroalimentación humana y entre pares conserva funciones valiosas (Escalante et al., 2023; Khusniyah et al., 2026; Lee et al., 2025). La combinación debe responder a necesidades identificadas. Agregar varias fuentes sin definir su función podría aumentar la carga de interpretación y multiplicar comentarios contradictorios.

Desempeño inmediato, retención y transferencia

La segunda contribución es diferenciar lo que permanece de lo que mejora durante la asistencia. Las ventajas inmediatas que desaparecen en algunos seguimientos (Fan et al., 2026; Yuan, 2026; Zheldibayeva, 2025) contrastan con resultados diferidos favorables en otras intervenciones (S. Albaqami & Alahmadi, 2026; Ates, 2026; Daniel et al., 2026; Syromiatnykov et al., 2026). Otras comparaciones con docentes o expertos resultaron menos favorables en retención, transferencia o casos complejos (Çiçek et al., 2025; Deriba et al., 2026; Zhang et al., 2026). Este patrón es consistente con la distinción entre aprendizaje y desempeño de Soderstrom y Bjork (2015), y obliga a especificar el intervalo, la novedad de la tarea y la disponibilidad de apoyos.

Una prueba sin herramienta al finalizar una sesión aporta información sobre ejecución independiente inmediata, pero no demuestra estabilidad prolongada. Una medición diferida puede evaluar retención del contenido practicado sin medir transferencia a un problema nuevo. Por ello, los cursos deberían combinar un producto asistido, una tarea individual posterior y una aplicación nueva. La evaluación conjunta permite distinguir la calidad del apoyo, la comprensión conservada y la flexibilidad de uso. También evita que la velocidad de producción o la satisfacción se conviertan en sustitutos de aprendizaje.

Heterogeneidad, equidad y pertinencia contextual

La tercera contribución es mostrar que la personalización puede reducir o ampliar diferencias. Algunos estudios describen mayores beneficios entre estudiantes inicialmente rezagados; otros encuentran ventajas concentradas en quienes ya poseen mejores fundamentos (Chen & Keng, 2026; He, Chen, et al., 2026; Omeh et al., 2025; Sami et al., 2026; Wang et al., 2026). Esta tensión resulta plausible: una explicación adaptativa puede resolver una carencia puntual, mientras la evaluación crítica de una respuesta exige conocimientos previos. La intervención necesita anticipar ambos casos mediante apoyos graduados, acceso equivalente y comprobaciones de comprensión.

La aplicabilidad a contextos latinoamericanos debe evaluarse a partir de condiciones concretas. Las experiencias de inteligencia de mercados en Colombia, programación en Chile y fisiología del ejercicio en Brasil ilustran resultados distintos (de la Puente et al., 2026; Mellado & Cubillos, 2025; Ribeiro Da Silva Junior et al., 2026). No justifican una extrapolación regional uniforme. La disponibilidad de dispositivos, conectividad, acompañamiento y materiales adecuados influye en la implementación. También importa la lengua de interacción y la pertinencia cultural del contenido. En esta revisión, todos los artículos incluidos se publicaron en inglés, aunque sus entornos educativos fueran diversos.

La amplitud disciplinar constituye una fortaleza y una limitación. Permite identificar principios transversales, pero reduce la pertinencia de un efecto promedio único. La precisión en escritura, la comunicación con un paciente y la ejecución motriz tienen exigencias diferentes. La conclusión útil es funcional: decidir qué ayuda requiere la tarea y con qué evidencia se evaluará. Esta lectura responde a la necesidad de mayor atención metodológica y contextual planteada por Bond et al. (2024), sin suponer que resultados de una disciplina se transfieran automáticamente a otra.

Alcances y limitaciones

La búsqueda se limitó a Scopus, artículos y acceso abierto, por lo que pudo omitir evidencia pertinente en otras bases, actas, literatura no publicada o publicaciones de acceso restringido. Los filtros lingüísticos tampoco garantizan representatividad geográfica. La concentración temporal en 2025 y 2026 refleja un campo reciente y restringe la evidencia sobre continuidad de los efectos. Además, las herramientas, versiones y configuraciones difieren entre estudios, y una marca comercial no identifica por sí sola una intervención reproducible.

Los diseños con clases intactas, cohortes históricas y participación voluntaria dificultan separar la intervención de diferencias previas, motivación o cambios curriculares. Las muestras pequeñas y las pérdidas también limitan la precisión. La síntesis narrativa conserva información sobre contextos y

mecanismos, aunque incorpora decisiones interpretativas de agrupación. Su fuerza reside en contrastar resultados y reconocer los límites de cada medida. Las recomendaciones deben entenderse como orientaciones condicionadas por las tareas estudiadas. Se requieren comparaciones con alternativas pedagógicas robustas, evaluaciones independientes y seguimientos que permitan examinar retención, transferencia y desigualdades. La actualización de esta evidencia será necesaria a medida que cambien los sistemas y se acumulen intervenciones más prolongadas.

Implicaciones para la práctica universitaria

Las experiencias revisadas permiten proponer una secuencia de implementación: definir el aprendizaje esperado; solicitar una actuación inicial del estudiante; ofrecer ayuda delimitada; exigir una decisión razonada sobre sus sugerencias; y comprobar el aprendizaje mediante una tarea posterior. La intensidad del acompañamiento debe ajustarse al conocimiento previo y al riesgo de error de la actividad. En habilidades clínicas, visuales y motrices, la validación experta resulta especialmente relevante. En escritura y programación, la trazabilidad de revisiones puede mostrar qué comprendió y modificó el estudiante.

La selección institucional de herramientas debería considerar la función pedagógica antes que la novedad de la plataforma. El acceso generalizado puede no añadir beneficios en ciertos cursos (Bauer et al., 2025; Kabba & Ejbari, 2026; Melanou et al., 2026; Thoeni & Fryer, 2026). Conviene comenzar con necesidades específicas, comparar con una alternativa adecuada y valorar resultados independientes. Una integración responsable conserva la interacción humana cuando aporta contexto, diálogo y juicio profesional, y utiliza la asistencia generativa allí donde amplía oportunidades de práctica verificable.

Conclusiones

Los 134 estudios incluidos muestran que la tutoría generativa puede contribuir a resultados educativos específicos en educación superior, especialmente cuando ofrece retroalimentación utilizable, práctica reiterada y apoyos alineados con la tarea. Los efectos dependen del diseño pedagógico, del comparador y del conocimiento previo. Las mejoras en productos, satisfacción y autonomía percibida no equivalen automáticamente a aprendizaje independiente.

La evidencia sobre retención y transferencia es heterogénea. Existen resultados favorables de configuraciones reflexivas y supervisadas, junto a ventajas inmediatas que se debilitan y comparaciones que favorecen a docentes o expertos. La principal condición para interpretar el valor educativo consiste en observar qué puede hacer el estudiante después de la asistencia y en situaciones nuevas. La evaluación debe conservar esa distinción desde el diseño de la intervención hasta la formulación de conclusiones.

La valoración MMAT se completó para la totalidad del corpus y los 134 estudios permanecieron en la síntesis. Sus limitaciones metodológicas orientan la interpretación, sin borrar los hallazgos discordantes. Para la docencia universitaria, la prioridad es organizar una colaboración en la que el estudiante planifique, compruebe y revise, y en la que el acompañamiento humano sostenga el juicio disciplinar. El criterio de éxito es el aprendizaje demostrado, no la cantidad de respuestas generadas.

Declaraciones editoriales

Disponibilidad de datos

Los datos y materiales que respaldan los resultados de este estudio se encuentran incluidos en el artículo y, cuando corresponde, en sus archivos suplementarios. Las solicitudes de información adicional podrán dirigirse al autor de correspondencia y estarán sujetas a los principios de confidencialidad, protección de datos y ética de la investigación.

Declaración de ética

La investigación se desarrolló de conformidad con la legislación nacional vigente, las disposiciones institucionales y los principios éticos aplicables a estudios con participación humana. Debido a su naturaleza social y no experimental, sin intervención clínica, biomédica ni manipulación de variables, no requirió la aprobación de un comité de ética. Todos los participantes recibieron información suficiente sobre los objetivos y procedimientos del estudio y otorgaron su consentimiento informado por escrito. Asimismo, se garantizó la confidencialidad, el anonimato y el uso exclusivamente académico de la información recopilada.

Contribuciones de autoría

Las contribuciones de autoría se establecieron

de acuerdo con la taxonomía CRediT:

OLTV: Conceptualización; metodología; redacción del borrador original; redacción, revisión y edición.

MNMS: Metodología; investigación; redacción del borrador original; redacción, revisión y edición.

SPCS: Investigación; redacción del borrador original; redacción, revisión y edición.

DEGN: Conceptualización; supervisión; administración del proyecto; redacción, revisión y edición.

Financiamiento

Los autores declaran que no recibieron financiamiento específico de organismos públicos, instituciones privadas ni entidades sin fines de lucro para la realización o publicación de este estudio.

Conflicto de intereses

Los autores declaran que la investigación se desarrolló en ausencia de relaciones comerciales, financieras, profesionales o personales que pudieran influir de manera inapropiada en sus resultados, interpretación o publicación.

Declaración sobre el uso de inteligencia artificial generativa

Los autores declaran que no utilizaron herramientas de inteligencia artificial generativa para producir el contenido científico, analizar o interpretar los datos, formular los resultados o elaborar las conclusiones del manuscrito. La responsabilidad sobre la originalidad, exactitud, integridad y validez académica del contenido corresponde exclusivamente a los autores.

Responsabilidad editorial

Las afirmaciones, interpretaciones y conclusiones expresadas en el artículo son responsabilidad exclusiva de sus autores y no representan necesariamente la posición institucional de ARTI EDITORIAL S.A.S., de la revista, de sus editores o de las entidades a las que se encuentran afiliados los autores.

Referencias bibliográficas

Abdelhalim, S. M., & Almaneea, M. O. (2026). GenAI-supported collaborative project-based language learning: Effects on EFL undergraduates' achievement and autonomy. *Asian-Pacific Journal of Second and Foreign Language Education*, *11*(1), 50. https://doi.org/10.1186/s40862-026-00430-8

Abdelmagid, A. S., Al-Mohaya, A. Y., Ibrahim, A. M., Teleb, A. A., & Jabli, N. M. (2025). The impact of artificial intelligence applications on developing levels of cognitive depth of information among postgraduate students. *International Journal of Learning, Teaching and Educational Research*, *24*(4), 316–333. https://doi.org/10.26803/ijlter.24.4.15

Abdelrady, A. H., Khalil, A. A., Ali El Deen, A. A. M. M., Alnofal, A. I. S., & Mohamed, A. M. (2026). Speaking with AI: The impact of Microsoft 365 Copilot voice chat on EFL learners' fluency, interactional competence, and autonomy. *Frontiers in Education*, *11*, 1851080. https://doi.org/10.3389/feduc.2026.1851080

Abdulla, S., Ismail, S., Fawzy, Y., & Elhaj, A. (2024). Using ChatGPT in teaching computer programming and studying its impact on students performance. *Electronic Journal of e-Learning*, *22*(6), 66–81. https://doi.org/10.34190/EJEL.22.6.3380

Al-khresheh, M. H., & Alruwaili, S. F. (2026). Examining ChatGPT-supported speaking practice among Saudi EFL students: A quasi-experimental study. *Humanities and Social Sciences Communications*, *13*(1), 1433. https://doi.org/10.1057/s41599-026-08106-x

Alalwi, F. S., Alqahtani, N., Jamshed, M., & Ahmmed, F. (2026). AI-driven feedback for pedagogical innovation: Assessing ChatGPT's impact on Saudi EFL learners' writing proficiency. *World Journal of English Language*, *16*(5), 197–207. https://doi.org/10.5430/wjel.v16n5p197

Albaqami, S., & Alahmadi, A. (2026). Immediate and delayed effects of ChatGPT-enhanced vocabulary instruction on Saudi EFL learners. *Information (Switzerland)*, *17*(9), 812. https://doi.org/10.3390/info17090812

Albaqami, S. E., Azim, S., Jamshed, M., & Mohammed, G. P. (2026). Improving engineering students' English proficiency through ChatGPT-assisted instruction: A quasi-experimental study. *Theory and Practice in Language Studies*, *16*(7), 2366–2374. https://doi.org/10.17507/tpls.1607.20

Aldamen, H., Almashour, M., Al-Deaibes, M., & Alsharefeen, R. (2026). Comparing AI-assisted and teacher-led reading strategy instruction in an EFL context: A quasi-experimental study. *Frontiers in Education*, *11*, 1828564. https://doi.org/10.3389/feduc.2026.1828564

Ali, Y. A., Ali, M. M., Alharbi, W., Hamid, S., & Abbas, S. (2025). The influence of ChatGPT on English grammar among Saudi ESL learners. *Arab World English Journal, 2025*(Special Issue), 358–371. https://doi.org/10.24093/awej/AI.21

Alkhalifah, E., & Almutlaq, S. (2026). The combined impact of ChatGPT and teacher feedback on the syntactic complexity of EFL learners' writing. *SAGE Open*, *16*(2), 01–12. https://doi.org/10.1177/21582440261453413

Allehyani, F., Alqahtani, N., Jamshed, M., & Warda, W. U. (2026). Measuring the impact of ChatGPT feedback on Saudi undergraduate EFL learners' writing skills. *Theory and Practice in Language Studies*, *16*(2), 670–679. https://doi.org/10.17507/tpls.1602.32

Alshammari, M. T. (2025). An investigation into ChatGPT-enhanced adaptive E-learning systems. *TEM Journal*, *14*(1), 503–510. https://doi.org/10.18421/TEM141-45

Alshehri, O. A. O., Lamouchi, A., Abdellatif, M. S., Al-Dosari, M. N. A., Mekheimer, M., & Nemt-allah, M. A. (2026). From tool to scaffold: Structured human–AI collaboration and its effects on academic writing and digital critical thinking among Saudi EFL learners. *Frontiers in Psychology*, *17*, 1–12. https://doi.org/10.3389/fpsyg.2026.1830103

Andewi, W., Waziana, W., Wibisono, D., Putra, K. A., Hastomo, T., & Oktarin, I. B. (2025). From prompting to proficiency: A mixed-methods analysis of prompting with ChatGPT versus lecturer interaction in an EFL classroom. *Journal of Studies in the English Language*, *20*(2), 210–238. https://doi.org/10.64731/jsel.v20i2.282318

Ates, H. (2026). Human-centered GenAI feedback design in higher education: A multisite experiment on direct, reflective, and hybrid approaches to scientific argumentation. *International Journal of Educational Technology in Higher Education*, *23*(1), 38. https://doi.org/10.1186/s41239-026-00614-9

Atta, A., Esmat, M., Amasha, N., Elayat, E., & Elsaid, W. K. (2025). A smart ChatGPT mobile application for improving C# programming skills for students in educational institutions. *Qubahan Academic Journal*, *5*(2), 49–62. https://doi.org/10.48161/qaj.v5n2a1772

Bauer, E., Richters, C., Pickal, A. J., Klippert, M., Sailer, M., & Stadler, M. (2025). Effects of AI-generated adaptive feedback on statistical skills and interest in statistics: A field experiment in higher education. *British Journal of Educational Technology*, *56*(5), 1735–1757. https://doi.org/10.1111/bjet.13609

Bojorque, R., Plaza, A., Morquecho, P., & Moscoso, F. (2026). A scalable and low-cost mobile RAG architecture for AI-augmented learning in higher education. *Applied Sciences (Switzerland)*, *16*(2), 963. https://doi.org/10.3390/app16020963

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, *21*, 4. https://doi.org/10.1186/s41239-023-00436-z

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, *368*, l6890. https://doi.org/10.1136/bmj.l6890

Chen, J., & Alibakhshi, G. (2026). AI-powered applications' effects on English language learners' cognitive, metacognitive, and resource management strategies, and language achievement. *Journal of Computer Assisted Learning*, *42*(1), e70171. https://doi.org/10.1002/jcal.70171

Chen, X., Liu, J., Ye, P., Li, Y., Sun, H., Li, R., Hou, Z., & Yan, J. (2026). Effects of ChatGPT-assisted teacher feedback on college students' critical thinking skills and perceptions of argumentative writing. *Journal of Computer Assisted Learning*, *42*(5), e70315. https://doi.org/10.1002/jcal.70315

Chen, Y.-H., & Keng, K. (2026). Bridging proficiency gaps: ChatGPT-enhanced STAD for achievement, attitudes, and equity in multilingual higher education. *Journal of Information Technology Education: Research*, *25*, 27. https://doi.org/10.28945/5824

Chiang, Y.-H., Tseng, H.-P., Lu, C.-H., & Huang, T.-C. (2026). Behavioral drivers of ChatGPT-assisted TOEIC reading: Evidence from an extended UTAUT model. *International Journal of Information and Education Technology*, *16*(7), 1978–1992. https://doi.org/10.18178/ijiet.2026.16.7.2659

Chu, X., Dai, H., & Zhai, X. (2026). Many hands make light work? Individual and

collaborative interaction with a GenAI-powered multi-agent system: Effects on learning performance and critical thinking. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.70089

Chun, J., Kim, J., Kim, H., Lee, G., Cho, S., Kim, C., Chung, Y., & Heo, S. (2025). A comparative analysis of on-device AI-driven, self-regulated learning and traditional pedagogy in university health sciences education. *Applied Sciences (Switzerland)*, *15*(4), 1815. https://doi.org/10.3390/app15041815

Çiçek, F. E., Ülker, M., Özer, M., & Kıyak, Y. S. (2025). ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: A randomized controlled trial. *Postgraduate Medical Journal*, *101*(1195), 458–463. https://doi.org/10.1093/postmj/qgae170

Dahri, N. A., Al-Rahmi, W. M., Alhashmi, K. A., & Bashir, F. (2025). Enhancing mobile learning with AI-powered chatbots: Investigating ChatGPT's impact on student engagement and academic performance. *International Journal of Interactive Mobile Technologies* , *19*(11), 17–38. https://doi.org/10.3991/ijim.v19i11.54643

Dai, Z. (2026). Rethinking AI-assisted writing instruction: Feedback literacy scripts, calibration training, and student writing development. *Frontiers in Psychology*, *17*, 1829268. https://doi.org/10.3389/fpsyg.2026.1829268

Daniel, R. A., V, N., BN, S., Daniel, A., & R, V. (2026). Integrating ChatGPT into knowledge-retrieval tutorials in undergraduate medical education: A prospective evaluation of higher-order learning and feasibility. *Medical Education Online*, *31*(1), 2639203. https://doi.org/10.1080/10872981.2026.2639203

de la Puente, M., Hernandez, D. P. E., Reynel, J. L. E., Pereira, J. E. F., Sanchez, D. G., & Monserrat, D. C. U. (2026). Effects of AI guided experiential learning on market intelligence skills among university students in Colombia. *Discover Education*, *5*(1), 752. https://doi.org/10.1007/s44217-026-01619-2

Deriba, F. G., Saqr, M., & Tukiainen, M. (2026). Does the impact of GenAI persist in learning? Comparing AI and instructor feedback on accessibility. *Universal Access in the Information Society*, *25*(1), 29. https://doi.org/10.1007/s10209-025-01268-6

Dinçer, S., Mavaşoğlu, M., & Kurt, Ç. (2026). Acoustic fluency patterns in AI-mediated task-based French instruction. *SAGE Open*, *16*(3). https://doi.org/10.1177/21582440261492392

Ekşi, S., Furtun, S., & Kibar Furtun, M. H. (2026). Artificial intelligence or the teacher? The effect of ChatGPT and teacher feedback on writing achievement in foreign language instruction. *Frontiers in Education*, *11*, 1860972. https://doi.org/10.3389/feduc.2026.1860972

Elshall, A. S., Badir, A., & Gebremedhin, M. A. (2026). AI-affordance alignment drives authentic learning gains without foundational erosion: A quasi-experimental pilot study from an environmental data science course. *Frontiers in Education*, *11*, 1834090. https://doi.org/10.3389/feduc.2026.1834090

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Fan, H., Du, W., Yan, J., & Jiang, H. (2026). Application of DeepSeek-based AI teaching assistant in teaching anesthesiology theories. *BMC Medical Education*, *26*(1), 138. https://doi.org/10.1186/s12909-025-08494-9

Fayid, F. M. B., & Kiraz, A. (2026). Human–machine cooperation in environmental education: Experimental evidence from AI-supported learning in higher education. *Systems*, *14*(5), 504. https://doi.org/10.3390/systems14050504

Filatova, A., Didenko, A., & Ravedovskaya, U. (2025). From lecturer to chatbot: Evaluating a hybrid teaching model. *Journal of Teaching English for Specific and Academic Purposes*, *13*(3), 529–538. https://doi.org/10.22190/JTESAP250815040F

Gasaymeh, A.-M. M., & Almohtadi, R. M. (2024). The effect of flipped interactive learning (FIL) based on ChatGPT on students' skills in a large programming class. *International Journal of Information and Education Technology*, *14*(11), 1516–1522. https://doi.org/10.18178/ijiet.2024.14.11.2182

Gokkurt Yilmaz, B. N., Ozbey, F., & Yilmaz, B. E. (2025). Effect of artificial intelligence-assisted personalized feedback on radiographic diagnostic performance of dental students: A controlled study. *BMC Medical Education*, *25*(1), 1403. https://doi.org/10.1186/s12909-025-07875-4

Gu, Z., Xi, T., & Zhang, C. (2026). DeepSeek-assisted problem-based learning for glaucoma education in an undergraduate ophthalmology clerkship: A randomized educational pilot study. *Scientific Reports*, *16*(1), 24184. https://doi.org/10.1038/s41598-026-58447-5

Guo, Y., & Lin, Z. (2025). Generative AI as a catalyst for data-driven learning: Efficacy, equity, and engagement in translation education. *SAGE Open*, *15*(4). https://doi.org/10.1177/21582440251406099

Hao, H., Razali, A. B., & Zuo, R. (2026). Exploring the impact of integrated AWE and generative AI feedback on Chinese EFL undergraduates' higher-order thinking in argumentative writing. *SAGE Open*, *16*(1). https://doi.org/10.1177/21582440251413884

He, Y., Chen, C., Yin, R., Zhang, W., Chang, H., Yang, W., Li, F., Li, X., Xia, Z., Xie, X., Huang, J., Zeng, Q., Yang, G., & Wu, J. (2026). Effect of large language model–powered virtual standardized patients on history-taking among undergraduate medical students: Propensity-matched cohort study. *JMIR Medical Education*, *12*, e92486. https://doi.org/10.2196/92486

He, Y., Niu, T., & Ashari, Z. M. (2026). Exploring the impact of multimodal generative artificial intelligence on CT dispositions in inquiry-based academic writing: A quasi-experimental analysis in higher education. *Frontiers in Education*, *11*, 1926173. https://doi.org/10.3389/feduc.2026.1926173

Hong, D.-R., Huang, C.-Y., & Gao, J. (2026). ChatGPT-4o with faculty guidance outperforms AI-only and traditional learning in ultrasonography training: A randomized trial. *Frontiers in Digital Health*, *8*, 1772965. https://doi.org/10.3389/fdgth.2026.1772965

Hong, H., Vate-U-Lan, P., & Viriyavejakul, C. (2025). Cognitive offload instruction with generative AI: A quasi-experimental study on critical thinking gains in English writing. *Forum for Linguistic Studies*, *7*(7), 325–334. https://doi.org/10.30564/fls.v7i7.10072

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285–291. https://doi.org/10.3233/EFI-180221

Hou, X., Liu, Z., Wang, G., & Shi, N. (2026). Bridging awareness and behavior: Decoding implicit metacognitive behaviors in AI-assisted programming via fine-grained log analysis. *Frontiers in Psychology*, *17*, 1790196. https://doi.org/10.3389/fpsyg.2026.1790196

Hu, H., Gong, Q., & Hashim, H. (2026). The impact of ChatGPT-generated feedback on English academic writing proficiency and engagement: A mixed-methods study. *SAGE Open*, *16*(3). https://doi.org/10.1177/21582440261480859

Huang, C., Zhong, Y., Li, Y., Wang, X., Han, Z., Zhang, D., & Liu, M. (2025). Enhancing student reading performance through a personalized two-tier problem-based learning approach with generative artificial intelligence. *Humanities and Social Sciences Communications*, *12*(1), 645. https://doi.org/10.1057/s41599-025-04919-4

Huang, G., & Chen, H. (2026). Beyond grammar: How ChatGPT-4.0 shapes writing self-efficacy and performance in CSL writing. *Language Learning and Technology*, *30*(1), 1–23. https://doi.org/10.64152/10125/73706

Huang, Y., Xu, B.-B., Wang, X.-Y., Luo, Y.-C., Teng, M.-M., & Weng, X. (2024). Implementation and evaluation of an optimized surgical clerkship teaching model utilizing ChatGPT. *BMC Medical Education*, *24*(1), 1540. https://doi.org/10.1186/s12909-024-06575-9

Huesca, G., Martínez-Treviño, Y., Molina-Espinosa, J. M., Sanromán-Calleros, A. R., Martínez-Román, R., Cendejas-Castro, E. A., & Bustos, R. (2024). Effectiveness of using ChatGPT as a tool to strengthen benefits of the flipped learning strategy. *Education Sciences*, *14*(6), 660. https://doi.org/10.3390/educsci14060660

Huesca, G., Rodriguez-Flores, T. C., Echeverria, G., Pérez-Fuertes, L., Constantino-Gonzalez, M. L. A., Martinez-Trevino, Y., Navarrete, C., Félix-Herrán, L. C., Cedillo-Hernandez, A., Ayala, A. G., & Delgado, D. A. C. (2026). Modeling factors influencing learning gains in generative artificial intelligence-supported undergraduate programming courses. *Education Sciences*, *16*(7), 1037. https://doi.org/10.3390/educsci16071037

Hussein, S. M., & Kurt, M. (2026). A multimodal human–AI instructional framework for productive vocabulary development: A classroom evaluation of a coordinated LLM–ASR system. *Systems*, *14*(5), 474. https://doi.org/10.3390/systems14050474

Ilyas, M., Khan, R. A., Yasmeen, R., Kamal, Z., & Ul Ain, N. (2026). Impact of ChatGPT-assisted

personalized learning on teaching acute abdomen to undergraduate medical students: A randomized crossover study. *Pakistan Journal of Medical Sciences*, *42*(8), 1943–1948. https://doi.org/10.12669/pjms.42.8.14827

Ironsi, C. S., & Bostanci, H. B. (2026). Enhancing the reading comprehension and autonomy of foreign language students using generative AI-integrated assistant. *Frontiers in Education*, *11*, 1763238. https://doi.org/10.3389/feduc.2026.1763238

Jamshed, M., Albedah, F., Ansari, M. S., & Banu, S. (2025). Assessing the efficacy of AI-driven corrective feedback via WhatsApp application to improve ESL learners' writing skills: An experimental study. *International Journal of Interactive Mobile Technologies* , *19*(7), 190–205. https://doi.org/10.3991/ijim.v19i07.52709

Jay-Jímenez, B. O., Martínez-Islas, D. A., Marroquin-Aguilar, A. T., Avelino-Vivas, F., Solis-Galván, D. M., Laguna-González, A. A., García-García, B. M., Minaya-Pérez, E., Quiñones-Lara, E., Martínez-Bonilla, I., Méndez-Cruz, A. R., & Saldívar-Cerón, H. I. (2026). AI-supported adaptive simulation for diagnostic disclosure in medical students: A randomized controlled trial. *International Medical Education*, *5*(2), 35. https://doi.org/10.3390/ime5020035

Jia, D., Chen, X., & Wang, Y. (2026). AI-mediated informal digital learning of English for speaking development: Longitudinal effects on ability and affect and implications for adaptive intelligence. *Journal of Intelligence*, *14*(9), 204. https://doi.org/10.3390/jintelligence14090204

Kabba, F. E., & Ejbari, Z. (2026). ChatGPT as scaffold: Quiz performance across session complexity. *International Journal of Evaluation and Research in Education*, *15*(4), 3172–3181. https://doi.org/10.11591/ijere.v15i4.39915

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kawinkoonlasate, P. (2025). A comparative study of Google Gemini and ChatGPT in enhancing English language learning for EFL learners: A case study of the English research writing course. *Pedagogical Research*, *10*(4), em0251. https://doi.org/10.29333/pr/17670

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), 17458. https://doi.org/10.1038/s41598-025-97652-6

Khusniyah, N. L., Hakim, L., & Muslimin, A. I. (2026). Comparing AI-based and peer-based feedback in teaching the CaRS model: A quasi-experimental study on postgraduate academic writing. *International Journal of Information and Education Technology*, *16*(3), 687–694. https://doi.org/10.18178/ijiet.2026.16.3.2541

Kinder, A., Briese, F. J., Jacobs, M., Dern, N., Glodny, N., Jacobs, S., & Leßmann, S. (2025). Effects of adaptive feedback generated by a large language model: A case study in teacher education. *Computers and Education: Artificial Intelligence*, *8*, 100349. https://doi.org/10.1016/j.caeai.2024.100349

Lai, C., Pan, M., Guo, K., & Cui, Y. (2026). What task to offload to GenAI in the writing feedback process? – effects of task-offloading approaches on EFL learners' writing skill development. *Computers and Education*, *252*, 105675. https://doi.org/10.1016/j.compedu.2026.105675

Lai, Z. C.-C. (2025). Enhancing EFL oral proficiency through a ChatGPT-integrated BOPPPS learning framework. *International Journal of Online Pedagogy and Course Design*, *15*(1). https://doi.org/10.4018/IJOPCD.383301

Laksana, I. P. Y., Ratminingsih, N. M., Santosa, M. H., & Kusuma, I. P. I. (2026). Aligning ChatGPT with E-portfolio assessment as EFL learning model: Its effect on student's speaking performance and feedback literacy. *Technology in Language Teaching and Learning*, *8*, 103616. https://doi.org/10.29140/tltl.2026.103616

Lee, H.-Y., Chen, P.-H., Wang, W.-S., Huang, Y.-M., & Wu, T.-T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education*, *21*(1), 16. https://doi.

org/10.1186/s41239-024-00447-4

Lee, H.-Y., Kim, J., Choi, H., Bae, H., Jeong, A., Choi, S., Kim, J.-H., & Kim, C.-E. (2025). Comparing AI chatbot simulation and peer role-play for OSCE preparation: A pilot randomized controlled trial. *BMC Medical Education*, *25*(1), 1755. https://doi.org/10.1186/s12909-025-08308-y

Lei, Q., Chen, Y., Chen, Y., Zhang, X., & Park, J. (2026). Evaluating an AI-scaffolded intervention for L2 vocabulary learning: Affordances, constraints, and pedagogical implications. *SN Social Sciences*, *6*(10), 461. https://doi.org/10.1007/s43545-026-01725-w

Li, L., Zhang, W., Zhang, K., Yang, Y., Wang, L., Zuo, L., Sun, Y., & Peng, Q. (2025). The role of generative AI tools in case-based learning and teaching evaluation of medical biochemistry. *BMC Medical Education*, *25*(1), 1185. https://doi.org/10.1186/s12909-025-07567-z

Li, Z., Li, J., & Dai, Z. (2026). Integrating GenAI chatbots into the flipped classroom to promote emotional engagement and enhance overall speaking proficiency in EFL contexts. *Discover Computing*, *29*(1), 428. https://doi.org/10.1007/s10791-026-10322-7

Lim, A., Stokes, E., Yeap, L.-L., Mazeh, A. C., Kreutzer, E., & Moore, J. (2026). Using an educator-guided generative artificial intelligence (GenAI) tool for developing communication skills in undergraduate pharmacy students. *Currents in Pharmacy Teaching and Learning*, *18*(9), 102675. https://doi.org/10.1016/j.cptl.2026.102675

Lin, C.-C., & Lin, V. (2026). Tool or tutor? A dynamic assessment-informed comparison of pedagogical AI agents and general-purpose GenAI for EFL speaking development and anxiety. *System*, *143*, 104157. https://doi.org/10.1016/j.system.2026.104157

Luan, L., Lin, X., & Dai, Y. (2025). Bridging the gap: ChatGPT'sroleinenhancingSTEMeducation. *Open Praxis*, *17*(1), 108–128. https://doi.org/10.55982/openpraxis.17.1.685

Lukešová, A., & Jennings, P. J. (2026). AI-assisted language learning: ChatGPT's impact on mastering English tenses at the university level. *Social Sciences and Humanities Open*, *13*, 102532. https://doi.org/10.1016/j.ssaho.2026.102532

Mahmud, F. A. (2025). Examining the efficacy of ChatGPT and human-derived corrective feedback in addressing grammatical errors in Saudi EFL students' compositions. *World Journal of English Language*, *16*(2), 258–270. https://doi.org/10.5430/wjel.v16n2p258

McCarrick, C. A., McEntee, P. D., Boland, P. A., Donnelly, S., O'Meara, Y., Heneghan, H., & Cahill, R. A. (2025). A randomized controlled trial of a deep language learning model-based simulation tool for undergraduate medical students in surgery. *Journal of Surgical Education*, *82*(9), 103629. https://doi.org/10.1016/j.jsurg.2025.103629

Melanou, C., Beege, M., & Kimmig, M. (2026). Generative AI and learning dynamics in higher education: A longitudinal empirical study. *Journal of Computer Assisted Learning*, *42*(5), e70322. https://doi.org/10.1002/jcal.70322

Mellado, R., & Cubillos, C. (2025). Can generative artificial intelligence outperform self-instructional learning in computer programming?: Impact on motivation and knowledge acquisition. *Applied Sciences (Switzerland)*, *15*(11), 5867. https://doi.org/10.3390/app15115867

Mohamed, A. M., Deen, A. A. M. M. A. E., Abukhait, R. O., Dahleb, F., Khan, Y. B., & Jmaiel, H. A. (2025). ChatGPT's impact on ESP writing proficiency and learner autonomy: An experimental study. *Technology in Language Teaching and Learning*, *7*(3), 102964. https://doi.org/10.29140/tltl.v7n3.102964

Mujeeb, N., Arfan Lodhi, M., Ali, A., Mashwani, W. K., Mostafa, S. M., & Ahmadini, A. A. H. (2026). Beyond red ink: AI feedback and ESL academic writing development in higher education. *Language Testing in Asia*, *16*(1), 24. https://doi.org/10.1186/s40468-026-00442-8

Nartthanarung, A., Plangsiri, K., & Kongmalai, P. (2026). Comparative impact of ChatGPT and conventional search tools on clinical reasoning performance: A randomized crossover study in preclinical medical students. *Advances in Medical Education and Practice*, *17*, 603679. https://doi.org/10.2147/AMEP.S603679

Nelson, E. C. W. (2026). Structured AI workflows in Saudi EMI research writing: Performance gains, equity, and academic integrity across three semesters. *Journal of International Students*, *16*(15), 81–102. https://doi.org/10.32674/e7jdhc67

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher*

Education, *31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Niu, Y., Xu, J., Yang, W., Yuan, H., Liu, Z., Zhang, W., & Liu, L. (2026). Application of an iterative generative AI-augmented teaching model in medical immunology education. *Frontiers in Artificial Intelligence*, *9*, 1788369. https://doi.org/10.3389/frai.2026.1788369

Oktarin, I. B., Saputri, M. E. E., Magdalena, B., Hastomo, T., & Maximilian, A. (2024). Leveraging ChatGPT to enhance students' writing skills, engagement, and feedback literacy. *Edelweiss Applied Science and Technology*, *8*(4), 2306–2319. https://doi.org/10.55214/25768484.v8i4.1600

Omeh, C. B., Ayanwale, M. A., Mnguni, L. E., & Olelewe, C. J. (2025). Fostering programming skill and critical thinking through AI-assisted PBL integration. *Journal of New Approaches in Educational Research*, *14*(1), 22. https://doi.org/10.1007/s44322-025-00041-0

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pan, M., Lai, C., Guo, K., & Lin, C.-H. (2026). Unpacking student interactions in GenAI-assisted social annotation: An epistemic network analysis of cognitive and social presence. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.70088

Pan, Y. (2025). Leveraging generative AI powered rubric-indexed feedback as a formative assessment strategy for enhancing medical English education. *Discover Computing*, *28*(1), 284. https://doi.org/10.1007/s10791-025-09830-9

Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, *11*(1), 2410101. https://doi.org/10.1080/2331186X.2024.2410101

Puteri, C. G., Cahyono, B. Y., Widiati, U., & Suryati, N. (2025). Using ChatGPT as a writing assistant: A study on essay quality development among Indonesian EFL university students. *Indonesian Journal of Applied Linguistics*, *15*(2), 332–346. https://doi.org/10.17509/3ebmy158

Qiu, Y. (2024). The impact of LLM hallucinations on motor skill learning: A case study in badminton. *IEEE Access*, *12*, 139669–139682. https://doi.org/10.1109/ACCESS.2024.3444783

Qu, Z., & Chen, L. (2026). Simulating the "cultural other": The impact of GenAI interlocutors on EFL learners' intercultural communicative competence, speaking anxiety, and learning engagement. *Frontiers in Psychology*, *17*, 1799695. https://doi.org/10.3389/fpsyg.2026.1799695

Rahman, G., Almutairi, E. A. A., Almutairi, M., & Mudhsh, B. A. (2025). The impact of AI-powered text generation on undergraduate students' critical thinking: A study on ChatGPT usage. *Journal of University Teaching and Learning Practice* , *22*(7). https://doi.org/10.53761/srkt0510

Rakha, A. H. (2026). Digital feedback via free ChatGPT within the reciprocal teaching style: Improving fundamental handball skills and students' attitudes among university beginners. *Frontiers in Sports and Active Living*, *8*, 1772502. https://doi.org/10.3389/fspor.2026.1772502

Randombage, P. J. S., Ahamed, M. Y. S., Kulathunga, T. D., & Wickramasekara, H. (2026). Effectiveness and students' perception of tutor-guided AI navigation in undergraduate medical education in Sri Lanka: A quasi-experimental study. *BMC Medical Education*, *26*(1), 1104. https://doi.org/10.1186/s12909-026-09467-2

Ren, B., Dong, Y., & Zhang, B. (2025). From errors to excellence: AI-driven scaffolding for advancing L2 writing skills. *Cogent Education*, *12*(1), 2588512. https://doi.org/10.1080/2331186X.2025.2588512

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., Koffel, J. B., & PRISMA-S Group. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, *10*, 39. https://doi.org/10.1186/s13643-020-01542-z

Ribeiro Da Silva Junior, R., Souza, L. N. d., Cavalcante, G. A., Fernandes, F. L., da Silva Júnior, J. A., Hipólito, R. G., Guzen, F., Fonseca Pinto, E., Knackfuss, M. I., & Allyrio Araújo De Medeiros Fernandes, T. (2026). AI-assisted learning in exercise physiology: A quasi-experimental study using PhysioExercise GPT. *Frontiers in Physiology*, *17*, 1894046. https://doi.org/10.3389/fphys.2026.1894046

Rother, J. F., Nissen, L., Laupichler, M. C., Raupach, T., & Sommer, N. (2026). Enhancing clinical reasoning in surgical education: A randomized crossover trial on AI-generated individual feedback for key feature question-based online assignments. *Journal of Surgical Education*, *83*(9), 104068. https://doi.org/10.1016/j.jsurg.2026.104068

Samat, C., Kwangmuang, P., Vongtathum, P., Kanjug, I., Wannasen, T., Duangngern, P., Gessala, N., Sarakan, P., Pitakaso, R., & Wu, T. T. (2026). Generative AI for adaptive curriculum design enhances student performance and engagement in technology enhanced learning environments.*DiscoverArtificialIntelligence*, *6*(1), 555. https://doi.org/10.1007/s44163-026-01264-6

Sami, A., Adkins, M., McLeod, A., Leung, F.-H., & Gilchrist, C. (2026). Artificial intelligence teaching assistants: A scalable solution for supporting struggling medical students. *Academic medicine : journal of the Association of American Medical Colleges*, *101*(10), 1265–1273. https://doi.org/10.1093/acamed/wvag208

Sarzhanova, G., & Gulnur, S. (2026). Integrating artificial intelligence into foreign language education: Review, systematization, and implementation mechanisms for academic staff. *Journal of Advanced Pharmacy Education and Research*, *16*(1), 183–190. https://doi.org/10.51847/dl9tTYezNI

Selvi, A. E., Irgatoğlu, A., Yurttadur, O., & Dağbaşı, G. (2026). Sustainable L2 writing pedagogy in Turkish higher education: Effects of AI-mediated feedback on self-regulated learning and writing performance. *PLOS One*, *21*(7), e0344618. https://doi.org/10.1371/journal.pone.0344618

Shi, H., Chai, C. S., Zhou, S., & Aubrey, S. (2026). Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer Assisted Language Learning*, *39*(4), 1148–1175. https://doi.org/10.1080/09588221.2025.2454541

Slijepcevic, N., & Yaylali, A. (2025). Leveraging "Khanmigo" generative AI-powered tool for personalized tutoring to learn scientific concepts. *Journal of Teaching and Learning*, *19*(4), 155–178. https://doi.org/10.22329/jtl.v19i4.10052

Smithsarakarn, P. (2026). AI-assisted feedback and EFL writing development: Evidence from Thai university students in an English for business course. *World Journal of English Language*, *16*(6), 323–332. https://doi.org/10.5430/wjel.v16n6p323

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199. https://doi.org/10.1177/1745691615569000

Sok, S., & Shin, H. W. (2025). Do interactions with ChatGPT influence L2 learners' oral speaking ability, summarization ability, and perceptions of generative AI tasks? *TESOL Quarterly*, *59*(S1), S19–S51. https://doi.org/10.1002/tesq.70001

Syromiatnykov, Y., Mamatov, F., Chuyanov, K., Batirov, Z., Chuyanov, D., Shomirzaev, M., Shadieva, D., Ilkhomov, K., Jo'rayeva, G., & Egamov, M. (2026). Independent engineering transfer after traceable generative-AI-assisted learning: A six-university controlled trial with deterministic cluster allocation in agricultural engineering education. *Applied Sciences (Switzerland)*, *16*(18), 8949. https://doi.org/10.3390/app16188949

Taveros, A. M., Peñaranda, J., & Mendoza, M. A. T. (2026). Agentic generative AI as a dialogic scaffold for problem-posing pedagogy. *International Journal of Learning, Teaching and Educational Research*, *25*(7), 1063–1084. https://doi.org/10.26803/ijlter.25.7.48

Thoeni, A., & Fryer, L. K. (2026). AI chatbots in higher education: Comparing expectations to evidence. *Computers in Human Behavior Reports*, *22*, 101061. https://doi.org/10.1016/j.chbr.2026.101061

Veerabhadrappa, S. K., Kumar Vadivel, J., Marya, A., Roodmal, S. Y., Kolarkodi, S. H., Tabnjh, A. K., Jain, V. K., Porntaveetus, T., Pruksametanan, A., & Selvaraj, S. (2026). A preliminary exploratory study on AI chatbot–assisted learning in enhancing dental students comprehension of intraoral radiographic techniques. *Frontiers in Education*, *11*, 1853688. https://doi.org/10.3389/feduc.2026.1853688

Vo, T. L., & Ho, T. M. P. (2026). A quasi-experimental study on enhancing EFL learners' grammar precision in writing skills under supported-ChatGPT approach at a university in a remote region of Vietnam. *Arab World English Journal*, *2026-January*(Special Issue 3), 361–375. https://doi.org/10.24093/awej/A3.23

Wang, W.-S., Lin, C.-J., Huang, Y.-M., & Wu, T.-T. (2026). Boosting self-directed learning

in virtual reality: The role of GPT-driven feedback for low achievers. *Education and Information Technologies*, *31*(14), 5423–5459. https://doi.org/10.1007/s10639-026-14015-y

Wang, X., Zheng, Z., Zhang, J., Hou, X., & Zhu, Z. (2026). The IDEA framework for metacognitively regulated GenAI use in higher education: Development and exploratory pilot evidence. *Computers and Education: Artificial Intelligence*, *11*, 100657. https://doi.org/10.1016/j.caeai.2026.100657

Wang, Z., Fan, T.-T., Li, M.-L., Zhu, N.-J., & Wang, X.-C. (2025). Feasibility study of using GPT for history-taking training in medical education: A randomized clinical trial. *BMC Medical Education*, *25*(1), 1030. https://doi.org/10.1186/s12909-025-07614-9

Wen, Y., Wen, C., Huang, S., Chong, Y., & Yang, D. (2026). Effects of generative AI-supported virtual reality dental skill training on learning performance, visual behavior, and cortical activation among dental students: Stratified randomized controlled trial. *Journal of medical Internet research*, *28*, e90938. https://doi.org/10.2196/90938

Wiboolyasarin, W., Chomchuen, F., Kamonsawad, R., Boonprawes, S., Wiboolyasarin, K., & Jinowat, N. (2026). Timing is everything: Investigating the impact of immediate, delayed, and combined AI-generated feedback on L2 writing proficiency. *Social Sciences and Humanities Open*, *13*, 102790. https://doi.org/10.1016/j.ssaho.2026.102790

Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence*, *6*, 100228. https://doi.org/10.1016/j.caeai.2024.100228

Xiao, Z., & Guo, W. (2026). Generative artificial intelligence in university sports training: A mixed-methods study. *International Journal of Information and Education Technology*, *16*(5), 1153–1160. https://doi.org/10.18178/ijiet.2026.16.5.2584

Xu, H., Sun, Y., Mo, Z.-W., Man, Y.-N., Liu, S.-M., Wei, Y.-N., & He, M.-L. (2026). A comparative evaluation of ChatGPT-assisted and traditional methods of teaching diagnostic assessments to medical students: A randomized controlled study. *Medicine (United States)*, *105*(36), e50512. https://doi.org/10.1097/MD.0000000000050512

Yang, Y., Huang, L., Lin, W., Li, Y., Xu, Y., & Cheng, L. (2025). Enhancing sustainable English writing instruction through a generative AI-based virtual teacher within a co-regulated learning framework. *Sustainability (Switzerland)*, *17*(19), 8770. https://doi.org/10.3390/su17198770

Yang, Y., Hussin, S., & Hashim, H. (2025). Revising with intelligence: ChatGPT feedback and its impact on EFL students' revision and self-efficacy. *Forum for Linguistic Studies*, *7*(7), 352–367. https://doi.org/10.30564/fls.v7i7.9845

Yang, Y., Zhang, H., & Jiang, Y. (2025). An intelligent teaching assistant system for enhanced online engineering education: A dual-teacher model. *Computer Applications in Engineering Education*, *33*(6), e70098. https://doi.org/10.1002/cae.70098

Yuan, B. (2026). Adaptive versus fixed metacognitive prompting in GenAI-assisted EFL writing: Effects on self-regulated learning and writing performance. *SAGE Open*, *16*(3). https://doi.org/10.1177/21582440261494288

Yusof, I. J. (2025). ChatGPT-assisted retrieval practice and exam scores: Does it work? *Journal of Information Technology Education: Research*, *24*, 8. https://doi.org/10.28945/5474

Zargaran, Z., & Aldhahli, S. S. B. (2026). A stage-specific affordance model of ChatGPT deployment in EFL writing: Criterion-level evidence from a quasi-experimental design. *Journal of International Students*, *16*(17), 339–358. https://doi.org/10.32674/ggjnen46

Zhang, J., Xu, B., Liu, Y., Cai, W., Wu, L., Qi, S., Peng, Y., Yu, J., Long, T., & Ge, B. (2026). Exploring ChatGPT's potential in dialogic teaching: A comparative neurocognitive study of AI and human instruction. *Acta Psychologica*, *268*, 107268. https://doi.org/10.1016/j.actpsy.2026.107268

Zhao, C., Zhu, J., Liu, J., Zhao, W., & Pang, Y. (2026). Effectiveness of a generative AI-powered digital tutor integrated with a knowledge graph in anatomy education for nursing students: A randomized controlled trial. *BMC Medical Education*, *26*(1), 1026. https://doi.org/10.1186/s12909-026-09469-0

Zheldibayeva, R. (2025). GenAI as a learning buddy for non-English majors: Effects on listening and writing performance. *Educational Process: International Journal*, *14*, e2025051. https://doi.org/10.22521/edupij.2025.14.51

Zhifang, X., & Wentao, G. (2026). Research on the innovative integration of scenario creation and generative artificial intelligence in college physical fitness classes. *Multidisciplinary Science Journal*, *8*(8), e2026549. https://doi.org/10.31893/multiscience.2026549

Zhou, Q., Hashim, H., & Sulaiman, N. A. (2025). Integrating AI chatbots in informal digital English learning: Impacts on listening competencies in Chinese higher education. *Education and Information Technologies*, *30*(18), 27031–27059. https://doi.org/10.1007/s10639-025-13811-2

Zhou, X., Zhao, S., Wu, P., & Chen, Y. (2026). System design and evaluation of RAG-enhanced digital humans in design education: Analyzing cognitive load and instructional efficiency. *Applied Sciences (Switzerland)*, *16*(2), 1068. https://doi.org/10.3390/app16021068

Zhu, Z., Luo, C., & Li, L. (2026). Adaptive teaching assistance model combining generative AI and big data analytics. *Frontiers in Psychology*, *17*, 1–21. https://doi.org/10.3389/fpsyg.2026.1844703