

NEUROEDUCACIÓN Y APRENDIZAJE EN CONTEXTO CONTEMPORÁNEOS:

PROCESOS COGNITIVOS Y TECNOLOGÍA EDUCATIVA
DESDE LA EVIDENCIA CIENTÍFICA



Autores

Dra. Nathalie Chacón Hurtado

Dr. Angel Onzari Luna Santos

Mg. Jenny Margot Cáceres Soto

Mg. Susana Pilar Heredia García

Dra. Socorro Lecarnaqué Sánchez

Mg. Vania Nicole Milagros Bazán Chávez

Mg. Patricia Roxana Ibáñez Almonacid

Mg. Gregoria Dueñas Huaman



Edición: Primera Edición 2026

ISBN: 978-9907-834-01-7

Editorial: ALEMA Casa Editora-Editorial Internacional S.A.S.D

Ciudad, País: Jipijapa – Ecuador.

URL: <https://editorialalema.org/libros/index.php/alema>

Diseño y diagramación:

Mgtr. Wilter Leonel Solórzano Álava

Corrección de contenidos:

Dr. C. Omar Mar Cornelio

Diseño, montaje y producción editorial:

ALEMA Casa Editora-Editorial Internacional S.A.S.D, Ecuador

Hecho en Ecuador, Made in Ecuador

Este texto ha sido sometido a un proceso de evaluación por pares externos.

Advertencia: “Quedan todos los derechos reservados. Se prohíbe la reproducción, el registro o la transmisión parcial o total de esta obra por cualquier sistema de recuperación de información existente o por existir, sin el permiso previo por escrito del titular de los derechos correspondientes”.

ISBN: 978-9907-834-01-7



NEUROEDUCACIÓN Y APRENDIZAJE EN CONTEXTOS CONTEMPORÁNEOS: PROCESOS COGNITIVOS Y TECNOLOGÍA EDUCATIVA DESDE LA EVIDENCIA CIENTÍFICA

AUTORES:

Nathalie Chacón Hurtado (Perú)

Dra en educación, Magíster Scientiarum en Gerencia Empresarial, Ingeniera Industrial y Licenciada en Administración mención Gerencia, con sólida formación académica y enfoque estratégico orientado a resultados. Cuenta con amplia experiencia en dirección y gestión operativa en sectores educativos e industriales, liderando procesos de mejora continua, productividad y aseguramiento de la calidad. Ha desempeñado cargos de alto impacto como CEO, directora Académica, Gerente de Proyectos, Gerente Técnico y Gerente de Operaciones, fortaleciendo la planificación estratégica y la optimización de recursos. Cuenta con trayectoria como docente de postgrado en la Universidad César Vallejo, la Universidad Tecnológica del Perú (UTP) y la Escuela de Postgrado Newman como director de tesis, participando en la formación de profesionales en investigación, gestión y educación. Ha contribuido con publicaciones científicas y ponencias internacionales, además de desempeñarme como revisora en revistas académicas indexadas, autora de libros y artículos científicos, capítulos de libros y patentes de invención. Ganadora del premio al profesional exitoso, orgullo peruano año 2024, por su labor destacada como docente universitario.

E-mail: nathalie@ucvvirtual.edu.pe

ORCID: <https://orcid.org/0000-0002-6541-893X>

Filiación: Universidad César Vallejo, Perú

Dr. Angel Onzari Luna Santos (Perú)

Doctor en Contabilidad por la Universidad Nacional José Faustino Sánchez Carrión, Maestro en Administración de la Educación por la Universidad César Vallejo, Contador Público de la Universidad Nacional José Faustino Sánchez Carrión, Licenciado en Educación con mención en Administración y Contabilidad de la Universidad Privada San Pedro de Chimbote, Docente de Pregrado y Posgrado de la Facultad de Ciencias Económicas, Contables y Financieras de la Universidad Nacional José Faustino Sánchez Carrión.

E-mail: aluna@unjfsc.edu.pe

ORCID: <https://orcid.org/0000-0001-6751-504X>

Filiación: Universidad Nacional José Faustino Sánchez Carrión

Mg. Jenny Margot Cáceres Soto (Perú)

Licenciada en Educación secundaria en la especialidad de Idiomas Extranjeros Inglés – francés por la Universidad Nacional Daniel Alcides Carrión. Maestría en Psicología Educativa y Segunda Especialidad en Neuroeducación por la Universidad Privada Cesar Vallejo. Diplomado en Desarrollo del Pensamiento Crítico en Comprensión Lectora en la Universidad Nacional Mayor de San Marcos.

E-mail: jmcaceres@ucvvirtual.edu.pe

ORCID: <https://orcid.org/0009-0006-8950-8071>

Filiación: Universidad César Vallejo, Perú

Mg. Susana Pilar Heredia García (Perú)

Profesora de educación secundaria especializada en historia y geografía, egresada del Instituto Superior Pedagógico de Huaraz (1995), bachiller en educación por la Universidad Marcelino Champagnat (2000), magíster en educación con mención en docencia y gestión educativa por la Universidad César Vallejo (2015) y posee una segunda especialidad en neuroeducación por la Universidad César Vallejo (2025). Cuenta con diplomado en gestión escolar con liderazgo pedagógico por la Universidad Nacional de Trujillo (2017 y 2014), diplomado en currículum nacional en EBR, EBA Y EBE por la Universidad Nacional de Trujillo (2022), entre otros.

E-mail: spheredia@ucvvirtual.edu.pe

ORCID: <https://orcid.org/0009-0002-9609-9844>

Filiación: Universidad César Vallejo, Perú

Dra. Socorro Lecarnaqué Sánchez (Perú)

Licenciada en Educación Inicial por la Universidad Nacional de Piura con segunda especialidad en Neuroeducación por la Universidad César Vallejo. Magíster en Docencia Universitaria por la universidad Nacional de Piura y estudios concluidos de Doctorado en Ciencias de la Educación por la misma universidad. Se desempeña como directora de la Institución Educativa particular Pequeñitos, que forma a niños y niñas de 2 a 5 años. También es docente universitaria hace más de 17 años en prestigiosas universidades de Piura. Actualmente dicta en la escuela de Educación inicial en la Universidad Tecnológica del Perú (UTP) y también en la Escuela de Inicial de la Universidad César Vallejo (UCV).

E-mail: jdlecarnque@ucvvirtual.edu.pe

ORCID: <https://orcid.org/0009-0001-7265-5925>

Filiación: Universidad César Vallejo, Perú

Mg. Vania Nicole Milagros Bazán Chávez: Licenciada en Educación Inicial por la Universidad Privada Antenor Orrego y Magíster en Educación con mención en Didáctica de la Educación Superior por la misma casa de estudios. Es Especialista en Neuroeducación por la Universidad César Vallejo (título en trámite) y cuenta con formación en Métodos de Investigación y Publicaciones Académicas por la Universidad Autónoma de Chile. Se desempeña como docente del Nivel Inicial en el Colegio Privado Bilingüe La Asunción, promoviendo la autonomía, la creatividad y el desarrollo integral en la primera infancia. Es autora del capítulo de investigación “Relación entre práctica docente, inteligencia emocional y autonomía en estudiantes de Educación Inicial de una universidad privada”, publicado en el libro internacional Exploraciones y descubrimientos en educación: resultados de investigación para el cambio (ISBN 978-628-97217-1-3), tras evaluación por pares académicos bajo el sistema de doble ciego. Ha participado como ponente en jornadas de investigación de posgrado y en eventos académicos vinculados a innovación y sostenibilidad educativa.

E-mail: nbazanch23@ucvvirtual.edu.pe

ORCID: <https://orcid.org/0009-0009-2970-2253>

Filiación: Universidad César Vallejo, Perú

Mg. Patricia Roxana Ibáñez Almonacid: Licenciada en Obstetricia de la UNMSM y Licenciada en Educación Secundaria especialidad de Informática y Computación Universidad Particular Garcilaso de la Vega, Magíster en Pedagogía Virtual de la Universidad Católica de Trujillo. Es Especialista en Neuroeducación por la Universidad César Vallejo y ganadora del Concurso Maestra Digital Perú del año 2018. Se desempeña como docente del Nivel Secundario en la I.E.E. Nro. 1070 Melitón Carvajal en la especialidad de Informática y Cómputo, como especialista de Google Microsoft Office 365 prepara a estudiantes a la Certificación Internacional de MS, trabajando con realidad aumentada e inteligencia artificial para el emprendimiento de sus estudiantes, haciéndola ganadora del concurso Crea y Emprende a nivel de Ugel 03 en la categoría diseño industrial año 2024.

E-mail: pribanez@ucvvirtual.edu.pe

ORCID: <https://orcid.org/0000-0003-1142-9235>

Filiación: Universidad César Vallejo, Perú

Mg. Gregoria Dueñas Huaman: Inició de formación académica en el I.S.P.P. Nuestra Señora del Rosario de Madre de Dios docente en Educación Inicial, actualmente culminando la Licenciatura en Educación Inicial en la Universidad Privada de Trujillo (UPRIT). Asimismo, segunda especialidad en Neuroeducación por la Universidad César Vallejo, En el sector financiero y legal, es Contadora Pública por la Universidad Privada Telesup, teniendo el grado de Magíster en Tributación por la Escuela de Postgrado de Newman de Tacna. con especialización en Auditoría Externa, reconocida por el Colegio de Contadores Públicos de Madre de Dios. Se desempeña en la Administración de Recursos Humanos y la gestión administrativa de instituciones de prestigio. Su labor asegura el estricto cumplimiento de las obligaciones tributarias y la normativa laboral vigente, manteniendo una relación eficiente ante organismos reguladores como Sunat, Sunafil y el Ministerio de Trabajo. en I.E.P. Talentos Madre de Dios * I.E.P. Emanuel. teniendo experiencia desde el 2015.

E-mail: duenhuagr@ucvvirtual.edu.pe

ORCID: <https://orcid.org/0009-0008-7150-2936>

Filiación: Universidad César Vallejo, Perú

RESUMEN

El aprendizaje contemporáneo se configura como un fenómeno complejo, influido por la interacción de variables cognitivas, emocionales y tecnológicas, lo que ha evidenciado las limitaciones de los modelos pedagógicos tradicionales centrados en la transmisión de contenidos. En este contexto, la neuroeducación emerge como un enfoque integrador que articula aportes de la neurociencia, la psicología cognitiva y la educación para comprender los procesos de aprendizaje desde el funcionamiento cerebral. El presente libro tiene como objetivo analizar cómo distintos factores neuroeducativos, cognitivos y tecnológicos inciden en el aprendizaje en diversos niveles educativos, a partir de estudios empíricos y revisiones sistemáticas.

La obra se estructura en siete capítulos: El primer capítulo presenta una revisión sistemática sobre el efecto de los chatbots educativos en la motivación y el compromiso estudiantil desde la perspectiva de la neuroeducación. El segundo capítulo desarrolla una revisión sistemática del uso de la inteligencia artificial generativa en los procesos de aprendizaje, abordada desde un enfoque neuroeducativo. El tercer capítulo examina la relación entre la calidad del sueño y su influencia en el rendimiento académico de los estudiantes de secundaria. El cuarto capítulo analiza la asociación entre la neuroeducación y el aprendizaje multisensorial en estudiantes del nivel inicial. El quinto capítulo explora el efecto de la música clásica en la neuroplasticidad y el aprendizaje de la lectoescritura en niños. El sexto capítulo estudia la relación entre la neuroeducación y el uso de tecnologías educativas con realidad aumentada. Finalmente, el séptimo capítulo aborda la influencia del aula virtual y su articulación con la neuroeducación en los estudiantes del nivel inicial.

En conjunto, los hallazgos evidencian que el aprendizaje es un proceso multifactorial, en el que variables biológicas, cognitivas y tecnológicas actúan de manera interdependiente. Se identifican efectos diferenciados según el nivel educativo y el tipo de intervención, así como la necesidad de enfoques integradores basados en evidencia.

El libro aporta una síntesis articulada que contribuye al desarrollo de prácticas educativas más informadas, contextualizadas y coherentes con los avances de la neuroeducación.

Palabras clave: Neuroeducación; aprendizaje; inteligencia artificial; chatbots educativos; realidad aumentada; neuroplasticidad.

ABSTRACT

Contemporary learning is a complex phenomenon shaped by the interplay of cognitive, emotional, and technological variables, which has highlighted the limitations of traditional pedagogical models focused on the transmission of content. In this context, neuroeducation emerges as an integrative approach that brings together insights from neuroscience, cognitive psychology, and education to understand learning processes from the perspective of brain function. This book aims to analyze how various neuroeducational, cognitive, and technological factors influence learning at different educational levels, based on empirical studies and systematic reviews.

The book is structured into seven chapters: The first chapter presents a systematic review of the effect of educational chatbots on student motivation and engagement from a neuroeducational perspective. The second chapter presents a systematic review of the use of generative artificial intelligence in learning processes, approached from a neuroeducational perspective. The third chapter examines the relationship between sleep quality and its influence on the academic performance of secondary school students. The fourth chapter analyzes the association between neuroeducation and multisensory learning in early childhood students. The fifth chapter explores the effect of classical music on neuroplasticity and literacy learning in children. The sixth chapter studies the relationship between neuroeducation and the use of educational technologies with augmented reality. Finally, the seventh chapter addresses the influence of the virtual classroom and its integration with neuroeducation in early childhood students.

Taken together, the findings demonstrate that learning is a multifactorial process in which biological, cognitive, and technological variables act interdependently. Differentiated effects are identified based on educational level and type of intervention, as well as the need for evidence-based, integrative approaches.

The book provides a coherent synthesis that contributes to the development of educational practices that are more informed, contextualized, and consistent with advances in neuroeducation.

Keywords: Neuroeducation; learning; artificial intelligence; educational chatbots; augmented reality; neuroplasticity.

ÍNDICE

RESUMEN.....	VII
<i>ABSTRACT</i>	VIII
ÍNDICE	IX
CAPÍTULO 1 REVISIÓN SISTEMÁTICA DEL EFECTO DE LOS CHATBOTS EDUCATIVOS EN LA MOTIVACIÓN Y EL COMPROMISO DESDE LA PERSPECTIVA DE LA NEUROEDUCACIÓN.....	5
CAPÍTULO 2 REVISIÓN SISTEMÁTICA DEL USO DE INTELIGENCIA ARTIFICIAL GENERATIVA EN PROCESOS DE APRENDIZAJE DESDE UN ENFOQUE NEUROEDUCATIVO.....	70
CAPÍTULO 3 LA CALIDAD DEL SUEÑO Y SU INFLUENCIA EN EL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES DE SECUNDARIA.....	138
CAPÍTULO 4 LA NEUROEDUCACIÓN Y SU ASOCIACIÓN EN EL APRENDIZAJE MULTISENSORIAL EN LOS ESTUDIANTES DEL NIVEL INICIAL.....	156
CAPÍTULO 5 EFECTO DE LA MÚSICA CLÁSICA EN LA NEUROPLASTICIDAD Y APRENDIZAJE DE LECTOESCRITURA EN NIÑOS	183
CAPÍTULO 6 RELACIÓN DE LA NEUROEDUCACIÓN Y EL USO DE TECNOLOGÍAS EDUCATIVAS CON REALIDAD AUMENTADA.....	204
CAPÍTULO 7 AULA VIRTUAL Y LA NEUROEDUCACIÓN EN LOS ESTUDIANTES DE NIVEL INICIAL.....	225
CONCLUSIONES GENERALES	256

INTRODUCCIÓN

En las últimas décadas, los sistemas educativos a nivel global han experimentado transformaciones profundas impulsadas por la digitalización, la expansión del conocimiento científico y la creciente complejidad de los contextos socioculturales en los que ocurre el aprendizaje. Este escenario ha puesto en evidencia una brecha persistente entre los modelos pedagógicos tradicionales —centrados en la transmisión de contenidos— y las demandas contemporáneas de formación, que exigen el desarrollo de competencias cognitivas complejas, pensamiento crítico, autorregulación y adaptabilidad. En este contexto, el aprendizaje deja de ser entendido como un proceso lineal y homogéneo para concebirse como un fenómeno multidimensional, en el que convergen variables cognitivas, emocionales, biológicas y tecnológicas. Esta complejidad plantea la necesidad de enfoques integradores capaces de articular distintos niveles de análisis y de ofrecer explicaciones más completas sobre cómo aprenden los estudiantes en escenarios educativos reales.

La neuroeducación emerge precisamente como un campo interdisciplinario orientado a responder a esta necesidad. Al integrar aportes de la neurociencia, la psicología cognitiva y las ciencias de la educación, este enfoque propone comprender los procesos de enseñanza-aprendizaje a partir del funcionamiento del cerebro, traduciendo hallazgos científicos en principios pedagógicos aplicables. Desde esta perspectiva, el aprendizaje no se limita a la adquisición de información, sino que implica modificaciones estructurales y funcionales en el sistema nervioso, enmarcadas en procesos de neuroplasticidad. Asimismo, la neuroeducación enfatiza el papel de variables como la emoción, la atención y la motivación como condiciones necesarias para el aprendizaje significativo, desafiando visiones reduccionistas que han predominado en la práctica educativa. De este modo,

se configura como un marco teórico con potencial para redefinir tanto el diseño instruccional como la evaluación de los aprendizajes.

El análisis de los procesos cognitivos constituye un componente central dentro de este enfoque. La atención, la memoria y las funciones ejecutivas operan como sistemas interdependientes que permiten la codificación, el almacenamiento y la recuperación de la información. En particular, la memoria —en sus distintos sistemas— juega un papel determinante en la consolidación del aprendizaje, proceso mediante el cual las experiencias se transforman en representaciones estables. La evidencia científica ha mostrado que este proceso no ocurre de manera aislada, sino que se encuentra influido por factores biológicos y contextuales, como la calidad del sueño, la estimulación sensorial y el entorno de aprendizaje. Asimismo, la neuroplasticidad permite explicar cómo el cerebro se reorganiza en respuesta a la experiencia, facilitando la adquisición de nuevas habilidades y conocimientos a lo largo del tiempo. Desde esta perspectiva, el rendimiento académico no puede entenderse únicamente como resultado de la instrucción formal, sino como la manifestación de procesos cognitivos complejos mediados por múltiples variables.

En paralelo, el desarrollo de la tecnología educativa ha introducido nuevas formas de interacción con el conocimiento, modificando las condiciones bajo las cuales se produce el aprendizaje. La incorporación de herramientas como la realidad aumentada, los sistemas basados en inteligencia artificial y los entornos digitales interactivos ha ampliado las posibilidades pedagógicas, permitiendo diseñar experiencias de aprendizaje más dinámicas, personalizadas y multisensoriales. Sin embargo, el impacto de estas tecnologías no es intrínsecamente positivo; su efectividad depende de la forma en que se integran en el proceso educativo y de su alineación con principios cognitivos y neuroeducativos. En este sentido, la tecnología puede actuar como un mediador que potencia procesos como la atención, la motivación y la comprensión, o, por el contrario, generar

sobrecarga cognitiva y distracción si no se utiliza de manera adecuada. Por ello, resulta fundamental analizar el papel de la tecnología no como un fin en sí mismo, sino como un componente articulado dentro de un enfoque pedagógico basado en evidencia.

La literatura científica reciente ha documentado una amplia variedad de factores que influyen en los procesos de aprendizaje desde una perspectiva neuroeducativa. Entre ellos, la calidad del sueño ha sido identificada como un elemento clave en la consolidación de la memoria y el funcionamiento cognitivo, aunque los hallazgos empíricos no siempre son consistentes, lo que sugiere la presencia de variables mediadoras. De igual forma, el aprendizaje multisensorial ha mostrado efectos positivos en el desarrollo cognitivo temprano, particularmente en contextos de educación inicial, al favorecer la integración de múltiples canales sensoriales. En el ámbito de la intervención pedagógica, se han explorado estrategias como el uso de la música en procesos de lectoescritura, evidenciando su potencial para estimular la neuroplasticidad. Por otro lado, el uso de tecnologías emergentes —como la realidad aumentada, los chatbots educativos y la inteligencia artificial generativa— ha sido objeto de creciente atención en la investigación, destacándose sus efectos sobre la motivación, el compromiso y la personalización del aprendizaje. No obstante, a pesar del crecimiento sostenido de la evidencia, persisten limitaciones importantes, entre ellas la fragmentación del conocimiento, la heterogeneidad metodológica y la escasez de estudios integradores en contextos específicos.

En este marco, el presente libro se propone contribuir a la comprensión de los procesos de aprendizaje desde una perspectiva integradora que articule factores neuroeducativos, cognitivos y tecnológicos. Su objetivo es analizar, a partir de evidencia empírica y revisiones sistemáticas, cómo distintos elementos —como el sueño, la estimulación multisensorial, la música y las tecnologías

educativas— influyen en el aprendizaje en diversos niveles educativos. Para ello, la obra se estructura en seis capítulos que abordan estas temáticas desde enfoques complementarios.

El primer capítulo presenta una revisión sistemática sobre el efecto de los chatbots educativos en la motivación y el compromiso estudiantil desde la perspectiva de la neuroeducación. El segundo capítulo desarrolla una revisión sistemática del uso de la inteligencia artificial generativa en los procesos de aprendizaje, abordada desde un enfoque neuroeducativo. El tercer capítulo examina la relación entre la calidad del sueño y el rendimiento académico en estudiantes de educación secundaria. El cuarto capítulo analiza la asociación entre estrategias neuroeducativas y el aprendizaje multisensorial en estudiantes del nivel inicial. El quinto capítulo explora el efecto de la música clásica en la neuroplasticidad y el aprendizaje de la lectoescritura en niños. El sexto capítulo estudia la relación entre la neuroeducación y el uso de tecnologías educativas con realidad aumentada. Finalmente, el séptimo capítulo aborda la influencia del aula virtual y su articulación con la neuroeducación en estudiantes del nivel inicial.

En conjunto, esta obra busca no solo sistematizar el conocimiento existente, sino también ofrecer una visión articulada que permita comprender la complejidad del aprendizaje en el contexto educativo contemporáneo. Al integrar diferentes niveles de análisis —desde los procesos neurocognitivos hasta las mediaciones tecnológicas—, el libro aspira a constituirse en un aporte relevante tanto para la investigación como para la práctica educativa, contribuyendo al desarrollo de enfoques pedagógicos más informados, efectivos y contextualizados.

Capítulo 1

Revisión sistemática del efecto de los chatbots educativos en la motivación y el compromiso desde la perspectiva de la neuroeducación

Nathalie Chacón Hurtado¹

Resumen

El presente estudio tuvo como objetivo analizar sistemáticamente la evidencia empírica reciente sobre el efecto de los chatbots educativos en la motivación y el compromiso estudiantil desde una perspectiva neuroeducativa. Se desarrolló una revisión sistemática con síntesis narrativa temática, siguiendo lineamientos PRISMA 2020. La búsqueda se centró en estudios empíricos publicados entre 2020 y 2026 sobre chatbots educativos, agentes conversacionales, tutores virtuales o sistemas afines utilizados con intención formativa. Se incluyeron estudios cuantitativos, cualitativos y mixtos que evaluaran motivación, engagement o constructos próximos, como satisfacción, autoeficacia, confianza, disposición a comunicarse e interés de aprendizaje. La evidencia revisada muestra una tendencia predominantemente favorable hacia el uso de chatbots educativos para fortalecer variables motivacionales, afectivas y conductuales, especialmente cuando las intervenciones ofrecen retroalimentación inmediata, práctica segura, personalización, presencia social y apoyo a la autorregulación. Sin embargo, los efectos son menos consistentes en rendimiento académico, aprendizaje profundo, transferencia y resultados longitudinales. La calidad metodológica del corpus fue heterogénea, con riesgos frecuentes asociados a diseños no

¹ Docente Universidad César Vallejo, Perú. E-mail: nathalie@ucvvirtual.edu.pe, ORCID: <https://orcid.org/0000-0002-6541-893X>

aleatorizados, muestras pequeñas, autoinformes y limitada disponibilidad de datos. Desde una lectura neuroeducativa prudente, los chatbots parecen actuar sobre condiciones psicológicas compatibles con el aprendizaje, como autonomía, competencia, emoción, feedback y autorregulación, aunque la evidencia neurobiológica directa sigue siendo limitada.

Palabras clave: chatbots educativos, motivación académica, engagement estudiantil, neuroeducación, inteligencia artificial educativa, revisión sistemática

1. Introducción

1.1. Relevancia del problema: IA conversacional, motivación y engagement en educación

La inteligencia artificial conversacional se ha incorporado progresivamente a los entornos educativos como parte de una transformación más amplia de las formas de interacción, retroalimentación y apoyo al aprendizaje. Los chatbots, agentes conversacionales y tutores virtuales han dejado de ser aplicaciones periféricas para convertirse en recursos utilizados en tareas de enseñanza, orientación académica, práctica guiada, apoyo personalizado y acompañamiento del estudiante. A diferencia de otras tecnologías educativas centradas principalmente en la distribución de contenidos, los sistemas conversacionales introducen una dimensión interactiva que puede modificar la forma en que los estudiantes solicitan ayuda, reciben retroalimentación, practican habilidades y regulan su proceso de aprendizaje. Revisiones recientes muestran que los chatbots educativos han sido estudiados en múltiples campos, plataformas y roles pedagógicos, incluyendo enseñanza, tutoría, apoyo entre pares, aprendizaje de lenguas, informática, ingeniería y educación general (Kuhail et al., 2023; Schei et al., 2024; Debets et al., 2025).

La relevancia de esta expansión no radica únicamente en la novedad técnica de los sistemas conversacionales, sino en su posible relación con variables centrales del aprendizaje. La motivación y el engagement estudiantil constituyen condiciones críticas para la participación, la persistencia,

la autorregulación y la disposición del estudiante a sostener esfuerzos cognitivos en tareas académicas. Desde esta perspectiva, los chatbots educativos resultan especialmente relevantes porque prometen intervenir en dimensiones que no son puramente instrumentales: pueden ofrecer feedback inmediato, práctica repetida, interacción flexible, apoyo a la autonomía, sensación de acompañamiento y oportunidades para fortalecer la percepción de competencia. La literatura reciente ha señalado que la motivación es crucial para casi todos los aspectos de la experiencia de aprendizaje, pero también que la forma en que los chatbots apoyan la motivación estudiantil continúa siendo un campo en desarrollo (Huang et al., 2025; Yin et al., 2021; Septiyanti et al., 2024).

No obstante, la promesa educativa de los chatbots no debe confundirse con una evidencia uniforme de eficacia. Aunque diversos estudios y revisiones reportan beneficios potenciales en aprendizaje, interés, participación o percepción de utilidad, los efectos sobre motivación y engagement parecen depender del diseño de la intervención, la calidad pedagógica del feedback, el tipo de interacción, la duración del uso, el contexto educativo y las características de los estudiantes. Deng y Yu (2023) encontraron efectos globales favorables de la tecnología chatbot sobre resultados de aprendizaje, pero la literatura posterior ha mostrado que los efectos motivacionales no siempre son lineales ni equivalentes entre contextos. De manera similar, Huang et al. (2025) advierten que la relación entre chatbots y motivación estudiantil requiere mayor clarificación, mientras que Debets et al. (2025) señalan que muchas implementaciones siguen siendo evaluadas principalmente mediante factores perceptuales y con fundamentos teóricos desiguales.

1.2. Delimitación conceptual: qué se entiende por chatbot educativo

En este artículo, el concepto de **chatbot educativo** se entiende de manera amplia, pero no indiscriminada. Se refiere a sistemas conversacionales utilizados con finalidad formativa para

apoyar procesos de enseñanza, aprendizaje, tutoría, práctica, retroalimentación, orientación académica, simulación o acompañamiento del estudiante. Esta delimitación incluye denominaciones afines como agentes conversacionales, *pedagogical agents*, sistemas de diálogo, tutores virtuales, *AI tutors* y asistentes conversacionales cuando cumplen una función pedagógica explícita. Esta amplitud conceptual resulta necesaria porque la literatura del campo utiliza una terminología inestable: investigaciones funcionalmente próximas pueden aparecer bajo etiquetas distintas, mientras que el uso exclusivo del término *chatbot* podría reducir la sensibilidad de una revisión sistemática.

Esta delimitación también exige diferenciar los chatbots educativos de la inteligencia artificial educativa en general. No toda herramienta de IA utilizada en educación constituye un chatbot educativo, ni toda interacción con IA implica una intervención conversacional con diseño pedagógico. Sistemas administrativos, analíticas de aprendizaje, motores de recomendación, plataformas adaptativas sin diálogo o usos incidentales de IA generativa pueden formar parte del ecosistema de IA educativa, pero no necesariamente corresponden al fenómeno analizado aquí. Esta distinción es especialmente importante en el contexto reciente de expansión de la IA generativa, porque los *AI chatbots* y los modelos conversacionales de gran escala han ampliado el campo, pero también han aumentado la necesidad de diferenciar entre uso tecnológico general e intervención pedagógica conversacional específica (Schei et al., 2024; Li & Chiu, 2025).

Más allá de su arquitectura técnica, el interés educativo de los chatbots depende de las funciones que desempeñan. Algunos sistemas operan como tutores o asistentes de tareas; otros facilitan práctica lingüística, simulación profesional, retroalimentación personalizada, apoyo metacognitivo o presencia social en entornos online. Desde una perspectiva pedagógica, los mecanismos más relevantes incluyen feedback inmediato, personalización, práctica repetida, interacción segura,

apoyo emocional, regulación metacognitiva, autonomía y competencia. Estudios recientes sobre emociones inducidas por chatbots y feedback empático sugieren que la interacción conversacional puede relacionarse con motivación, autorregulación y razonamiento metacognitivo, aunque estos efectos dependen del diseño de la interacción y no de la tecnología por sí sola (Yin et al., 2024; Ortega-Ochoa et al., 2024).

Esta delimitación permite situar la neuroeducación como marco interpretativo, no como promesa de medición neurobiológica directa. Los chatbots educativos pueden incidir en procesos relevantes para el aprendizaje —atención, emoción, recompensa, carga cognitiva, feedback, autorregulación, percepción de competencia y compromiso activo—, pero la mayor parte de la literatura disponible trabaja con indicadores psicológicos, educativos y conductuales. Por tanto, una lectura neuroeducativa rigurosa debe integrar estos procesos con prudencia, evitando inferir mecanismos cerebrales directos cuando los estudios no los miden explícitamente. La revisión bibliográfica previa muestra precisamente que los estudios suelen recurrir a teorías motivacionales, afectivas y de autorregulación que dialogan con la neuroeducación, aunque rara vez incorporan medidas neurocientíficas directas.

1.3. Fragmentación del campo: tecnologías, diseños, outcomes y teorías

1.3.1. Fragmentación tecnológica y terminológica

El campo de los chatbots educativos se caracteriza por una marcada fragmentación tecnológica y terminológica. La literatura alterna entre conceptos como *chatbot*, *conversational agent*, *educational bot*, *pedagogical agent*, *virtual tutor*, *dialogue system*, *AI tutor* y, más recientemente, *generative AI chatbot*. Esta pluralidad refleja la evolución técnica del área, pero también dificulta la recuperación, comparación y acumulación de evidencia. Bajo una misma etiqueta pueden agruparse sistemas basados en reglas, modelos de recuperación, agentes con rutas conversacionales

predeterminadas, tutores inteligentes, pacientes simulados, asistentes generativos o agentes multimodales. Kuhail et al. (2023) ya habían mostrado la necesidad de analizar los chatbots educativos según múltiples dimensiones —campo educativo, plataforma, principios de diseño, rol, estilo de interacción, evidencia y limitaciones—, mientras que Debets et al. (2025) identifican falta de claridad acumulada sobre objetivos, tecnologías, teorías, criterios de evaluación e impactos.

Esta dispersión no es solo terminológica, sino conceptual. Si se restringe demasiado el fenómeno, se corre el riesgo de excluir estudios relevantes que usan etiquetas alternativas. Si se amplía sin criterios, se corre el riesgo opuesto: agrupar tecnologías con funciones pedagógicas distintas y producir conclusiones excesivamente generales. Por ello, una revisión centrada en motivación y engagement requiere una delimitación suficientemente amplia para capturar el ecosistema conversacional, pero lo bastante precisa para diferenciar chatbots educativos de otras formas de IA educativa sin interacción pedagógica conversacional.

1.3.2. Fragmentación metodológica

La fragmentación también se observa en los diseños de investigación. Los estudios sobre chatbots educativos emplean diseños experimentales, cuasi experimentales, cualitativos, mixtos, transversales, estudios de caso, pilotos y encuestas de percepción. Esta diversidad permite observar el fenómeno desde múltiples aproximaciones, pero reduce la comparabilidad entre investigaciones y limita la posibilidad de establecer conclusiones causales robustas. Además, varias revisiones recientes advierten que buena parte de la evidencia se apoya en autoinformes, cuestionarios de percepción, satisfacción o utilidad, mientras que los estudios longitudinales y las evaluaciones de efectos sostenidos siguen siendo menos frecuentes. Huang et al. (2025) reportan que la investigación sobre chatbots y motivación se ha basado predominantemente en estudios

cuantitativos con cuestionarios, con escasa evidencia longitudinal, y Debets et al. (2025) señalan que muchos chatbots educativos son evaluados principalmente mediante factores perceptuales.

Esta situación tiene consecuencias para la interpretación del campo. Un aumento en satisfacción, interés o percepción de utilidad no equivale necesariamente a un cambio sostenido en motivación académica, engagement profundo o autorregulación. De igual modo, una interacción conversacional atractiva puede mejorar la experiencia inmediata del estudiante sin producir necesariamente efectos duraderos en aprendizaje, transferencia o persistencia. Por ello, la síntesis de la evidencia requiere distinguir entre resultados perceptuales, afectivos, conductuales, cognitivos y académicos.

1.3.3. Fragmentación de outcomes

La literatura también presenta dispersión en las variables evaluadas. Motivación y engagement rara vez aparecen como constructos aislados; suelen coexistir con satisfacción, interés, autoeficacia, utilidad percibida, facilidad de uso, ansiedad, presencia social, disposición a comunicarse, autorregulación, carga cognitiva, rendimiento académico o transferencia. Esta amplitud es teóricamente valiosa, pero dificulta determinar qué se sabe específicamente sobre motivación y engagement. Huang et al. (2025) identifican múltiples variables motivacionales en la investigación sobre chatbots, entre ellas engagement, interés y motivación intrínseca, mientras que la revisión bibliográfica previa muestra que otros estudios incorporan necesidades psicológicas, emociones, feedback, metacognición, apoyo docente y diseño conversacional como variables relacionadas.

Esta fragmentación obliga a evitar conclusiones agregadas demasiado simples. Un chatbot puede aumentar interés situacional sin mejorar motivación académica sostenida; puede favorecer participación conductual sin promover engagement cognitivo profundo; o puede reducir ansiedad en contextos comunicativos sin producir mejoras equivalentes en desempeño. Por ello, una revisión

centrada en motivación y engagement debe distinguir entre subdimensiones motivacionales, afectivas, cognitivas y conductuales, así como entre outcomes primarios y resultados secundarios.

1.3.4. Fragmentación teórica

La integración teórica del campo también permanece incompleta. Los estudios sobre chatbots educativos utilizan marcos como la teoría de la autodeterminación, la aceptación tecnológica, la presencia social, la autorregulación del aprendizaje, la carga cognitiva y el engagement multidimensional. Sin embargo, estos marcos no siempre se articulan entre sí ni se conectan con una explicación integrada de los procesos afectivos, cognitivos y motivacionales implicados en la interacción con chatbots. Debets et al. (2025) advierten que muchos chatbots se integran en escenarios educativos sin una base teórica sólida, aun cuando suelen mostrar efectos positivos en sus objetivos inmediatos.

Esta debilidad teórica es especialmente relevante para una lectura neuroeducativa. La literatura sobre chatbots y motivación trabaja con constructos como autonomía, competencia, emoción, feedback, presencia social, autorregulación, ansiedad y carga cognitiva, todos ellos vinculados con procesos de aprendizaje sensibles a la interacción entre emoción, cognición y conducta. Yin et al. (2024) muestran que las emociones inducidas durante la interacción con chatbots pueden afectar la motivación y que el feedback metacognitivo puede reducir emociones negativas. Ortega-Ochoa et al. (2024), por su parte, muestran que el feedback empático de un chatbot puede apoyar competencias, motivación, autorregulación y razonamiento metacognitivo en educación superior online. Sin embargo, la revisión bibliográfica previa indica que estas aproximaciones suelen basarse en mediciones psicológicas y educativas, no en evidencia neurobiológica directa.

1.4. Vacío de investigación: necesidad de una síntesis motivacional, de engagement y neuroeducativa

La literatura reciente ya ofrece contribuciones importantes para comprender los chatbots educativos. Kuhail et al. (2023) sistematizaron estudios sobre interacción con chatbots educativos y mostraron la diversidad de campos, plataformas, roles e interacciones. Deng y Yu (2023) aportaron una revisión sistemática y metaanálisis sobre el efecto de la tecnología chatbot en educación, con resultados favorables para el aprendizaje general. Debets et al. (2025) ampliaron el panorama al revisar objetivos, tecnologías subyacentes, teorías, criterios de evaluación e impactos. Huang et al. (2025) focalizaron específicamente la relación entre chatbots y motivación estudiantil, mostrando que el tema ha crecido, pero que sigue metodológicamente fragmentado y con necesidad de mayor desarrollo.

A pesar de estos avances, persiste un vacío específico. La literatura disponible continúa distribuida entre revisiones generales sobre chatbots educativos, estudios sobre IA generativa, investigaciones centradas en aprendizaje o percepción, y trabajos que abordan motivación o engagement como resultados parciales. Falta una síntesis sistemática que examine simultáneamente los chatbots educativos como intervenciones pedagógicas conversacionales, la motivación y el engagement como ejes de análisis, los tipos de sistemas utilizados, las variables evaluadas, la calidad metodológica de la evidencia y los mecanismos interpretables desde una perspectiva neuroeducativa.

Este vacío no implica ausencia de estudios, sino insuficiente integración crítica. La revisión bibliográfica previa muestra que el campo ha avanzado en identificar mecanismos como autonomía, competencia, feedback, emoción, presencia social, apoyo empático y autorregulación; sin embargo, estos mecanismos suelen aparecer dispersos entre marcos teóricos, diseños y niveles de evidencia distintos. También persisten tensiones importantes: los chatbots pueden ser percibidos como motivadores y útiles, pero sus efectos pueden depender de la calidad del diseño

conversacional, del tipo de estudiante, del contexto pedagógico y de la presencia o ausencia de interacción humana significativa.

En consecuencia, el problema central no es simplemente determinar si los chatbots educativos “funcionan”, sino comprender **qué tipo de evidencia existe, qué dimensiones de motivación y engagement han sido evaluadas, qué convergencias y divergencias muestra la literatura, qué grado de confianza permiten los diseños disponibles y qué mecanismos pueden interpretarse desde la neuroeducación sin sobredimensionar la evidencia disponible**. Una revisión sistemática con síntesis narrativa temática resulta pertinente para responder a esta necesidad, especialmente en un campo donde la heterogeneidad tecnológica, metodológica y conceptual impide reducir la evidencia a una conclusión única.

1.5. Objetivo y preguntas de investigación

El objetivo de este artículo es **analizar sistemáticamente la evidencia empírica reciente sobre el efecto de los chatbots educativos en la motivación y el engagement estudiantil, interpretando los hallazgos desde una perspectiva neuroeducativa**.

A partir de este objetivo, la revisión se orienta por las siguientes preguntas de investigación:

1. ¿Qué tipos de chatbots educativos y agentes conversacionales han sido evaluados en relación con la motivación y el engagement estudiantil?
2. ¿Qué variables motivacionales, afectivas, cognitivas y conductuales han sido consideradas en la literatura?
3. ¿Qué convergencias y divergencias muestra la evidencia empírica reciente sobre el uso de chatbots educativos?
4. ¿Qué nivel de confianza ofrecen los estudios disponibles según su calidad metodológica?

5. ¿Qué mecanismos pueden interpretarse desde una perspectiva neuroeducativa en la relación entre chatbots educativos, motivación y engagement?

Estas preguntas se formulan para responder a la fragmentación del campo, no para confirmar una hipótesis tecnológica predefinida. La revisión parte de la premisa de que los chatbots educativos no constituyen una intervención homogénea y de que sus efectos potenciales deben analizarse considerando diseño, función pedagógica, contexto de uso, outcome evaluado y solidez metodológica de los estudios disponibles.

1.6. Contribución esperada del artículo

Este artículo busca realizar tres contribuciones. En primer lugar, ofrece una contribución teórica al integrar literatura sobre chatbots educativos, motivación, engagement y neuroeducación. Esta integración permite superar una lectura puramente tecnológica del fenómeno y situar los sistemas conversacionales dentro de procesos de aprendizaje asociados con autonomía, competencia, feedback, emoción, presencia social, carga cognitiva y autorregulación. En este sentido, la revisión responde a la necesidad de articular marcos conceptuales que en la literatura aparecen con frecuencia de manera separada o parcialmente conectada (Huang et al., 2025; Debets et al., 2025). En segundo lugar, aporta una contribución metodológica al diferenciar tipos de chatbots, funciones pedagógicas, variables de resultado y niveles de confianza. Esta aproximación evita conclusiones globales simplificadas y permite distinguir entre evidencia robusta, evidencia preliminar y resultados todavía inciertos. Tal distinción es relevante en un campo donde los estudios varían ampliamente en diseño, muestra, instrumentos, duración, intensidad de intervención y forma de reporte.

En tercer lugar, la revisión ofrece una contribución práctica para el diseño de experiencias educativas mediadas por IA conversacional. Al organizar la evidencia en torno a mecanismos como

feedback inmediato, personalización, práctica segura, apoyo emocional, presencia social y autorregulación, el artículo puede orientar decisiones pedagógicas más precisas sobre cuándo, cómo y para qué integrar chatbots educativos. Esta contribución no supone promover su adopción indiscriminada, sino aportar una base crítica para diseñar intervenciones conversacionales que favorezcan motivación y engagement de manera pedagógicamente fundamentada.

2. Metodología

2.1. Diseño de la revisión, marco de reporte y pregunta de investigación

2.1.1. Tipo de revisión sistemática

El presente estudio se diseñó como una **revisión sistemática con síntesis narrativa temática**, orientada a analizar la evidencia empírica disponible sobre el efecto de los chatbots educativos en la motivación y el compromiso estudiantil (*engagement*) desde una perspectiva neuroeducativa. Esta modalidad de revisión fue considerada pertinente debido a la heterogeneidad esperada del campo, tanto en términos de diseños metodológicos como de niveles educativos, tipos de chatbot, variables de resultado, instrumentos de medición y contextos disciplinares.

El corpus previsto y posteriormente sistematizado incluyó estudios cuasi experimentales, cualitativos, mixtos, exploratorios, pilotos y observacionales, lo que hacía metodológicamente inadecuado asumir una estructura homogénea de evidencia. Por ello, la revisión no se planteó como metaanálisis, sino como una síntesis sistemática orientada a identificar patrones, convergencias, divergencias y vacíos en la literatura. La delimitación conceptual del fenómeno se centró en chatbots educativos, agentes conversacionales, *pedagogical agents*, sistemas de diálogo, tutores virtuales y sistemas afines utilizados con intención formativa.

2.1.2. Marco de reporte PRISMA 2020

La revisión se estructuró siguiendo los principios generales de transparencia y trazabilidad propuestos por PRISMA 2020 para revisiones sistemáticas. Este marco orientó la identificación de registros, el cribado por título y resumen, la evaluación a texto completo, la documentación de decisiones de inclusión/exclusión y la presentación del flujo de selección de estudios.

El uso de PRISMA se asumió como guía de reporte, no como una garantía automática de calidad metodológica. En consecuencia, la revisión documentó de manera explícita las bases consultadas, las ecuaciones de búsqueda, los filtros aplicados, los criterios de elegibilidad, el procedimiento de extracción, la evaluación de calidad metodológica y la estrategia de síntesis. El diagrama de flujo PRISMA se presenta posteriormente en la sección de Resultados, junto con la descripción del proceso de selección.

2.1.3. Pregunta de investigación y operacionalización SPIDER

La pregunta de investigación fue formulada mediante el modelo SPIDER, apropiado para revisiones que integran evidencia cuantitativa, cualitativa y mixta. La pregunta orientadora fue:

¿Cómo influyen los chatbots educativos en la motivación y el engagement de estudiantes en contextos educativos formales o no formales, y qué patrones pueden interpretarse desde una perspectiva neuroeducativa?

Tabla 1. Operacionalización de la pregunta de investigación mediante SPIDER

Elemento SPIDER	Definición en esta revisión
S — Sample	Estudiantes de cualquier nivel educativo
PI — Phenomenon of Interest	Uso de chatbots educativos, agentes conversacionales, <i>pedagogical agents</i> , tutores virtuales, AI tutors o sistemas conversacionales con finalidad formativa
D — Design	Estudios empíricos cuantitativos, cualitativos o mixtos
E — Evaluation	Motivación, engagement, satisfacción, autoeficacia, confianza, disposición a comunicarse, interés de aprendizaje, participación o constructos próximos

R — Research type	Evidencia cuantitativa, cualitativa y mixta
-------------------	---

Esta operacionalización se corresponde con la delimitación conceptual de la revisión, que define como resultados primarios la motivación y el engagement, incluyendo subdimensiones como motivación intrínseca, extrínseca, académica y engagement conductual, emocional o cognitivo.

2.1.4. Protocolo, registro y transparencia metodológica

La revisión no contó con un protocolo previamente registrado en una plataforma pública. Para mitigar esta limitación, se documentaron de forma explícita la estrategia de búsqueda, las bases consultadas, las ecuaciones booleanas, los filtros aplicados, los criterios de inclusión y exclusión, la matriz de extracción, la evaluación de calidad metodológica y la estrategia de síntesis.

La ausencia de registro prospectivo se reconoce como una limitación de transparencia metodológica. No obstante, la trazabilidad interna del proceso se preservó mediante tablas de elegibilidad, extracción, evaluación crítica y síntesis narrativa. Estas matrices permitieron vincular cada decisión metodológica con los criterios previamente definidos para delimitar el fenómeno tecnológico, el contexto educativo, los resultados de interés y el tipo de evidencia admisible.

2.2. Estrategia de búsqueda

2.2.1. Bases de datos consultadas y justificación

La búsqueda se realizó en bases de datos seleccionadas por su cobertura en educación, tecnología educativa, inteligencia artificial aplicada y sistemas conversacionales. En particular, se consultaron **ScienceDirect** e **IEEE Xplore**.

ScienceDirect fue incluida por su cobertura en educación, psicología del aprendizaje, inteligencia artificial educativa, ciencias sociales aplicadas y tecnología educativa. IEEE Xplore fue incorporada por su relevancia en informática educativa, inteligencia artificial, agentes conversacionales, sistemas de diálogo y tecnologías emergentes aplicadas al aprendizaje. Esta

combinación permitió cubrir tanto la literatura pedagógica y psicológica como la literatura tecnológica vinculada al diseño y evaluación de sistemas conversacionales.

2.2.2. Términos, sinónimos y variantes terminológicas

La estrategia de búsqueda no se limitó al término *chatbot*, dado que el campo utiliza denominaciones heterogéneas para referirse a sistemas conversacionales educativos. Por ello, se incorporaron variantes como *chat bot*, *chatbots*, *conversational agent*, *conversational AI*, *dialogue system*, *pedagogical agent*, *virtual tutor*, *intelligent tutor*, *AI tutor* y *virtual assistant*.

Los términos tecnológicos se combinaron con descriptores educativos como *education*, *learning*, *teaching*, *student*, *classroom*, *course*, *instruction* y *training*. A su vez, se incorporaron términos relacionados con motivación y engagement, entre ellos *motivation*, *intrinsic motivation*, *extrinsic motivation*, *academic motivation*, *engagement*, *student engagement*, *learner engagement*, *participation*, *persistence*, *self-determination*, *autonomy*, *competence* y *relatedness*.

Esta amplitud terminológica responde a la inestabilidad conceptual del campo. El protocolo de revisión estableció que restringir la búsqueda exclusivamente al término “chatbot” podía reducir la sensibilidad y excluir estudios relevantes que emplearan etiquetas como agente conversacional, tutor virtual o sistema de diálogo.

2.2.3. Ecuaciones de búsqueda por base de datos

Las ecuaciones de búsqueda fueron adaptadas a la sintaxis de cada base de datos. En ScienceDirect se utilizó una ecuación centrada en tres bloques: tecnología conversacional, contexto educativo y resultados motivacionales o de engagement.

En IEEE Xplore se empleó una ecuación ampliada en metadatos para capturar una mayor diversidad terminológica.

Tabla 2. Estrategia de búsqueda por base de datos

Base de datos	Ecuación de búsqueda	Campos utilizados	Justificación
ScienceDirect	("chatbot" OR "conversational agent" OR "pedagogical agent") AND (education OR learning) AND (motivation OR engagement OR "student engagement")	Búsqueda general/metadatos según disponibilidad de la base	Recuperar estudios educativos y psicológicos sobre chatbots, motivación y engagement
IEEE Xplore	Ecuación ampliada con chatbot, chatbots, conversational agent, pedagogical agent, dialogue system, virtual tutor, AI tutor, combinados con términos educativos y motivacionales	All Metadata	Recuperar literatura tecnológica y de IA educativa con terminología diversa

Las ecuaciones documentadas en el protocolo de búsqueda sirvieron como base para la identificación inicial de registros y para la posterior depuración temática del corpus.

2.2.4. Periodo de cobertura y filtros aplicados

La búsqueda ejecutada se concentró en el periodo **2020–2026**, con el propósito de capturar la expansión reciente de la inteligencia artificial conversacional aplicada a educación. Aunque el protocolo conceptual consideraba ventanas temporales más amplias para análisis de sensibilidad, la ejecución documentada de la búsqueda se focalizó en el periodo reciente, de acuerdo con los filtros aplicados en las bases consultadas.

Los filtros aplicados incluyeron año de publicación, tipo documental, idioma y acceso. En ScienceDirect se aplicaron filtros de artículos de investigación, idioma inglés/español y acceso abierto. En IEEE Xplore se aplicaron filtros por año, acceso abierto y búsqueda en metadatos. El uso de filtros buscó aumentar la pertinencia de los registros recuperados y reducir la inclusión de estudios no empíricos, no educativos o centrados en aplicaciones conversacionales ajenas al aprendizaje.

2.2.5. Alcance y límites de la estrategia de búsqueda

La estrategia de búsqueda priorizó la sensibilidad terminológica y la precisión temática. Sin embargo, presenta límites. El uso de filtros de acceso abierto pudo excluir estudios relevantes no disponibles bajo esa modalidad. La restricción lingüística pudo introducir sesgo idiomático. Asimismo, la dependencia de términos explícitos en metadatos pudo dejar fuera investigaciones que evaluaran sistemas conversacionales educativos con denominaciones distintas a las previstas en la ecuación de búsqueda.

Estas limitaciones fueron consideradas al interpretar el alcance de la revisión. Por ello, los hallazgos se entienden como representativos de la literatura empírica recuperada bajo la estrategia definida, no como una cobertura exhaustiva absoluta de todo el campo.

2.3. Gestión de registros y selección de estudios

2.3.1. Organización inicial de registros

Los registros recuperados fueron organizados en una matriz de cribado, en la que se documentaron los datos bibliográficos básicos, la base de procedencia, la decisión de elegibilidad y la justificación correspondiente. Esta matriz permitió conservar trazabilidad entre la búsqueda inicial, el cribado por título/resumen, la lectura a texto completo y la decisión final de inclusión, exclusión o clasificación dudosa.

No se atribuye el uso de software especializado de cribado cuando este no se encuentra documentado. La gestión se basó en matrices sistemáticas de control y en la aplicación explícita de criterios de elegibilidad.

2.3.2. Criterios de inclusión

Se incluyeron estudios que cumplieran los siguientes criterios:

1. Ser estudios empíricos primarios.
2. Incluir estudiantes como población principal.

3. Desarrollarse en un contexto educativo formal o no formal.
4. Evaluar el uso de un chatbot educativo, agente conversacional, sistema de diálogo, *pedagogical agent*, tutor virtual o AI tutor con intención formativa.
5. Presentar una intervención, uso, efecto o asociación relacionada con aprendizaje, enseñanza, tutoría, retroalimentación, práctica o apoyo académico.
6. Medir explícitamente motivación, engagement o constructos próximos.
7. Proporcionar información suficiente sobre muestra, intervención, método y resultados.
8. Estar dentro del periodo, idioma, tipo documental y filtros definidos en la búsqueda.

Estos criterios se derivan de la delimitación conceptual del protocolo, que establece como foco de la revisión los estudios empíricos primarios sobre chatbots educativos y su relación con motivación o engagement.

2.3.3. Criterios de exclusión

Se excluyeron revisiones sistemáticas, revisiones narrativas, ensayos teóricos, editoriales, artículos de opinión, estudios conceptuales sin datos empíricos y trabajos técnicos centrados exclusivamente en arquitectura, precisión algorítmica o desarrollo de software sin datos de usuarios.

También se excluyeron estudios sobre chatbots no educativos, aplicaciones conversacionales en salud, banca, marketing o atención al cliente, trabajos sobre inteligencia artificial generativa en educación sin intervención chatbot claramente delimitada, estudios centrados solo en usabilidad general sin vínculo con motivación o engagement, y estudios sin muestra o procedimiento empírico suficientemente descrito.

La tabla de elegibilidad documentó las decisiones de inclusión, exclusión y duda, acompañadas de una justificación específica para cada estudio evaluado.

2.3.4. Cribado por título y resumen

El primer nivel de selección se realizó mediante cribado por título y resumen. En esta fase se verificó si el registro hacía referencia a un chatbot, agente conversacional o sistema afín; si el contexto era educativo; si la población correspondía a estudiantes; si el documento tenía carácter empírico; y si se mencionaban variables de motivación, engagement o constructos próximos.

Los registros que no cumplían estos criterios fueron excluidos. Los casos con información insuficiente o ambigua en título y resumen fueron conservados para evaluación a texto completo.

2.3.5. Lectura completa y resolución de casos dudosos

Los textos potencialmente elegibles fueron evaluados a texto completo. En esta fase se aplicaron criterios más estrictos relacionados con la claridad de la intervención conversacional, la función pedagógica activa del sistema, el contexto educativo, la presencia de resultados motivacionales o de engagement, la suficiencia de la evidencia empírica y la trazabilidad de muestra y procedimiento.

Los casos dudosos fueron clasificados considerando si el estudio cumplía de manera sustantiva el núcleo de la pregunta de revisión. Por ejemplo, Ahmed et al. (2024) fue excluido porque abordaba la IA generativa en educación de forma amplia, sin centrarse en un chatbot educativo específico con evaluación directa de motivación o engagement como intervención principal. En cambio, He (2025) fue clasificado como dudoso debido a la ambigüedad metodológica de la intervención, la muestra reducida y la limitada claridad del procedimiento empírico.

2.3.6. Trazabilidad mediante diagrama PRISMA

El flujo de identificación, cribado, evaluación a texto completo e inclusión se documentó mediante un diagrama de flujo PRISMA. Este diagrama se presenta en la sección de Resultados para informar de manera visual el número de registros identificados, excluidos, evaluados e incluidos.

2.4. Extracción de datos

2.4.1. Matriz de extracción

La información de los estudios incluidos fue registrada en una matriz estructurada de extracción de datos. Esta matriz dio lugar a cuatro productos principales: la tabla de elegibilidad, la tabla de características de los estudios incluidos, la tabla de calidad metodológica/riesgo de sesgo y la tabla de síntesis narrativa de la evidencia.

La extracción se diseñó para permitir una caracterización integral de los estudios, la comparación de diseños, la identificación de dominios temáticos, la evaluación de tipos de intervención conversacional y la síntesis de resultados relacionados con motivación y engagement.

2.4.2. Variables bibliográficas, metodológicas y contextuales

Se extrajeron variables bibliográficas, metodológicas y contextuales: autor, año, país, revista, diseño del estudio, metodología, tamaño muestral, unidad de análisis, periodo de estudio y disponibilidad de datos o código. Estas variables permitieron describir la composición del corpus y valorar su diversidad geográfica, disciplinar y metodológica.

La tabla de características de los estudios incluidos contiene estos campos y permite observar la variedad de diseños, contextos educativos, muestras y formas de reporte presentes en el corpus.

2.4.3. Variables tecnológicas, educativas y de resultado

También se extrajeron variables relacionadas con la intervención tecnológica y el contexto educativo: tipo de IA o chatbot, plataforma o sistema utilizado, función pedagógica del chatbot, nivel educativo, área disciplinar, variables dependientes, instrumentos de recolección de datos y resultados principales.

En relación con los outcomes, se registraron variables como motivación académica, motivación intrínseca, motivación extrínseca, engagement conductual, engagement emocional, engagement cognitivo, engagement social, satisfacción, utilidad percibida, facilidad de uso, autoeficacia,

confianza, disposición a comunicarse, ansiedad comunicativa, interés de aprendizaje y rendimiento o desempeño cuando este aparecía asociado a la intervención. Estos campos se vinculan tanto con la caracterización de los estudios como con la síntesis narrativa de la evidencia.

2.4.4. Verificación de consistencia de la extracción

La extracción fue revisada mediante un procedimiento iterativo de consistencia interna. Para ello, se contrastaron las decisiones de elegibilidad, las características metodológicas, los outcomes reportados, los resultados principales y los juicios de calidad entre las matrices elaboradas.

Este procedimiento permitió detectar inconsistencias entre la clasificación de los estudios, la descripción de sus variables, la evaluación de riesgo de sesgo y la síntesis narrativa. No se afirma la existencia de revisión independiente por múltiples revisores cuando esta no se encuentra documentada.

2.5. Evaluación de la calidad metodológica y riesgo de sesgo

2.5.1. Herramientas utilizadas según diseño

La evaluación de calidad metodológica se adaptó al diseño de cada estudio. Dado que el corpus incluyó estudios no aleatorizados, cualitativos, mixtos, pilotos y cuasi experimentales, no se aplicó una única herramienta de evaluación.

Según la matriz de calidad metodológica, se utilizaron ROBINS-I para estudios no aleatorizados, CASP cualitativo para estudios cualitativos, MMAT para estudios de métodos mixtos y JBI cuasi experimental para diseños cuasi experimentales o antes/después específicos. Esta selección permitió valorar la calidad de cada estudio de acuerdo con su naturaleza metodológica.

2.5.2. Dominios evaluados

Los estudios fueron evaluados en dominios comunes de calidad o riesgo, ajustados a las herramientas correspondientes. Los dominios registrados fueron selección, medición, confusión, análisis, reporte y riesgo global.

Esta estructura permitió identificar limitaciones recurrentes en el corpus, como sesgo de selección, ausencia de aleatorización, control limitado de variables de confusión, dependencia de autoinformes, muestras reducidas o baja disponibilidad de datos/código. La matriz de calidad metodológica recoge estos juicios por estudio y proporciona una valoración global del riesgo.

2.5.3. Asignación de juicios de calidad

Los juicios por dominio se clasificaron como bajo, moderado o alto según la información disponible en cada estudio. El riesgo global se determinó considerando el patrón dominante de limitaciones metodológicas, con especial atención a la selección de participantes, el control de confusión, la validez de la medición, la adecuación del análisis y la transparencia del reporte.

En estudios cualitativos o mixtos, los juicios fueron interpretados en función de la coherencia del diseño, la adecuación de la muestra, la claridad del procedimiento analítico y la trazabilidad de la interpretación. La finalidad no fue homogeneizar artificialmente diseños distintos, sino valorar la solidez relativa de cada contribución.

2.5.4. Uso de la calidad metodológica en la síntesis

La evaluación de calidad metodológica no se utilizó como criterio automático de exclusión de los estudios incluidos. Su función fue modular el peso interpretativo de cada estudio dentro de la síntesis narrativa.

Los estudios con mayor riesgo de sesgo fueron tratados como evidencia preliminar o exploratoria, especialmente cuando presentaban muestras pequeñas, ausencia de grupo comparador, falta de aleatorización, dependencia de autoinforme o control limitado de variables de confusión. En

cambio, los estudios con diseños comparativos, instrumentos estructurados, triangulación metodológica o análisis más sólidos recibieron mayor peso interpretativo en la síntesis de patrones.

2.6. Estrategia de síntesis

2.6.1. Justificación de la síntesis narrativa temática

Se optó por una síntesis narrativa temática debido a la heterogeneidad del corpus. Los estudios diferían en diseño, muestra, nivel educativo, tipo de chatbot, duración e intensidad de la intervención, instrumentos de medición, variables dependientes y formato de reporte de resultados. Estas diferencias impedían combinar los resultados en un metaanálisis sin incurrir en comparaciones artificiales. Por ello, la síntesis se orientó a identificar patrones temáticos, convergencias, divergencias y niveles de confianza, más que a estimar un tamaño de efecto agregado. La tabla de síntesis narrativa organiza los estudios según diseño, tipo de evidencia, foco, resultado, contribución, dominio, utilidad práctica y confianza.

2.6.2. Procedimiento de agrupación temática y conceptual

El análisis se desarrolló mediante agrupación temática y conceptual de los estudios incluidos. Primero, los estudios fueron organizados por dominio educativo, incluyendo aprendizaje de lenguas, educación superior, formación clínica o profesional, programación, matemáticas, educación online y experiencias multimodales o de infancia temprana.

En segundo lugar, se agruparon según el tipo de intervención conversacional: chatbots orientados a tareas, agentes conversacionales pedagógicos, sistemas de práctica lingüística, pacientes simulados, agentes multimodales, sistemas basados en ejemplos trabajados y sistemas generativos o basados en ChatGPT.

En tercer lugar, se organizaron los outcomes en categorías motivacionales, afectivas, conductuales, cognitivas y perceptuales. Esta agrupación permitió analizar no solo qué resultados reportaba cada

estudio, sino también qué función pedagógica cumplía el chatbot y en qué condiciones parecía asociarse con motivación o engagement.

2.6.3. Clasificación de resultados y nivel de contribución

Los resultados fueron clasificados según su dirección e interpretación. Se consideraron **positivos** cuando el estudio reportaba mejora, efecto favorable o asociación positiva en motivación, engagement o constructos próximos. Se consideraron **mixtos** cuando coexistían beneficios con resultados nulos, limitados, ambiguos o posibles costos cognitivos. Se consideraron **preliminares** cuando los hallazgos eran favorables, pero el diseño, la muestra o la calidad del reporte limitaban la inferencia. Se consideraron **no concluyentes** cuando la evidencia era insuficiente para sostener una interpretación clara.

Además de la dirección del resultado, se registró la contribución de cada estudio, su utilidad práctica y el nivel de confianza atribuido en la síntesis narrativa.

2.6.4. Integración de calidad metodológica y confianza

La síntesis no se basó únicamente en la dirección de los resultados. Cada hallazgo fue interpretado considerando la calidad metodológica del estudio que lo sustentaba. Así, un resultado positivo no recibió el mismo peso si provenía de un estudio cuasi experimental con grupo control que si procedía de un piloto con muestra reducida, un estudio transversal, un estudio cualitativo exploratorio o un trabajo con limitada trazabilidad metodológica.

La integración entre la evaluación de calidad y la síntesis narrativa permitió diferenciar evidencia con confianza moderada, moderada-baja, baja o preliminar. Esta estrategia buscó evitar conclusiones causales excesivas y mantener una interpretación proporcional a la solidez de los diseños disponibles.

2.6.5. Integración de la perspectiva neuroeducativa

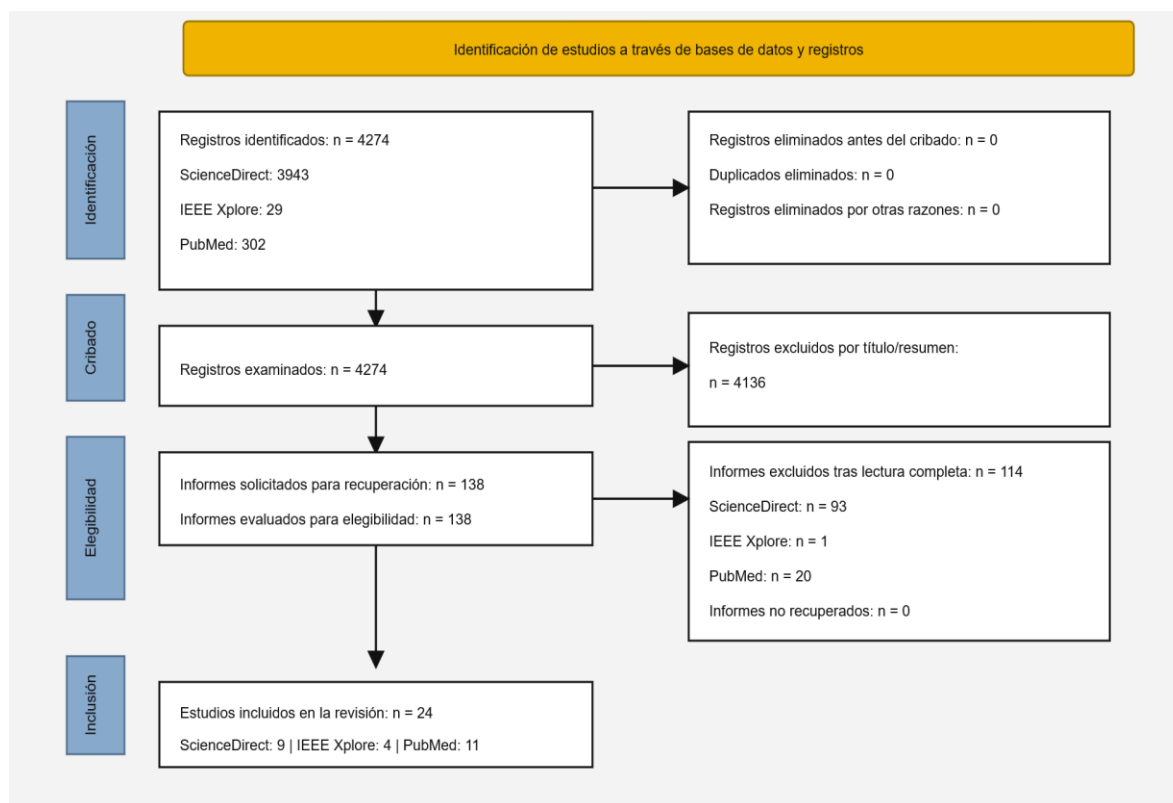
La perspectiva neuroeducativa no se utilizó como filtro primario de búsqueda, sino como eje interpretativo de la síntesis. El protocolo estableció que términos como neuroeducación, neurociencia, atención, recompensa, carga cognitiva, emoción, feedback, autorregulación y personalización debían orientar la extracción e interpretación, sin restringir la búsqueda principal. Los resultados fueron interpretados considerando procesos relevantes para el aprendizaje, como atención, motivación, recompensa, emoción, carga cognitiva, feedback, autorregulación, autonomía, competencia y engagement conductual, emocional, cognitivo y social. Esta interpretación se realizó a partir de indicadores psicológicos y educativos reportados por los estudios, no a partir de mediciones neurobiológicas directas.

3. Resultados

3.1. Selección de estudios y conformación del corpus

La búsqueda se diseñó para recuperar literatura empírica sobre chatbots educativos, agentes conversacionales, pedagogical agents y sistemas afines utilizados con intención formativa y asociados a variables de motivación, engagement o constructos próximos. El protocolo combinó términos tecnológicos —como *chatbot*, *conversational agent* y *pedagogical agent*— con términos educativos y resultados motivacionales, aplicando filtros temporales, lingüísticos, documentales y de acceso según la estrategia definida para la revisión. En ScienceDirect, por ejemplo, la búsqueda inicial arrojó un volumen amplio de registros, que fue reducido mediante filtros y cribado sucesivo por título/resumen y lectura completa. El protocolo documenta también búsquedas en IEEE Xplore con una ecuación ampliada para capturar variantes terminológicas del ecosistema conversacional educativo.

Figura 1. Diagrama de flujo PRISMA del proceso de selección de estudios.



El proceso de elegibilidad permitió distinguir entre estudios centrados en chatbots educativos como intervención pedagógica y trabajos más generales sobre inteligencia artificial generativa en educación. Esta delimitación fue decisiva para excluir estudios cuyo foco era amplio o perceptual, sin evaluación directa de un chatbot educativo sobre motivación o engagement. Por ejemplo, Ahmed et al. (2024) fue excluido por tratar la IA generativa de forma amplia, mientras que He (2025) fue clasificado como caso dudoso debido a la ambigüedad metodológica de la intervención y a la limitada trazabilidad empírica reportada.

En consecuencia, el corpus final quedó conformado por estudios empíricos primarios que evaluaban chatbots, agentes conversacionales o sistemas pedagógicos afines en contextos educativos, con resultados explícitos sobre motivación, engagement, satisfacción, disposición a comunicarse, confianza, interés de aprendizaje o variables relacionadas.

3.2. Caracterización general, temática y conceptual de los estudios incluidos

Los estudios incluidos configuran un campo internacional, interdisciplinar y metodológicamente heterogéneo. La evidencia procede de contextos diversos, entre ellos Egipto, Arabia Saudita, Emiratos Árabes Unidos/Malasia, Estados Unidos, Japón, Taiwán, Brasil, Australia, China y Hong Kong. Aunque esta distribución sugiere una expansión global del interés por los chatbots educativos, el corpus se concentra principalmente en educación superior. Los estudios en infancia temprana, primaria o secundaria son menos frecuentes, lo que indica que la evidencia disponible está más desarrollada en contextos universitarios y profesionales que en etapas escolares iniciales. Vease el anexo 1.

Desde el punto de vista metodológico, predominan diseños cuasi experimentales, estudios exploratorios, estudios de caso, métodos mixtos, pilotos y diseños transversales. Esta diversidad permite observar el fenómeno desde múltiples ángulos, pero también dificulta comparar directamente los resultados. Los tamaños muestrales son variables y, en varios casos, reducidos. Además, la disponibilidad de datos o código aparece mayoritariamente como no reportada, lo que limita la reproducibilidad y la verificación independiente de los hallazgos.

Temáticamente, el corpus se organiza en varios dominios. El primero es el aprendizaje de lenguas extranjeras, donde los chatbots son utilizados para practicar inglés, coreano u otras lenguas, con énfasis en motivación, disposición a comunicarse, ansiedad comunicativa, confianza y competencia percibida. En esta línea se ubican Annamalai et al. (2023), Ayedoun et al. (2020), Chien et al. (2022) y Kim y Su (2024). La relevancia de este dominio parece asociarse a la naturaleza conversacional de los chatbots, que permite práctica repetida, interacción flexible y reducción de barreras afectivas.

Un segundo dominio corresponde a la educación superior y el aprendizaje autorregulado. Aquí los chatbots se emplean para apoyar motivación académica, estrategias cognitivas y metacognitivas,

establecimiento de metas, presencia social y participación en entornos online. Al-Abdullatif et al. (2023), Hew et al. (2023) y Li (2025) representan esta línea. En estos estudios, el chatbot no funciona solo como herramienta de respuesta, sino como mediador de organización, acompañamiento y persistencia académica.

Un tercer dominio se relaciona con la formación clínica y profesional. Abdelwahab et al. (2025) y Hall et al. (2026) muestran usos de chatbots en enfermería, partería y farmacia clínica, donde la intervención se vincula con razonamiento, confianza, desempeño, comunicación profesional y autenticidad percibida. En este dominio, el valor pedagógico del chatbot reside en su capacidad para simular situaciones aplicadas, ofrecer práctica situada y apoyar el desarrollo de habilidades clínicas o comunicativas.

Un cuarto dominio agrupa estudios en STEM, programación y matemáticas. Garcia et al. (2025), Li (2025) y Ng et al. (2025) abordan variables como comprensión, interés de aprendizaje, autoeficacia, engagement, carga cognitiva y profundidad cognitiva. Esta línea se diferencia de las anteriores porque el chatbot se utiliza con mayor énfasis como apoyo cognitivo, fuente de ejemplos, guía de resolución o recurso para sostener la confianza ante tareas complejas.

Finalmente, aparecen líneas menos representadas pero conceptualmente relevantes. Aslan et al. (2024) amplía el campo hacia agentes pedagógicos multimodales e inmersivos en infancia temprana, incorporando interacción física, social y emocional. Liew et al. (2025), por su parte, examina cómo el estilo lingüístico del chatbot puede modular estados afectivos, engagement y carga cognitiva. Estos estudios muestran que el campo comienza a desplazarse desde la pregunta general sobre si los chatbots “funcionan” hacia preguntas más específicas sobre qué características de diseño producen determinados efectos.

En conjunto, las variables dominantes son motivación académica, motivación intrínseca y extrínseca, engagement conductual, emocional, cognitivo y social, satisfacción, utilidad percibida, facilidad de uso, autoeficacia, confianza, disposición a comunicarse, ansiedad comunicativa e interés de aprendizaje. El rendimiento o desempeño académico aparece en varios estudios, pero con menor centralidad que las variables motivacionales y perceptuales.

Conceptualmente, el campo se articula alrededor de la teoría de la autodeterminación, la autorregulación del aprendizaje, la presencia social, la aceptación tecnológica, la carga cognitiva y el engagement multidimensional. La neuroeducación opera como marco integrador para interpretar procesos de atención, emoción, recompensa, feedback, autonomía, competencia y regulación afectiva, aunque no todos los estudios incluidos adoptan explícitamente un diseño neuroeducativo.

3.3. Tipología de intervenciones conversacionales y funciones pedagógicas

Los estudios incluidos no evalúan una tecnología homogénea. Bajo la etiqueta general de chatbot educativo se agrupan sistemas basados en reglas, chatbots integrados en plataformas como WhatsApp o LINE, agentes conversacionales pedagógicos, pacientes virtuales simulados, sistemas basados en ejemplos trabajados, agentes multimodales y sistemas apoyados en ChatGPT o IA generativa. Esta heterogeneidad exige interpretar los resultados por función pedagógica y no solo por tipo de herramienta.

Una primera función es el andamiaje motivacional. En esta línea, el chatbot busca fortalecer motivación, autoeficacia, valor de tarea, confianza o disposición a participar. Abdelwahab et al. (2025), Al-Abdullatif et al. (2023), Ayedoun et al. (2020) y Kim y Su (2024) se ubican en esta orientación, aunque con diseños y contextos diferentes.

Una segunda función es la mediación del engagement. Hew et al. (2023), Aslan et al. (2024), Hall et al. (2026) y Liew et al. (2025) muestran que los chatbots pueden actuar como recursos para

sostener participación, interacción, presencia social, conexión emocional o implicación conductual. En estos casos, el chatbot no solo entrega información, sino que organiza una experiencia de interacción.

Una tercera función es la práctica segura y repetida. Esta función es especialmente visible en aprendizaje de lenguas y formación clínica. Annamalai et al. (2023), Chien et al. (2022), Kim y Su (2024) y Hall et al. (2026) muestran que los entornos conversacionales permiten repetir tareas, ensayar respuestas, practicar comunicación y reducir la exposición social que puede limitar la participación.

Una cuarta función es el apoyo cognitivo y autorregulatorio. Garcia et al. (2025), Hew et al. (2023), Li (2025) y He (2025) ilustran usos del chatbot como fuente de feedback, guía personalizada, apoyo para ejemplos trabajados, acompañamiento metacognitivo o soporte para el aprendizaje autónomo. Finalmente, algunos sistemas operan como simuladores situados. Hall et al. (2026), Abdelwahab et al. (2025) y Aslan et al. (2024) muestran intervenciones donde el valor del chatbot depende de su capacidad para recrear situaciones auténticas, favorecer interacción contextualizada y promover confianza en escenarios profesionales o multimodales.

En términos pedagógicos, los mecanismos más recurrentes son feedback inmediato, personalización, práctica repetida, presencia social, reducción de ansiedad, simulación auténtica, apoyo a la autonomía, apoyo a la competencia y andamiaje cognitivo. Estos mecanismos deben interpretarse como rutas plausibles de efecto, no como efectos universales demostrados en todos los estudios.

Tabla 3. Tipología funcional de las intervenciones conversacionales incluidas

Función pedagógica	Tipo de intervención conversacional	Estudios que la sustentan	Mecanismo pedagógico principal
--------------------	-------------------------------------	---------------------------	--------------------------------

Andamiaje motivacional	Chatbots orientados a motivación, confianza o disposición a participar	Abdelwahab et al., 2025; Al-Abdullatif et al., 2023; Ayedoun et al., 2020; Kim y Su, 2024	Apoyo a autonomía, competencia y valor de tarea
Mediación del engagement	Chatbots para presencia social, interacción o participación	Hew et al., 2023; Aslan et al., 2024; Hall et al., 2026; Liew et al., 2025	Incremento de participación, interacción y conexión social
Práctica segura y repetida	Chatbots de conversación o simulación	Annamalai et al., 2023; Chien et al., 2022; Kim y Su, 2024; Hall et al., 2026	Reducción de ansiedad y aumento de práctica autónoma
Apoyo cognitivo y autorregulatorio	Chatbots con feedback, ejemplos o guía personalizada	Garcia et al., 2025; Hew et al., 2023; Li, 2025; He, 2025	Feedback inmediato, regulación y andamiaje cognitivo
Simulación situada	Pacientes virtuales o agentes inmersivos	Hall et al., 2026; Abdelwahab et al., 2025; Aslan et al., 2024	Autenticidad, transferencia contextual y confianza profesional

3.4. Síntesis de la evidencia: convergencias, divergencias y patrones explicativos

El anexo 2 muestra la síntesis narrativa. Allí se aprecia una tendencia predominantemente favorable hacia el uso de chatbots educativos en variables motivacionales y de engagement. La mayoría de los estudios reporta resultados positivos o mixtos positivos, especialmente en motivación percibida, satisfacción, engagement conductual, confianza, autoeficacia, disposición a comunicarse, interés de aprendizaje y percepción de utilidad.

Las convergencias más claras se observan en contextos donde el chatbot ofrece interacción inmediata, práctica autónoma, feedback o acompañamiento personalizado. Abdelwahab et al. (2025) reporta mejoras en motivación académica, conocimiento, interpretación clínica, razonamiento y satisfacción en estudiantes de enfermería. Al-Abdullatif et al. (2023) muestra mayor motivación, valor de tarea, autoeficacia y estrategias cognitivas/metacognitivas en educación superior. En aprendizaje de lenguas, Annamalai et al. (2023), Ayedoun et al. (2020),

Chien et al. (2022) y Kim y Su (2024) sugieren que los chatbots pueden apoyar autonomía, competencia, disposición a comunicarse y confianza comunicativa.

También se observan hallazgos favorables en engagement. Aslan et al. (2024) reporta alto engagement conductual, emocional y social en una experiencia multimodal con agente pedagógico. Hew et al. (2023) muestra que chatbots diseñados para goal setting y presencia social pueden apoyar engagement en cursos online. Hall et al. (2026) encuentra que los pacientes simulados tipo chatbot fueron percibidos como auténticos, fáciles de usar y capaces de incrementar engagement y confianza en formación farmacéutica. Garcia et al. (2025) refuerza la utilidad de un agente pedagógico basado en ejemplos trabajados para apoyar comprensión, motivación, autoconfianza y engagement en programación.

Sin embargo, la evidencia no es uniforme. Los resultados son más consistentes para variables perceptuales, afectivas y conductuales que para rendimiento académico, aprendizaje profundo o transferencia. Chien et al. (2022), por ejemplo, muestra mejoras motivacionales y de práctica oral, pero sin ventajas significativas claras en rendimiento entre grupos. Liew et al. (2025) indica que ciertos estilos lingüísticos pueden aumentar engagement y emoción positiva, aunque con posible costo en carga cognitiva. Ng et al. (2025) plantea una tensión entre usabilidad, engagement y profundidad cognitiva: el chatbot puede ser atractivo y útil, pero no necesariamente mejorar rendimiento o pensamiento profundo.

Los hallazgos permiten distinguir tres niveles de inferencia. En primer lugar, algunos estudios cuasi experimentales o con comparación entre grupos sugieren efectos favorables, aunque la ausencia de aleatorización robusta en varios casos limita la fuerza causal de la interpretación. En segundo lugar, existe una asociación recurrente entre interacción con chatbots y aumento de motivación, satisfacción, engagement, confianza o utilidad percibida. En tercer lugar, pueden identificarse

patrones explicativos plausibles: el feedback inmediato puede fortalecer la sensación de progreso; la personalización puede apoyar autonomía y competencia; la práctica conversacional segura puede reducir ansiedad; la presencia social puede disminuir aislamiento en cursos online; la simulación auténtica puede favorecer confianza profesional; y los ejemplos guiados pueden apoyar comprensión y autoeficacia.

Desde una lectura neuroeducativa, estos patrones sugieren que los chatbots educativos pueden actuar sobre condiciones relevantes para el aprendizaje: atención sostenida, expectativa de recompensa, regulación emocional, percepción de competencia, autonomía y compromiso activo. No obstante, la evidencia disponible no mide directamente mecanismos neurobiológicos; más bien permite inferirlos a partir de variables psicológicas, conductuales y educativas. Por ello, la interpretación neuroeducativa debe mantenerse como marco explicativo prudente, no como demostración causal directa.

En conjunto, la síntesis indica que los chatbots educativos tienen mayor respaldo empírico como mediadores de motivación y engagement que como herramientas con efectos demostrados sobre aprendizaje profundo, rendimiento sostenido o transferencia a largo plazo.

3.5. Calidad metodológica, confianza de la evidencia y vacíos estructurales

La evaluación metodológica muestra que la evidencia disponible es prometedora, pero presenta limitaciones importantes. El riesgo global de sesgo se ubica con frecuencia entre moderado y alto, especialmente por problemas de selección, ausencia de aleatorización, confusión no controlada, muestras pequeñas, dependencia de autoinformes y limitada disponibilidad de datos o código. Véase el anexo 3.

Entre las fortalezas metodológicas del corpus se encuentran el uso de diseños cuasi experimentales en varios estudios, la presencia de grupos control en algunos casos, comparaciones pretest-postest,

instrumentos estructurados, métodos mixtos, registros conversacionales, entrevistas, observación y análisis estadísticos apropiados. Aslan et al. (2024), por ejemplo, destaca por triangulación multimodal, observación semiestructurada y codificación del engagement. Kim y Su (2024) combina comparación pre-post con apoyo cualitativo. Garcia et al. (2025) ofrece una descripción clara de muestra, instrumentos y análisis mixto.

No obstante, varios estudios presentan limitaciones que reducen la certeza de los resultados. Abdelwahab et al. (2025), Al-Abdullatif et al. (2023), Ayedoun et al. (2020) y Chien et al. (2022) muestran riesgos asociados a diseños no aleatorizados, selección y confusión. Hall et al. (2026) presenta una muestra efectiva muy pequeña y ausencia de grupo comparador. He (2025) muestra limitaciones severas por muestra reducida, falta de control y trazabilidad metodológica limitada. Li (2025), al ser transversal, permite identificar asociaciones, pero no establecer causalidad.

Por tipo de resultado, la confianza es mayor para motivación percibida, engagement conductual o perceptual, satisfacción y utilidad percibida. Estos outcomes aparecen de manera recurrente y con dirección predominantemente favorable. La confianza es moderada-baja para autoeficacia, confianza y disposición a comunicarse, porque aunque los resultados son positivos, dependen de contextos específicos y medidas heterogéneas. La confianza es baja para rendimiento académico, aprendizaje profundo, transferencia y efectos longitudinales, ya que estos resultados son menos frecuentes, menos consistentes o evaluados como outcomes secundarios.

Tabla 4. Nivel de confianza interpretativa por tipo de outcome

Tipo de outcome	Nivel de confianza sugerido	Fundamentación
Motivación percibida	Moderada	Resultado recurrente en varios estudios, aunque con diseños frecuentemente cuasi experimentales o basados en autoinforme

Engagement conductual o perceptual	Moderada	Evidencia convergente en cursos online, simulaciones, agentes multimodales y experiencias de aprendizaje asistidas
Satisfacción y utilidad percibida	Moderada	Alta recurrencia en estudios aplicados, pero dependiente de percepción estudiantil
Autoeficacia y confianza	Moderada-baja a moderada	Evidencia favorable, aunque con heterogeneidad de instrumentos y contextos
Disposición a comunicarse	Moderada	Evidencia relevante en aprendizaje de lenguas, con apoyo cuantitativo y cualitativo
Rendimiento académico	Baja a moderada-baja	Resultados menos consistentes y frecuentemente secundarios
Aprendizaje profundo y transferencia	Baja	Pocos estudios lo evalúan directamente
Efectos longitudinales	Baja	Escasa evidencia de seguimiento sostenido
Mecanismos neuroeducativos directos	Baja	Los mecanismos se infieren desde variables psicológicas y educativas, no desde mediciones neurobiológicas directas

Los vacíos estructurales son claros. El campo requiere más ensayos aleatorizados, estudios longitudinales, muestras amplias, comparación entre tipos de chatbot, análisis de moderadores, evaluación de efectos negativos, medición de carga cognitiva, estudios en educación primaria y secundaria, investigación sobre equidad y accesibilidad, y mayor transparencia de datos. También falta una operacionalización neuroeducativa más directa: la mayoría de los estudios permite interpretar mecanismos de atención, motivación, recompensa o autorregulación, pero no los mide de forma neurocognitiva directa.

3.6. Balance integrador de los resultados

Los resultados configuran un campo emergente, interdisciplinar y metodológicamente heterogéneo. La evidencia disponible muestra que los chatbots educativos se han estudiado principalmente como mediadores de motivación, engagement, práctica segura, feedback, presencia

social y autorregulación. Su uso parece especialmente prometedor en aprendizaje de lenguas, educación superior, formación clínica, programación, matemáticas y entornos online.

El conocimiento más consistente se concentra en variables perceptuales, afectivas y conductuales: motivación, satisfacción, utilidad percibida, confianza, disposición a comunicarse y engagement. En cambio, permanece incierto el impacto sobre rendimiento académico sostenido, aprendizaje profundo, transferencia, retención longitudinal y efectos diferenciales según edad, nivel educativo, diseño del chatbot o características del estudiante.

En síntesis, los chatbots educativos muestran potencial para activar condiciones motivacionales y de compromiso relevantes para el aprendizaje, pero la evidencia aún no permite sostener conclusiones causales fuertes o generalizables. La madurez del campo es, por tanto, intermedia: existe convergencia suficiente para reconocer una línea prometedora, pero se requieren diseños más robustos para establecer eficacia, mecanismos y condiciones de implementación.

4. Discusión

4.1. Interpretación central de los hallazgos

Los resultados de esta revisión sugieren que los chatbots educativos muestran un potencial más consistente para favorecer condiciones motivacionales y de compromiso que para producir mejoras académicas profundas o sostenidas. La evidencia converge en resultados favorables sobre motivación académica, satisfacción, engagement, confianza, autoeficacia, disposición a comunicarse e interés de aprendizaje, especialmente en contextos donde el chatbot ofrece práctica guiada, retroalimentación inmediata o interacción personalizada. Estudios como Abdelwahab et al. (2025), Al-Abdullatif et al. (2023), Hew et al. (2023), Kim y Su (2024) y Garcia et al. (2025) ilustran este patrón en dominios diversos, desde enfermería y educación superior online hasta lenguas y programación. Sin embargo, los efectos son menos concluyentes cuando se consideran

rendimiento académico, aprendizaje profundo, transferencia o efectos longitudinales. Esta diferencia indica que el aporte principal de los chatbots educativos parece ubicarse en la activación de disposiciones afectivas, conductuales y motivacionales para aprender, más que en la demostración robusta de efectos cognitivos duraderos.

4.2. Los chatbots como mediadores de autonomía, competencia y relación: lectura desde la Teoría de la Autodeterminación

La Teoría de la Autodeterminación ofrece una lente particularmente útil para interpretar estos hallazgos, porque permite comprender cómo ciertas características conversacionales pueden apoyar necesidades psicológicas básicas de autonomía, competencia y relación. La autonomía aparece vinculada con la posibilidad de practicar a ritmo propio, acceder a apoyo inmediato y repetir interacciones sin la presión evaluativa de una situación presencial. Este patrón es visible en estudios sobre aprendizaje de lenguas y educación online, donde los chatbots favorecen práctica flexible, disposición a comunicarse y participación autónoma. La competencia, por su parte, se relaciona con feedback, ejemplos guiados, resolución progresiva de tareas y percepción de mejora. En este sentido, estudios como Al-Abdullatif et al. (2023) y Garcia et al. (2025) sugieren que los chatbots pueden fortalecer motivación, estrategias cognitivas, autoconfianza y comprensión cuando el diseño ofrece andamiaje pedagógico claro.

La necesidad de relación también ayuda a explicar los efectos sobre engagement. Aunque un chatbot no sustituye la interacción humana, puede generar presencia conversacional, acompañamiento percibido y un espacio de práctica socialmente seguro. Hew et al. (2023) muestra el valor de los chatbots para goal setting y presencia social en cursos online; Kim y Su (2024) vincula la interacción con menor ansiedad y mayor confianza comunicativa; y Hall et al. (2026) muestra que los pacientes simulados pueden aumentar autenticidad percibida y engagement en

formación clínica. No obstante, esta interpretación debe formularse con cautela: los chatbots no satisfacen automáticamente necesidades psicológicas básicas. Solo pueden hacerlo cuando el diseño conversacional ofrece feedback significativo, interacción pedagógicamente orientada, apoyo emocional razonable y carga cognitiva manejable.

4.3. Alcance de la evidencia: convergencias, tensiones y límites metodológicos

Los resultados coinciden con literatura previa que describe a los chatbots educativos como recursos capaces de apoyar motivación, interés, feedback, interacción y engagement. Esta convergencia es consistente con revisiones recientes que señalan asociaciones positivas entre chatbots, motivación y experiencia estudiantil, pero también con trabajos que advierten que tales efectos dependen del diseño, la calidad de la interacción y el contexto pedagógico. La principal tensión es que un mayor engagement no implica necesariamente aprendizaje profundo. Liew et al. (2025), por ejemplo, muestra que ciertos rasgos lingüísticos pueden aumentar engagement y emoción positiva, pero también elevar la carga cognitiva. De forma similar, los resultados mixtos reportados en algunos estudios sugieren que la participación conductual o la satisfacción no deben interpretarse automáticamente como evidencia de transferencia, pensamiento profundo o rendimiento sostenido. Esta cautela se refuerza al considerar la calidad metodológica del corpus. La evidencia presenta riesgos frecuentes de selección y confusión, diseños no aleatorizados, muestras pequeñas, dependencia de autoinformes, ausencia de seguimiento longitudinal y baja disponibilidad de datos o código. Aunque varios estudios utilizan instrumentos estructurados, grupos de comparación, métodos mixtos o análisis estadísticos adecuados, el riesgo global se ubica con frecuencia entre moderado y alto. Por tanto, los hallazgos permiten identificar tendencias prometedoras, pero no sostener afirmaciones causales fuertes ni generalizaciones amplias sobre la eficacia de los chatbots educativos. La confianza resulta mayor para motivación percibida, engagement conductual,

satisfacción y utilidad percibida, y menor para rendimiento, aprendizaje profundo, transferencia y efectos sostenidos.

4.4. Implicaciones teóricas, prácticas y agenda futura

La principal implicación teórica es que los chatbots educativos no deben entenderse como herramientas motivadoras por sí mismas, sino como posibles **mediadores conversacionales** de autonomía, competencia, relación, emoción y autorregulación. Desde una perspectiva neuroeducativa prudente, su valor reside en activar condiciones psicológicas y pedagógicas compatibles con el aprendizaje —feedback, atención, emoción, percepción de progreso, seguridad afectiva y regulación—, no en demostrar mecanismos neurobiológicos directos. En términos prácticos, el diseño de chatbots debería priorizar feedback claro, personalización, práctica segura, presencia social, andamiaje cognitivo y control de carga cognitiva, evitando reducir la calidad de la experiencia a métricas de uso o satisfacción. La agenda futura requiere ensayos más robustos, estudios longitudinales, comparación entre tipos de chatbot, medición de engagement cognitivo, análisis de efectos adversos y una operacionalización más precisa de los mecanismos neuroeducativos. En síntesis, el potencial de los chatbots educativos depende menos de automatizar la interacción y más de diseñar conversaciones pedagógicamente significativas que sostengan motivación, engagement y autorregulación.

Conclusiones

La evidencia revisada permite concluir que los chatbots educativos constituyen una línea prometedora dentro de la inteligencia artificial aplicada a la educación, especialmente por su potencial para favorecer condiciones motivacionales y de compromiso estudiantil. Los hallazgos muestran efectos más consistentes en motivación percibida, satisfacción, engagement conductual,

confianza, autoeficacia, disposición a comunicarse e interés de aprendizaje que en resultados académicos complejos o sostenidos.

Desde una perspectiva teórica, los resultados se comprenden mejor cuando los chatbots no se interpretan como herramientas motivadoras por sí mismas, sino como **mediadores conversacionales** capaces de apoyar autonomía, competencia y relación. Esta lectura, basada principalmente en la Teoría de la Autodeterminación, permite explicar por qué los efectos son más visibles cuando el chatbot ofrece feedback significativo, práctica flexible, acompañamiento conversacional, personalización y oportunidades de interacción segura.

La perspectiva neuroeducativa aporta un marco útil para interpretar los hallazgos, siempre que se emplee con cautela. Los estudios revisados permiten inferir relaciones con procesos como atención, emoción, recompensa, carga cognitiva, autorregulación y percepción de competencia; sin embargo, la mayoría de la evidencia procede de indicadores psicológicos, educativos o conductuales, no de mediciones neurobiológicas directas. Por tanto, las conclusiones neuroeducativas deben formularse como interpretaciones teóricas plausibles, no como demostraciones causales cerebrales.

La revisión también muestra límites importantes. La confianza global en la evidencia es moderada o moderada-baja debido a la heterogeneidad de diseños, instrumentos, contextos y tipos de chatbot. Además, varios estudios presentan muestras pequeñas, ausencia de aleatorización, dependencia de autoinformes, falta de seguimiento longitudinal y escasa disponibilidad de datos o código. Estas limitaciones impiden sostener afirmaciones causales fuertes o generalizaciones amplias sobre la eficacia de los chatbots educativos.

En términos prácticos, los resultados sugieren que la integración de chatbots en educación debe orientarse por criterios pedagógicos claros. No basta con incorporar IA conversacional; es necesario diseñar interacciones que ofrezcan retroalimentación útil, reduzcan barreras afectivas,

promuevan práctica significativa, apoyen la autorregulación y controlen la carga cognitiva. El valor educativo del chatbot depende menos de su novedad tecnológica que de la calidad de su diseño instruccional y conversacional.

Finalmente, futuras investigaciones deberían priorizar ensayos más robustos, estudios longitudinales, comparación entre tipos de chatbot, análisis de efectos adversos, evaluación de engagement cognitivo profundo y medición más precisa de mecanismos neuroeducativos. También se requiere mayor atención a diferencias individuales, niveles educativos poco estudiados, interacción docente-chatbot y transparencia metodológica. En conjunto, el campo muestra un potencial considerable, pero todavía necesita evidencia más sólida para establecer bajo qué condiciones los chatbots educativos contribuyen de manera sostenida a la motivación, el engagement y el aprendizaje.

Referencias bibliográficas

- Abdelwahab, A. M. T., Aboraiah, M. I. H., & Elsayed, H. E. M. (2025). Effect of using artificial intelligence chatbot about electronic fetal monitoring on maternity nursing students' performance. *BMC medical education*. <https://doi.org/10.1186/s12909-025-08391-1>
- Ahmed, Z., Shanto, S. S., Rime, Most. H. K., Morol, Md. K., Fahad, N., Hossen, Md. J., & Abdullah-Al-Jubair, Md. (2024). The Generative AI Landscape in Education: Mapping the Terrain of Opportunities, Challenges, and Student Perception. *IEEE Access*, 12, 147023-147050. <https://doi.org/10.1109/ACCESS.2024.3461874>
- Al-Abdullatif, A. M., Al-Dokhny, A. A., & Drwish, A. M. (2023). Implementing the Bashayer chatbot in Saudi higher education: Measuring the influence on students' motivation and learning strategies. *Frontiers in psychology*. <https://doi.org/10.3389/fpsyg.2023.1129070>
- Annamalai, N., Eltahir, M. E., Zyoud, S. H., Soundrarajan, D., Zakarneh, B., & Al Salhi, N. R. (2023). Exploring English language learning via Chabot: A case study from a self determination theory perspective. *Computers and Education: Artificial Intelligence*, 5, 100148. <https://doi.org/10.1016/j.caeai.2023.100148>
- Aslan, S., Durham, L. M., Alyuz, N., Okur, E., Sharma, S., Savur, C., & Nachman, L. (2024). Immersive multi-modal pedagogical conversational artificial intelligence for early childhood education: An exploratory case study in the wild. *Computers and Education: Artificial Intelligence*, 6, 100220. <https://doi.org/10.1016/j.caeai.2024.100220>
- Ayedoun, E., Hayashi, Y., & Seta, K. (2020). Toward Personalized Scaffolding and Fading of Motivational Support in L2 Learner–Dialogue Agent Interactions: An Exploratory Study.

- IEEE Transactions on Learning Technologies*, 13(3), 604-616.
<https://doi.org/10.1109/TLT.2020.2989776>
- Chien, Y.-C., Wu, T.-T., Lai, C.-H., & Huang, Y.-M. (2022). Investigation of the Influence of Artificial Intelligence Markup Language-Based LINE ChatBot in Contextual English Learning. *Frontiers in psychology*. <https://doi.org/10.3389/fpsyg.2022.785752>
- Critical Appraisal Skills Programme. (2018). CASP qualitative checklist. <https://casp-uk.net/casp-tools-checklists/>
- Debets, T., Banihashem, S. K., Joosten-ten Brinke, D., Vos, T. E. J., Maillette de Buy Wenniger, G., & Camp, G. (2025). Chatbots in education: A systematic review of objectives, underlying technology and theory, evaluation criteria, and impacts. *Computers & Education*, 234, Article 105323. <https://doi.org/10.1016/j.compedu.2025.105323>
- Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability*, 15(4), Article 2940. <https://doi.org/10.3390/su15042940>
- Garcia, R. D. S., Villa, J. E. A., Miranda, A. L. M., Guedes, G. T. A., Oran, A. C., Souza, P. S. S. D., Vilela, R. F., Valle, P. H. D., & Silva, W. (2025). Theory Inspires, but Examples Engage: A Mixed-Methods Analysis of Worked Examples From CoderBot in Programming Education. *IEEE Access*, 13, 145007-145022. <https://doi.org/10.1109/ACCESS.2025.3599357>
- Hall, S., Grant, G. D., Hussey, S., Harwood, B., Austin, L., & Gospos, E. (2026). Evaluation of the use of chatbots as simulated patients in capstone clinical pharmacy education: A pilot study. *Currents in Pharmacy Teaching and Learning*, 18(5), 102588. <https://doi.org/10.1016/j.cptl.2026.102588>

- He, J. (2025). Standardizing AI-Assisted English Writing: ChatGPT's Opportunities, Challenges, and Transformer-Based Innovations for Scholarly Communication. *Journal of ICT Standardization*, 13(4), 383-404. <https://doi.org/10.13052/jicts2245-800X.1342>
- Hew, K. F., Huang, W., Du, J., & Jia, C. (2023). Using chatbots to support student goal setting and social presence in fully online activities: Learner engagement and perceptions. *Journal of Computing in Higher Education*, 35(1), 40-68. <https://doi.org/10.1007/s12528-022-09338-x>
- 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. *Education for Information*, 34(4), 285–291. <https://doi.org/10.3233/EFI-180221>
- Huang, W., Jiang, J., King, R. B., & Fryer, L. K. (2025). Chatbots and student motivation: A scoping review. *International Journal of Educational Technology in Higher Education*, 22, Article 26. <https://doi.org/10.1186/s41239-025-00524-2>
- Joanna Briggs Institute. (2020). Checklist for quasi-experimental studies. Joanna Briggs Institute. <https://jbi.global/critical-appraisal-tools>
- Kim, A., & Su, Y. (2024). How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. *System*, 122, 103256. <https://doi.org/10.1016/j.system.2024.103256>
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973–1018. <https://doi.org/10.1007/s10639-022-11177-3>

- Li, K. (2025). Factors influencing learners learning interest when using AI chatbots-assisted math leaning in higher education. *Frontiers in psychology*. <https://doi.org/10.3389/fpsyg.2025.1716913>
- Li, Y., & Chiu, T. K. F. (2025). The mediating effects of needs satisfaction on the relationship between teacher support and student engagement with generative artificial intelligence (GenAI) chatbots from a self-determination theory (SDT) perspective. *Education and Information Technologies*, 30, 20051–20070. <https://doi.org/10.1007/s10639-025-13574-w>
- Liew, T. W., Tan, S.-M., Pang, W. M., Gan, C. L., Chan, T. J., & Ahmad, F. (2025). Cringe, lit, or mid: Affective and cognitive effects of youth slang in an educational chatbot. *Acta Psychologica*, 256, 105036. <https://doi.org/10.1016/j.actpsy.2025.105036>
- Liu, S. (2026). Beyond the bot: Instructor facilitation, self-efficacy, and online English motivation in an AI-rich era. *Acta Psychologica*, 264, 106534. <https://doi.org/10.1016/j.actpsy.2026.106534>
- Ng, I. S. H., Siu, A., Han, C. S. J., Ho, O. S. H., Sun, J., Markiv, A., Knight, S., & Sagoo, M. G. (2025). Evaluating a Custom Chatbot in Undergraduate Medical Education: Randomised Crossover Mixed-Methods Evaluation of Performance, Utility, and Perceptions. *Behavioral sciences (Basel, Switzerland)*. <https://doi.org/10.3390/bs15091284>
- Or, A., Sukumar, S., Ma, A., Ang, D., Liu, M., Ritchie, H. E., & Sarrafpour, B. (2026). Enhancing Dental Students' History-Taking Skills With a Generative Artificial Intelligence Chatbot. *Journal of Dental Education*, 90(3), 384-390. <https://doi.org/10.1002/jdd.13952>
- Ortega-Ochoa, E., Quiroga Pérez, J., Arguedas, M., Daradoumis, T., & Marquès Puig, J. M. (2024). The effectiveness of empathic chatbot feedback for developing computer competencies,

- motivation, self-regulation, and metacognitive reasoning in online higher education. *Internet of Things*, 25, Article 101101. <https://doi.org/10.1016/j.iot.2024.101101>
- Ortega-Ochoa, E., Quiroga Pérez, J., Arguedas, M., Daradoumis, T., & Marquès Puig, J. M. (2024). The effectiveness of empathic chatbot feedback for developing computer competencies, motivation, self-regulation, and metacognitive reasoning in online higher education. *Internet of Things*, 25, 101101. <https://doi.org/10.1016/j.iot.2024.101101>
- 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., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71. <https://doi.org/10.1136/bmj.n71>
- Quah, B., Chia, N. W. C., Huynh, M. K. H., & Islam, I. (2026). Implementation of an artificial intelligence-powered training platform in oral medicine and pathology education: A comparative study. *BMC Medical Education*, 26, 462. <https://doi.org/10.1186/s12909-026-08815-6>
- Rahimi, A. R., Sheykhholeslami, M., & Mahmoudi Pour, A. (2025). Uncovering personalized L2 motivation and self-regulation in ChatGPT-assisted language learning: A hybrid PLS-SEM-ANN approach. *Computers in Human Behavior Reports*, 17, 100539. <https://doi.org/10.1016/j.chbr.2024.100539>
- Rücker, C. R., & Becker-Genschow, S. (2025). Enhancing enthusiasm for STEM education with AI: Domain-specific chatbot as personalized learning assistant. *Computers and Education Open*, 9, 100315. <https://doi.org/10.1016/j.caeo.2025.100315>

- Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and use of AI chatbots among students in higher education: A scoping review of empirical studies. *Education Sciences*, 14(8), Article 922. <https://doi.org/10.3390/educsci14080922>
- Septiyanti, N. D., Luthfi, M. I., & Darmawansah, D. (2024). Effect of chatbot-assisted learning on students' learning motivation and its pedagogical approaches. *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, 10(1), 69–77. <https://doi.org/10.23917/khif.v10i1.4246>
- Shen, Q. (2025). Artificial intelligence as a pedagogical communicator: Mixed-methods insights from Raffles International College Bangkok. *Computers and Education Open*, 9, 100297. <https://doi.org/10.1016/j.caeo.2025.100297>
- Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., Ramsay, C. R., Regidor, D., Rothstein, H. R., Sandhu, L., Santaguida, P. L., Schünemann, H. J., Shea, B., Shrier, I., Tugwell, P., Turner, L., Valentine, J. C., Waddington, H., Waters, E., Wells, G. A., Whiting, P. F., & Higgins, J. P. T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355, Article i4919. <https://doi.org/10.1136/bmj.i4919>
- Wei, W., Zhao, A., & Ma, H. (2025). Understanding How AI Chatbots Influence EFL Learners' Oral English Learning Motivation and Outcomes: Evidence From Chinese Learners. *IEEE Access*, 13, 56699-56716. <https://doi.org/10.1109/ACCESS.2025.3554545>
- Yin, J., Goh, T. T., & Hu, Y. (2024). Interactions with educational chatbots: The impact of induced emotions and students' learning motivation. *International Journal of Educational*

Technology in Higher Education, 21, Article 39. <https://doi.org/10.1186/s41239-024-00480-3>

Yin, J., Goh, T. T., Yang, B., & Yang, X. (2021). Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance. *Journal of Educational Computing Research*, 59(1), 154–177. <https://doi.org/10.1177/0735633120952067>

Zhai, C., Wibowo, S., & Li, L. D. (2024). Evaluating the AI dialogue System's intercultural, humorous, and empathetic dimensions in English language learning: A case study. *Computers and Education: Artificial Intelligence*, 7, 100262. <https://doi.org/10.1016/j.caeai.2024.100262>

Zhang, R., Zou, D., Xie, H., & Wang, F. L. (2025). Engagements with GPT responses and learner prompts in ChatGPT-based learning of English argumentative writing logic and their impacts. *Computers and Education: Artificial Intelligence*, 9, 100489. <https://doi.org/10.1016/j.caeai.2025.100489>

Anexos

Anexo 1

Tabla 1. Características de los estudios incluidos

	Author/Year	Country	Journal	Study Design	Methodology	Sample Size	Unit of Analysis	Type of AI	Educational Context	Variables	Data Collection Method	Key Findings	Study Period	Data/Code Availability
1	Abdelwahab et al., 2025	Egypt	BMC Medical Education	Cuasi experimental, pretest - posttest con grupo control	Cuasi experimental	84	Estudiantes universitarios de enfermería	Chatbot educativo de IA sobre monitoreo fetal electrónico	Educación superior formal; estudiantes de tercer nivel en curso de partería, Mansoura University	Conocimiento teórico, interpretación práctica, confianza en razonamiento clínico, motivación académica, satisfacción	Cuestionario autoadministrado, Academic Motivation Scale, encuesta de satisfacción, análisis con SPSS, pruebas t, chi-cuadrado, Fisher, tamaños del efecto	Mejoras significativas en conocimiento, interpretación, razonamiento clínico y motivación académica en posttest y seguimiento; 90.5% reportó alta satisfacción	Año académico 2024/2025	NR
2	Al-Abdullatif et al., 2023	Arabia Saudita	Frontiers in Psychology	Cuasi experimental con enfoque experimental de sujeto único	Cuasi experimental	60	Estudiantes de posgrado	Chatbot orientado a tareas, integrado en WhatsApp, modelo de recuperación	Educación superior; posgrado en Arabia Saudita	Motivación de aprendizaje, estrategias cognitivas y metacognitivas	Escala MSLQ, comparación entre grupo experimental y control, análisis descriptivo y t de muestras independientes	El grupo experimental mostró mayor motivación y mayor uso de estrategias cognitivas y metacognitivas que el grupo control	NR	NR
3	Annamalai et	Emiratos Árabes Unidos	Computers and Education	Estudio de caso	Cualitativo	25	Estudiantes universitarios de	Chatbot educativo para aprendizaje de inglés	Educación superior formal;	Motivación desde TAD: autonomía, competencia y	Entrevistas; análisis temático de Braun y Clarke	El chatbot apoyó competencia, autonomía y relación; también se reportaron falta de entorno emocional e información inexacta;	NR	NR

	al., 2023	os / Malasia	ation : Artificial Intelligence		iv a		pregrado		aprendizaje del inglés	relación; barreras percibidas		se sugirió uso combinado con enseñanza presencial o blended		
4	Aslan et al., 2024	Estados Unidos	Computers and Education : Artificial Intelligence	Cuasi experimental y estudio de caso exploratorio en entorno real	Métodos mixtos	17	14 estudiantes y 3 educadores	Agente conversacional pedagógico multimodal inmersivo con colaboración humano-IA	Educación infantil / primaria temprana; escuela elemental, aprendizaje matemático	Engagement, experiencia, rendimiento, percepciones de estudiantes y educadores	Video, audio, lidar, registros del sistema, cuestionarios, entrevistas, observación semiestructurada, codificación multimodal del engagement, posttest	Percepciones positivas; alto engagement con menor tiempo de pantalla, mayor actividad física y más interacción social; ganancia de aprendizaje del 24%	NR	Datos no identificables disponibles previa solicitud razonable
5	Ayedoun et al., 2020	Japón	IEEE Transactions on Learning Technologies	Estudio exploratorio empírico	Cuantitativa	NR	Estudiantes de L2	Agente conversacional con estrategias de comunicación y afective backchannels	Aprendizaje de inglés como lengua extranjera	Willingness to communicate, preferencias por estrategias conversacionales, expected willingness to communicate	Interacción con sistema y posttest sobre preferencias y expected WTC	La mayoría prefirió combinar estrategias; quienes tenían menor WTC tendieron a preferir affective backchannels y quienes tenían mayor WTC, communication strategies; el expected WTC tendió a ser mayor cuando el sistema incorporaba la estrategia preferida del estudiante.	NR	NR
6	Chen et al., 2022	Taiwán	Frontiers in Psychology	Cuasi experimental pretest - posttest con grupo control no aleatorizado	Cuantitativa	73	Estudiantes de secundaria	LINE ChatBot basado en AIML con estrategia competitiva en el grupo experimental	Aprendizaje contextual de inglés	Habilidad de speaking, habilidad de listening, motivación intrínseca, motivación extrínseca	Pretest y posttest; speaking evaluado por docente; listening; MSLQ; ANCOVA	El rendimiento de ambos grupos mejoró ligeramente sin diferencias significativas entre grupos en speaking y listening; la motivación extrínseca mejoró en ambos grupos; el componente competitivo incrementó la motivación intrínseca.	4 semanas	NR

7	García et al., 2025	Brasil	IEEE Access	Estudio exploratorio de métodos mixtos	Mixta	103	Estudiantes universitarios de cursos introductorios de programación	CoderBot, agente pedagógico basado en Example-Based Learning	Educación superior en programación	Aceptación tecnológica, autoeficacia, motivación, autoconfianza, percepciones de utilidad	Cuestionario TAM, cuestionario de autoeficacia, preguntas abiertas, estadística descriptiva y codificación cualitativa	Se reportó facilidad de uso, mejoras perceptibles en comprensión de conceptos, motivación y autoconfianza; cualitativamente se observó potencial pedagógico en claridad, accesibilidad y apoyo continuo, aunque con necesidad de ampliar ejemplos y mejorar respuestas.	NR	NR
8	Hall et al., 2026	Australia	Currants in Pharmacy Teaching and Learning	Estudio piloto de métodos mixtos	Mixta	9 participantes antes en la evaluación; 60 estudiantes matriculados en el curso	Estudiantes de cuarto año de farmacia	Microsoft Power Virtual Agents como pacientes virtuales simulados	Educación superior; curso capstone de farmacoterapia clínica	Usabilidad, engagement, confianza, percepción de autenticidad, mejora de habilidades clínicas	Cuestionario Likert adaptado del SUS, respuestas cualitativas, estadística descriptiva y análisis temático inductivo	Los chatbots fueron bien recibidos; aumentaron el engagement con el curso, ofrecieron una experiencia auténtica, fueron percibidos como fáciles de usar y apoyaron habilidades clínicas y de comunicación.	12 semanas	NR
9	He, 2025	China	Journal of ICT Standardization	Estudio de caso con análisis empírico pretest - posttest	Mixta	5	Estudiantes	ChatGPT como sistema de apoyo personalizado a la escritura en inglés	Enseñanza/aprendizaje de escritura en inglés	Calidad de escritura, velocidad de escritura, participación, satisfacción, motivación de aprendizaje	Análisis de errores y calidad antes/después, tiempo de escritura, seguimiento de uso, escala Likert de satisfacción/efecto/practicidad	Tras usar ChatGPT disminuyeron errores gramaticales y ortográficos, mejoró la calidad global de escritura, aumentó la velocidad de escritura y se incrementaron la participación, satisfacción y motivación de aprendizaje	NR	NR
10	How et al., 2023	Hong Kong, China	Journal of Computing in Higher	Dos estudios de caso exploratorios	Métodos mixtos	67	Estudiantes	Chatbots basados en reglas	Educación superior totalmente en línea	Engagement conductual, utilidad percibida, facilidad de uso, percepciones	Registros conversacionales, cuestionarios Likert, entrevistas	Experiencias positivas con ambos chatbots; evidencia de engagement conductual y percepciones favorables de utilidad y facilidad de uso.	NR	NR

			Educ ation											
1 1	Ki m y Su, 202 4	Austr alia	Syste m	Cuasi- experi mental con grupo control	M ét o d o s m i x t o s	65	Estudia ntes	Chatbot de IA	Aprendizaj e de coreano como lengua extranjera en educación superior	WTC, ansiedad, confianza comunicativa	Cuestionario WTC, entrevistas semiestructurada s	Mejoró significativamente la willingness to communicate, redujo ansiedad y aumentó confianza comunicativa en el grupo experimental.	Un semest re	NR
1 2	Li, 202 5	China	Front iers in Psyc holog y	Transve rsal correla cional	C ua nt it at iv a	247	Estudia ntes univers itarios	AI chatbots-assisted learning	Aprendizaj e de matemátic as en educación superior	Learning interest, self- efficacy, performance expectancy, effort expectancy, math anxiety	Encuesta; PLS- SEM	Performance expectancy, effort expectancy y self-efficacy predijeron significativamente el learning interest; effort expectancy mostró el mayor efecto.	NR	NR
1 3	Lie w et al., 202 5	Mala sia	Acta Psyc holog ica	Experi mental within- subject s contrab alancea do	M ét o d o s m i x t o s	NR	Estudia ntes univers itarios	ChatGPT-powered educational chatbot	Educación superior	Engagement, emociones positivas, arousal, human- likeness, carga cognitiva	Experimento de laboratorio, cuestionarios, comentarios cualitativos	El slang juvenil aumentó human- likeness, engagement y emociones positivas, pero también incrementó la dificultad percibida/carga cognitiva.	NR	NR
1 4	Ng et al., 202 5	Reino Unid o	Beha viora l Scien ces	Ensayo aleatori zado crossov er	M ét o d o s m i x t o s	20	Estudia ntes de medici na	Chatbot GPT-4o específico de dominio	Educación médica de pregrado	Performance, ease of use, satisfaction, engagement, confidence, quality, clarity, trust	SBA tests, encuestas Likert post-task, focus groups	El chatbot superó a los métodos convencionales en facilidad de uso, satisfacción, engagement, calidad y claridad, pero no mostró ganancias significativas de rendimiento.	2024	NR
1 5	Or et al., 202 6	Austr alia	Jour nal of Dent al	Estudio piloto evaluati vo	M ét o d o s	31	Estudia ntes de odonto logía	Chatbot generativo GPT-4/Turbo como paciente virtual	Educación odontológi ca preclínica	Motivación, disfrute, participación, competencia percibida, usabilidad	Encuestas, comparación con enseñanza presencial, análisis temático	Mejoró la competencia percibida; aumentó participación y oportunidades de práctica; fue visto como menos interesante que la	2024	NR

			Educ ation		m ixt os							enseñanza presencial, pero útil como complemento curricular.		
16	Ortega - Oc ho a et al., 2024	España	Inter net of Thin gs	Cuasi-experi mental post-test only con grupos intactos	Cua nti tat iva	196	Estu dian tes	Pedagogical conversational agent empático	Curso online de Distributed Systems en educación superior	Learning performance, motivation, self-regulation, metacognitive reasoning	Instrumentos posttest, estadísticas descriptivas e inferenciales	No hubo diferencias significativas en la mayoría de los indicadores de motivación y autorregulación frente al docente; el chatbot fue tan efectivo como el feedback humano y algunos tipos de feedback se asociaron fuertemente con razonamiento metacognitivo.	Un término académico reciente	NR
17	Quah et al., 2026	Singapur	BMC Medical Education	Estudio comparativo con fase piloto	Métodos mixtos	130	Estu dian tes de odonto logía	Plataforma GPT-4 con virtual patient y virtual preceptor	Currículo de pregrado en medicina oral y patología oral	Aceptación, engagement, confianza, conciencia de progreso, conocimiento	Formulario de feedback, análisis temático, pruebas pre y post	Más de 90% quiso más oportunidades similares; el grupo intervención tuvo puntuaciones posttest significativamente mayores, aunque la mejora entre grupos no fue significativa.	NR	NR
18	Rahimi et al., 2025	Irán	Computers in Human Behavior Reports	Transversal analítico	Cua nti tat iva	133	Estu dian tes EFL	ChatGPT	Aprendizaje de idiomas en instituto de lenguas	PEL2MSS, PESRL	Cuestionario; PLS-SEM y ANN	Current L2-self, digital self-authenticity y ought-to L2-self fueron factores motivacionales significativos que afectaron la autorregulación personalizada con ChatGPT.	NR	NR
19	Liu, 2026	China	Acta Psychologica	Estudio observacional longitudinal de tres olas	Cua nti tat iva; PL S-	463	Estu dian tes universitarios de EFL	Chatbot generativo de IA (ChatGPT) para práctica de inglés; el curso también incluyó un asistente de escritura con IA, pero el modelo estimó de forma	Educación superior; clases universitarias de inglés en tres universidades chinas	AI-Chatbot-Assisted English Practice (ACEP), e-Instructor Facilitation (EIF), e-Communication Self-Efficacy (ECSE), e-Social English Learning Exposure (ESEL), e-Motivation for English Mastery (EMM)	Encuesta de tres olas por autoinforme	La práctica con chatbot predijo positivamente la motivación en inglés ($\beta = 0.372$, $p < 0.01$) y la autoeficacia comunicativa ($\beta = 0.272$, $p < 0.01$); la autoeficacia medió el efecto del chatbot sobre la motivación ($\beta = 0.042$, $p = 0.03$). El efecto moderado por exposición social digital no fue significativo para la ruta del chatbot.	Un semestre	NR

					SE M			separada la variable ACEP						
20	Rücker y Becker - Genschow, 2025	Alemania	Computers and Education Open	Ensayo controlado aleatorizado por conglomerados	Cuantitativa	195	Estudiantes de noveno grado	Chatbot de IA específico de dominio (ADA)	Educación secundaria; matemáticas/STEM	Aceptación tecnológica (TAM), emociones, situacional, cognitiva, rendimiento	Evaluación pre-post; escalas TAM; medidas de emociones, interés situacional y carga cognitiva; prueba de rendimiento	Alta aceptación; mejora significativa del interés situacional; tendencia no significativa a mayor carga cognitiva; tendencias favorables en rendimiento y emociones.	Intervención de una sola lección	Material suplementario en línea; código/datos: NR.
21	Shen, 2025	Tailandia	Computers and Education Open	Diseño mixto secuencial explicativo	Mixta	300 en encuesta + 10 en entrevistas	Estudiantes universitarios	Agente conversacional basado en LLM	Educación superior multicultural	Facilidad de comunicación, presencia social, utilidad percibida, confianza, engagement, satisfacción/aceptación	Encuesta Likert; CFA/SEM; entrevistas semiestructuradas	La presencia social incrementa confianza y satisfacción; la utilidad percibida incrementa engagement y satisfacción; el uso fácil aumentó presencia y utilidad, pero también elevó expectativas.	Entrevistas: 19 abril–20 mayo 2025; periodo de encuesta: NR.	NR
22	Wei et al., 2025	China	IEEE Access	Estudio no aleatorizado, transversal, basado en encuesta	Cuantitativa	351	Estudiantes EFL usuarios de chatbots	Chatbots de IA para aprendizaje oral de inglés	Aprendizaje de inglés como lengua extranjera	Human likeness, autoeficacia, presencia social, motivación de aprendizaje, resultados de aprendizaje, competencia técnica	Cuestionario online; escalas Likert; SEM; análisis moderador con PROCESS	Human likeness, autoeficacia y presencia social predicen fuertemente la motivación; la motivación predice resultados; la competencia técnica modera dichas relaciones.	1–31 julio 2024	NR
23	Zhai et al., 2024	NR	Computers and Education	Estudio de caso cualitativo	Cualitativa	37	Aprendices EAL entrevistados	Sistema de diálogo con IA (MACHE-Bot)	Aprendizaje de inglés como lengua adicional	Dimensiones interculturales, humor, empatía, confianza, antropomorfismo, autodescubrimiento	Entrevistas semiestructuradas en profundidad por Zoom; análisis	El bot fue bien recibido; se identificaron cinco factores clave: dimensiones interculturales, confianza, antropomorfismo, autodescubrimiento recíproco y mayor engagement/motivación.	Prueba piloto de 14 días; reclutamiento	Datos disponibles a solicitud por correo

			: Artifi- cial Intell igenc e						recíproco, engagement motivación	y	cualitativo con NVivo 12		anunci- ado durant e 20 días	al autor corres- pondie- nte.
24	Zhang et al., 2025	China (Hong Kong SAR China, afiliación institucional)	Computers and Education: Artificial Intelligence	Estudio mixto	Mixta	42	Estudiantes universitarios EFL	Chatbot GPT-4 específico de disciplina (LogicalHamster)	Educación superior; escritura argumentativa en inglés	Engagement en respuestas GPT y prompts; conocimiento lógico; lógica en escritura argumentativa	Eye-tracking Tobii; pruebas pre-post-delayed; tareas de escritura pre-post; entrevistas semiestructuradas	El engagement con respuestas GPT favoreció el desarrollo del conocimiento lógico, especialmente la retención; el engagement con prompts podría aumentar carga cognitiva extrínseca; no hubo mejora significativa en la lógica de escritura.	45–75 minutos por sesión; fecha global del estudio: NR	NR

Fuente: elaboración propia

Anexo 2

Tabla 3 — SÍNTESIS NARRATIVA DE LA EVIDENCIA

	Author/Year	Design	Type of Evidence	Focus	Result	Contribution	Domain	Practical Utility	Confidence
1	Abdelwahab et al., 2025	Cuasi experimental con grupo control	Cuantitativa	Efecto de chatbot educativo en motivación académica y desempeño clínico en enfermería	Positivo	Evidencia de mejora sostenida en motivación académica, conocimiento, interpretación clínica, razonamiento y satisfacción	A, B, D	Alta	Moderada
2	Al-Abdullatif et al., 2023	Cuasi experimental	Cuantitativa	Influencia del chatbot Bashayer en motivación y estrategias de aprendizaje en posgrado	Positivo	Evidencia de aumento en motivación, valor de tarea, autoeficacia y estrategias cognitivas/metacognitivas	A, B, D	Alta	Moderada
3	Annamalai et al., 2023	Estudio de caso cualitativo	Cualitativa	Motivación para aprender inglés vía chatbot desde la teoría de la autodeterminación	Mixto positivo	Aporta explicación teórica de cómo el chatbot favorece autonomía, competencia y relación, pero también muestra límites afectivos y de precisión	A, B, D	Media	Moderada
4	Aslan et al., 2024	Cuasi experimental + caso exploratorio	Mixta	Engagement, experiencia y rendimiento con agente conversacional pedagógico multimodal en infancia temprana	Positivo	Evidencia de alto engagement conductual, emocional y social, con reducción de pantalla y mejora del rendimiento	A, C, E	Alta	Moderada
5	Ayedoun et al., 2020	Exploratorio empírico	Cuantitativa	Apoyo motivacional personalizado para aumentar la disposición a comunicarse en L2	Positivo	Aporta evidencia preliminar de que un agente conversacional puede modular un constructo motivacional relevante y que la preferencia por estrategias conversacionales importa para el efecto observado.	A, B, D	Media	Baja
6	Chien et al., 2022	Cuasi experimental	Cuantitativa	Uso de LINE ChatBot en aprendizaje contextual de inglés con componente competitivo	Mixto positivo	Muestra que el entorno con chatbot se asocia con mejoras en motivación y práctica oral, aunque la competencia no produjo ventajas significativas en rendimiento entre grupos.	A, B	Alta	Moderada
7	García et al., 2025	Exploratorio de métodos mixtos	Mixta	Agente pedagógico para programación basado en worked examples	Positivo	Refuerza la idea de que un chatbot pedagógico puede apoyar comprensión, motivación, autoconfianza y	A, B, C	Alta	Moderada

						engagement en programación, especialmente mediante ejemplos correctos y erróneos guiados.			
8	Hall et al., 2026	Piloto de métodos mixtos	Mixta	Pacientes simulados chatbot para entrenamiento clínico en farmacia	Positivo	Aporta evidencia aplicada de que los chatbots como pacientes simulados pueden aumentar engagement, confianza y autenticidad percibida en tareas clínicas de comunicación.	A, C, D	Alta	Moderada-baja
9	He, 2025	Estudio de caso empírico pretest-postest	Mixta	Uso de ChatGPT para apoyo personalizado en escritura en inglés y su efecto en motivación, participación y desempeño	Positivo	Aporta evidencia preliminar de que un sistema conversacional basado en ChatGPT puede apoyar la escritura en inglés mediante retroalimentación en tiempo real, con mejoras en calidad textual, participación, satisfacción y motivación	A, B, C	Media	Baja
10	Hew et al., 2023	Dos estudios de caso exploratorios	Mixta	Goal setting, social presence y engagement online	Positive	Muestra que un chatbot pedagógicamente diseñado puede apoyar engagement y autorregulación en cursos totalmente en línea.	A,C,D,E	High	Moderate
11	Kim y Su, 2024	Cuasi-experimental con apoyo cualitativo	Mixta	WTC, ansiedad y confianza comunicativa	Positive	Aporta evidencia de que el chatbot puede mejorar una variable motivacional central en aprendizaje de lenguas: la disposición a comunicarse.	A,B,C,D	High	Moderate
12	Li, 2025	Transversal correlacional	Cuantitativa	Interés de aprendizaje en matemáticas asistidas por chatbot	Positive	Integra aceptación tecnológica y motivación, identificando predictores del learning interest en contexto chatbot.	B,D,E	Medium	Moderate
13	Liew et al., 2025	Experimental within-subjects	Mixta	Efectos afectivos, engagement y carga cognitiva del estilo lingüístico	Mixed	Evidencia que el diseño lingüístico del chatbot puede aumentar engagement y emoción positiva, pero con coste cognitivo.	A,C,E	High	Moderate
14	Ng et al., 2025	Ensayo aleatorizado crossover	Mixta	Usabilidad, engagement y profundidad cognitiva	Mixed	Muestra una paradoja eficiencia-profundidad: el chatbot mejora engagement y usabilidad, pero no el rendimiento ni el pensamiento profundo.	A,C,E	High	Moderate
15	Or et al., 2026	Estudio piloto evaluativo	Mixta	Motivación, disfrute y participación en history-taking	Mixed	Sugiere valor del chatbot como recurso complementario para práctica segura y aumento de	A,B,C,D	High	Low

						participación, aunque menos interesante que la enseñanza presencial.			
16	Ortega-Ochoa et al., 2024	Cuasi-experimental	Cuantitativa	Feedback empático, motivación, autorregulación y metacognición	Mixed	Indica equivalencia general frente al feedback humano, con ventajas específicas de ciertos tipos de feedback para autorregulación y razonamiento metacognitivo.	A,B,D,E	High	Moderate
17	Quah et al., 2026	Comparativo con fase piloto	Mixta	Aceptación, engagement y aprendizaje en medicina/patología oral	Positive	Muestra alta aceptación, engagement e interés futuro, con resultados de conocimiento al menos equivalentes y potencialmente superiores al método tradicional.	A,B,C,E	High	Moderate
18	Rahimi et al., 2025	Transversal analítico	Cuantitativa	Motivación L2 personalizada y autorregulación con ChatGPT	Positive	Aporta un modelo explicativo de cómo factores motivacionales específicos predicen autorregulación en aprendizaje asistido por ChatGPT.	A,B,D	Medium	Moderate
19	Liu, 2026	Longitudinal de tres olas	Cuantitativa	Efecto de la práctica con chatbot sobre motivación y autoeficacia en inglés online	Positive	Aporta evidencia de que la práctica con chatbot, diferenciada analíticamente dentro de un entorno AI-rich, se asocia con mayor motivación y mayor autoeficacia comunicativa, y que la autoeficacia media esa relación. Su principal valor teórico es integrar chatbot, facilitación docente y exposición digital dentro de un marco ecológico.	A,B,D,E	Alta	Moderate
20	Rücker y Becker-Genschow, 2025	Ensayo controlado aleatorizado por conglomerados	Cuantitativa experimental	Efecto de un chatbot específico de dominio en interés situacional, aceptación, emociones, carga cognitiva y rendimiento en matemáticas	Positive	Evidencia experimental relativamente sólida de que un chatbot específico de dominio puede mejorar el interés situacional en educación STEM, con alta aceptación estudiantil.	A; B; C; D	High	Moderate
21	Shen, 2025	Mixto secuencial explicativo	Mixta (encuesta SEM + entrevistas)	Cómo la presencia social, confianza y utilidad percibida median engagement y satisfacción con un comunicador pedagógico de IA	Positive	Integra TAM con presencia social, confianza y engagement en un contexto multicultural, mostrando rutas mediadas hacia la satisfacción.	A; B; C; D	High	Moderate
22	Wei et al., 2025	Estudio transversal no	Cuantitativa observacional	Factores que influyen en la motivación de aprendizaje oral con chatbots de IA y su relación con resultados de aprendizaje	Positive	Aporta evidencia correlacional de que human likeness, autoeficacia y presencia social se asocian positivamente	D	Medium	Moderate

		aleatorizado por encuesta				con la motivación, y que la competencia técnica modera estas relaciones.			
23	Zhai et al., 2024	Estudio de caso cualitativo	Cualitativa	Rol del humor, empatía e interculturalidad en engagement y motivación en aprendizaje de inglés	Positive	Propone cinco factores clave para el uso de sistemas de diálogo en EAL y destaca la relevancia del humor, la empatía y la confianza en la experiencia de aprendizaje.	A; B; C; D	Medium	Low
24	Zhang et al., 2025	Estudio mixto	Mixta (eye-tracking, pruebas, tareas, entrevistas)	Diferencias de engagement entre respuestas GPT y prompts del estudiante, y sus efectos sobre conocimiento lógico y escritura argumentativa	Mixed	Distingue microcomponentes del engagement en aprendizaje con ChatGPT y muestra que el engagement con respuestas GPT favorece conocimiento lógico, mientras que el engagement con prompts puede aumentar carga cognitiva y no mejorar significativamente la escritura lógica.	A; B; C; D	High	Moderate

Anexo 3.

Tabla 2. Calidad metodológica / riesgo de sesgo

	Author/Year	Design	Tool	Selection	Measurement	Confounding	Analysis	Reporting	Overall Risk	Justification
1	Abdelwahab et al., 2025	No aleatorizado	ROBINS-I	Alto	Moderado	Alto	Bajo	Moderado	Alto	Muestreo intencional y asignación no aleatoria; uso de instrumentos estructurados y análisis estadístico sólido; posible sesgo de confusión propio del diseño cuasi experimental; reporte adecuado pero sin disponibilidad explícita de datos/código.
2	Al-Abdullatif et al., 2023	No aleatorizado	ROBINS-I	Alto	Moderado	Alto	Moderado	Moderado	Alto	Diseño cuasi experimental sin aleatorización reportada; comparación entre grupos con MSLQ y t de Student; riesgo de confusión y sesgo de selección; el reporte permite identificar intervención y resultados, pero no informa datos/código.
3	Annamalai et al., 2023	Estudio cualitativo	CASP cualitativo	Moderado	Moderado	Moderado	Moderado	Moderado	Moderado	El estudio es apropiado para explorar motivación, pero la muestra es pequeña y no probabilística; usa entrevistas y análisis temático; no reporta datos/código ni estrategias extensas de control de sesgo interpretativo.
4	Aslan et al., 2024	No aleatorizado / mixto	MMAT	Moderado	Bajo	Moderado	Bajo	Bajo	Moderado	Diseño mixto bien descrito, con triangulación, observación semiestructurada, codificación multimodal e IRR; persiste riesgo de

										confusión por ausencia de aleatorización y tamaño muestral reducido; el reporte es fuerte y la disponibilidad de datos está condicionada por privacidad.
5	Ayedoun et al., 2020	No aleatorizado exploratorio	ROBINS-I	Alto	Moderado	Alto	Moderado	Moderado	Alto	El estudio es empírico y el outcome motivacional está bien definido, pero en el texto disponible no se informa con claridad la muestra ni se observa aleatorización; además, la inferencia causal es limitada y los procedimientos de control de sesgos no se detallan suficientemente.
6	Chien et al., 2022	Cuasi experimental no aleatorizado	ROBINS-I	Alto	Moderado	Alto	Bajo	Moderado	Alto	Hay pretest-posttest y ANCOVA, pero no hubo asignación aleatoria individual; existe riesgo de sesgo de selección y de confusión entre clases. La medición usa instrumentos explícitos, incluida MSLQ, y el análisis estadístico es adecuado.
7	Garcia et al., 2025	Métodos mixtos exploratorio	MMAT	Moderado	Moderado	Moderado	Moderado	Moderado	Moderado	El estudio describe con claridad la muestra, instrumentos y análisis cuantitativo-cualitativo; sin embargo, se centra en percepciones y no presenta comparación experimental, por lo que persisten riesgos de sesgo de selección, autoinforme y ausencia de control de confusores.
8	Hall et al., 2026	Piloto de métodos mixtos	MMAT	Alto	Moderado	Alto	Moderado	Moderado	Alto	Aunque el contexto y los instrumentos están claramente descritos, la muestra efectiva fue muy pequeña, dependió de

										autoinforme y no hubo grupo comparador ni seguimiento a largo plazo; los propios autores reconocen estas limitaciones.
9	He, 2025	Estudio no aleatorizado tipo caso con comparación antes/después	JB1 cuasi experimental	Alto	Alto	Alto	Moderado	Alto	Alto	El estudio presenta contexto educativo y datos empíricos, pero la muestra es muy pequeña (cinco estudiantes), no se reporta procedimiento de selección robusto ni grupo control, y la intervención combina varios componentes algorítmicos. La medición parece basarse en indicadores internos del sistema y seguimiento de pocos casos, con escasa información sobre validez de instrumentos y control de variables de confusión. El reporte ofrece resultados descriptivos de mejora, participación y satisfacción, pero la trazabilidad metodológica sigue siendo limitada.
10	Hew et al., 2023	Mixto	MMAT	Moderado	Moderado	Alto	Moderado	Moderado	Moderate	Estudios de caso sin control ni aleatorización; muestra pequeña; medición multimétodo adecuada pero centrada en engagement/percepciones; control limitado de confundidores.
11	Kim y Su, 2024	Cuasi-experimental/mixto	MMAT	Moderado	Moderado	Moderado	Moderado	Bajo	Moderate	Hubo grupo control y comparación pre-post, pero clases desiguales y sin evidencia de cegamiento; outcomes bien definidos y apoyo cualitativo.

12	Li, 2025	Observacional transversal	ROBINS-I	Moderado	Moderado	Alto	Moderado	Bajo	Moderate	Estudio no experimental; no permite causalidad; medición por autoinforme; análisis PLS-SEM adecuado, pero con riesgo de confusión residual.
13	Liew et al., 2025	Experimental/mixto	MMAT	Moderado	Bajo	Moderado	Moderado	Bajo	Moderate	Diseño contrabalanceado fortalece la comparación, pero el tamaño muestral no está reportado en el extracto disponible; outcomes múltiples y posible efecto de novedad.
14	Ng et al., 2025	Ensayo aleatorizado crossover mixto	MMAT	Moderado	Bajo	Moderado	Moderado	Bajo	Moderate	Diseño aleatorizado crossover sólido, pero muestra muy pequeña y posible sesgo de autoselección por reclutamiento mediante pósteres; sin cegamiento evidente.
15	Or et al., 2026	Piloto mixto no aleatorizado	MMAT	Alto	Moderado	Alto	Moderado	Moderado	High	Muestra voluntaria pequeña fuera del horario curricular; probable sesgo de confirmación/early adopters; outcomes principalmente autoinformados y sin grupo control.
16	Ortega-Ochoa et al., 2024	Cuasi-experimental	ROBINS-I	Moderado	Moderado	Moderado	Moderado	Bajo	Moderate	Grupos intactos por idioma y sin aleatorización individual; instrumentos cuantitativos y análisis inferenciales adecuados; riesgo de confusión por diferencias entre clusters.
17	Quah et al., 2026	Mixto comparativo	MMAT	Moderado	Moderado	Moderado	Moderado	Bajo	Moderate	Fase piloto más fase comparativa; instructor no cegado y posible variación del tiempo de interacción; análisis estadístico apropiado y resultados claramente reportados.

18	Rahimi et al., 2025	Observacional transversal	ROBINS-I	Moderado	Moderado	Alto	Moderado	Bajo	Moderate	Encuesta transversal y autoinforme; análisis avanzado PLS-SEM-ANN, pero sin precedencia temporal ni control experimental; validez de constructo reportada.
19	Liu, 2026	Observacional longitudinal; riesgo moderado	ROBINS-I	Riesgo moderado-alto	Riesgo moderado	Riesgo alto	Riesgo bajo-moderado	Riesgo bajo	Moderate	El estudio tiene fortaleza analítica y temporal al usar encuesta en tres olas y PLS-SEM, además de pruebas de sesgo de método común y robustez. Sin embargo, usa muestreo por conveniencia, autoinforme para medir uso del chatbot en lugar de registros de plataforma, y el contexto instruccional incluyó también un asistente de escritura con IA, lo que introduce confusión potencial en la atribución del efecto específico al chatbot. Aun así, el artículo modeló explícitamente ACEP como constructo diferenciado y reportó resultados de validez y robustez.
20	Rücker y Becker-Genschow, 2025	Ensayo controlado aleatorizado por conglomerados	RoB 2	Bajo	Moderado	Moderado	Bajo	Moderado	Moderate	La asignación fue por conglomerados y hubo grupo control, lo que reduce sesgo de selección. Sin embargo, la intervención fue muy breve y el grupo control mostró diferencias iniciales en pre-test; además, parte de las medidas fueron autorreportadas y no se reporta disponibilidad de datos/código.
21	Shen, 2025	Mixto secuencial explicativo	MMAT	Moderado	Moderado	Moderado	Bajo	Moderado	Moderate	La encuesta incluyó N=300 y las entrevistas fueron intencionales por

										máxima variación; no obstante, no se informó periodo de la encuesta y no se incluyeron controles demográficos en el SEM principal. Las medidas y el análisis estuvieron claramente especificados, con CFA/SEM, bootstrap y triangulación.
22	Wei et al., 2025	Estudio transversal no aleatorizado por encuesta	ROBINS-I	Alto	Moderado	Alto	Moderado	Moderado	High	Se utilizó muestreo por conveniencia y autoinforme, lo que incrementa sesgo de selección y medición. Aunque el SEM y PROCESS fueron apropiados para el modelo planteado, el diseño transversal limita control de confusión y causalidad.
23	Zhai et al., 2024	Estudio de caso cualitativo	JB1	Alto	Moderado	Moderado	Moderado	Moderado	Moderate	El reclutamiento fue por conveniencia desde un grupo de Facebook, lo que eleva sesgo de selección. No obstante, hubo protocolo de entrevista, validación por expertos, análisis con NVivo y descripción del proceso de codificación; los datos están disponibles bajo solicitud.
24	Zhang et al., 2025	Estudio mixto	MMAT	Moderado	Bajo	Moderado	Bajo	Moderado	Moderate	El tamaño muestral fue pequeño (N=42) y la generalización es limitada, pero la medición combinó eye-tracking, pruebas, tareas y entrevistas, lo que fortalece la validez. El análisis fue multimétodo y el instrumento se pilotó previamente; no se reporta disponibilidad de datos/código.

Capítulo 2

**REVISIÓN SISTEMÁTICA DEL USO DE INTELIGENCIA ARTIFICIAL
GENERATIVA EN PROCESOS DE APRENDIZAJE DESDE UN ENFOQUE
NEUROEDUCATIVO**

Angel Onzari Luna Santos²

RESUMEN

Esta revisión sistemática analizó la literatura científica sobre el uso de inteligencia artificial generativa (GenAI) en procesos de aprendizaje e interpretó sus efectos desde un enfoque neuroeducativo. Se siguió el protocolo PRISMA 2020 mediante búsquedas en seis bases de datos —ScienceDirect, IEEE Xplore, SciELO, Emerald, PubMed y Frontiers— con cobertura temporal 2020–2026, de las cuales se seleccionaron 28 estudios empíricos tras un proceso de cribado en dos fases con evaluación de calidad metodológica mediante ROBINS-I, RoB 2, MMAT y JBI según el diseño de cada estudio. El corpus, con riesgo global moderado, se distribuyó en cinco dominios temáticos con predominancia del pedagógico/adopción, y reveló tres convergencias —efectos positivos sobre motivación y engagement, mediación de la utilidad percibida en la adopción, y personalización dinámica como atributo diferencial de GenAI— junto con divergencias sistemáticas en los efectos sobre rendimiento académico objetivo, carga cognitiva y dimensiones socioemocionales. El análisis integrado identificó tres patrones transversales interpretados desde el marco de la cognición extendida: la mediación afectivo-motivacional como mecanismo de entrada al acoplamiento cognitivo, la alfabetización en IA como condición de posibilidad del acoplamiento funcional, y el diseño conversacional

² Doctor en Contabilidad por la Universidad Nacional José Faustino Sánchez Carrión E-mail: aluna@unjfsc.edu.pe, ORCID: <https://orcid.org/0000-0001-6751-504X>

de la herramienta como determinante del modo de acoplamiento —sustitutivo o extensivo—. Se concluye que los efectos de GenAI en el aprendizaje son sistémicamente dependientes del perfil del aprendiz y del diseño instruccional, y que el avance del campo requiere diseños longitudinales, medidas neurocognitivas objetivas e integración explícita de marcos neuroeducativos.

Palabras clave: inteligencia artificial generativa, neuroeducación, cognición extendida, procesos de aprendizaje, revisión sistemática, carga cognitiva, metacognición, alfabetización en inteligencia artificial

INTRODUCCIÓN

1. Introducción

1.1 La inteligencia artificial generativa en educación: un fenómeno de expansión sin precedentes

La velocidad con que la inteligencia artificial generativa ha penetrado los entornos educativos supera, con diferencia, la capacidad del campo científico para comprender los efectos que esa penetración produce en los procesos cognitivos que sostienen el aprendizaje humano. Esta asimetría entre adopción y comprensión no es un déficit pasajero atribuible a la novedad del fenómeno: es, en cambio, el síntoma de una tensión estructural entre una tecnología que opera sobre los mecanismos del pensamiento y una investigación educativa que aún no dispone de los marcos teóricos necesarios para explicar, y no solo describir, lo que ocurre cuando un aprendiz externaliza parte de su procesamiento cognitivo a un sistema generativo. Comprender esa tensión es el problema científico que esta revisión sistemática se propone abordar.

La inteligencia artificial generativa —en adelante GenAI— designa una clase de sistemas de inteligencia artificial basados en modelos de lenguaje de gran escala (LLMs) capaces de producir contenido nuevo —texto, código, imágenes, retroalimentación— en respuesta a instrucciones en lenguaje natural (Gozalo-Brizuela & Garrido-Merchán, 2023). Lo que

distingue a GenAI de los sistemas de inteligencia artificial educativa anteriores no es solo su capacidad de generación, sino la naturaleza de la interacción que habilita: a diferencia de los sistemas expertos o de los tutores inteligentes basados en reglas, GenAI produce respuestas adaptadas al contexto conversacional del aprendiz, genera retroalimentación no predefinida, y puede asumir múltiples roles pedagógicos —tutor, interlocutor, evaluador, asistente de escritura— en el mismo entorno de aprendizaje (Gibson et al., 2023). Esta distinción es educativamente relevante porque implica que los efectos cognitivos de GenAI no son los de una herramienta de consulta estática sino los de un sistema que modifica activamente las condiciones en que el aprendiz procesa información, regula su atención y construye conocimiento. Para los fines de esta revisión, el término GenAI abarca los modelos de lenguaje de gran escala, los chatbots generativos, los asistentes de escritura y generación de contenido basados en LLMs, y los sistemas de retroalimentación automatizada de naturaleza generativa, aplicados en contextos de aprendizaje formal o no formal.

La expansión de estas herramientas en los entornos educativos ha sido tanto cuantitativa como cualitativamente significativa. Estudios con cobertura internacional documentan que más del sesenta por ciento de los estudiantes universitarios reconoce utilizar herramientas GenAI para tareas académicas, con tendencias de adopción que en algunos contextos superan el noventa por ciento en el período 2024-2025 (Ogunleye et al., 2024; Marzano, 2025). Las modalidades de uso son heterogéneas: asistencia en escritura académica, generación de materiales de estudio, obtención de retroalimentación inmediata, resolución de ejercicios, práctica conversacional en idiomas extranjeros, personalización de contenidos y apoyo en la fase de preescritura y planificación. Esa heterogeneidad de usos, presente en niveles educativos que van de la educación primaria a la formación posgraduada y en disciplinas tan diversas como la física, la música, la danza y la enseñanza de lenguas extranjeras, es precisamente la razón por la que una síntesis sistemática resulta necesaria: la multiplicidad de contextos no ha producido

un cuerpo de conocimiento internamente coherente sino, hasta ahora, una acumulación de hallazgos cuya consistencia y transferibilidad están por determinar.

1.2 El campo en construcción: avances, tensiones y fragmentación teórica

1.2.1 Lo que el campo ha establecido: convergencias con base empírica

La investigación sobre GenAI en educación ha avanzado de forma notable en los últimos tres años, y sería inexacto presentar el campo como un territorio sin hallazgos consolidados. La revisión de la literatura disponible permite identificar tres convergencias con suficiente respaldo empírico para ser consideradas tendencias del campo, aunque con las reservas metodológicas que se precisan a continuación.

La primera y más documentada es el efecto positivo de GenAI sobre la motivación y el engagement del aprendiz. Revisiones con cobertura amplia de estudios primarios coinciden en señalar que la interacción con herramientas GenAI incrementa el interés situacional, la disposición a participar en las tareas y la percepción de relevancia del aprendizaje, con independencia del nivel educativo y del contexto disciplinar (Yan et al., 2024; Qian, 2025). La segunda convergencia es el atributo diferencial de personalización dinámica: la capacidad de GenAI para adaptar materiales, ejemplos, retroalimentación y nivel de dificultad al perfil individual del aprendiz en tiempo real constituye su valor añadido más consistentemente documentado respecto a tecnologías educativas previas (Guettala et al., 2024; Bahroun et al., 2023). La tercera convergencia es la mejora en el desempeño en tareas inmediatas — producción escrita, resolución de ejercicios, práctica en lenguas extranjeras— cuando GenAI actúa como herramienta de apoyo activo durante la ejecución de la tarea (Li et al., 2025; Pesovski et al., 2024). Estas tres tendencias, sin embargo, comparten una característica metodológica que limita su fuerza inferencial: descansan mayoritariamente en diseños correlacionales transversales y en medidas autoinformadas, lo que impide descartar explicaciones alternativas —como el efecto de novedad tecnológica— y no permite establecer

si los efectos observados se sostienen más allá del período de intervención. Esta base metodológica es la que el siguiente bloque examina críticamente.

1.2.2 Tensiones internas: contradicciones no resueltas en la evidencia

La coexistencia de los hallazgos anteriores con resultados opuestos en la misma literatura no puede atribuirse a errores de estudios individuales. Las contradicciones identificadas son sistemáticas y revelan que el campo está midiendo constructos diferentes bajo los mismos nombres, en contextos cuya comparabilidad no ha sido establecida, sin marcos de referencia compartidos que permitan interpretar cuándo un resultado contradice a otro y cuándo simplemente mide algo distinto.

La primera tensión concierne a la relación entre GenAI y los procesos cognitivos de orden superior. Mientras algunos estudios documentan mejoras en pensamiento crítico, metacognición y autorregulación cuando GenAI actúa como interlocutor activo en el proceso de aprendizaje (Xu et al., 2025; Tomisu et al., 2025), otros evidencian el fenómeno opuesto: la "pereza metacognitiva", entendida como la tendencia del aprendiz a delegar en el sistema generativo las operaciones de monitoreo, evaluación y ajuste de la propia comprensión, reduciendo su implicación en los procesos metacognitivos que hacen el aprendizaje duradero (Fan et al., 2024). Estudios experimentales con condiciones controladas muestran que GenAI puede mejorar significativamente el desempeño en tareas inmediatas sin producir ganancias equivalentes en la transferencia del conocimiento ni en la motivación intrínseca, sugiriendo que el incremento en el rendimiento observable no es evidencia suficiente de aprendizaje profundo (Fan et al., 2024; Li et al., 2025).

La segunda tensión afecta a los efectos sobre la carga cognitiva. La teoría de la carga cognitiva predice que las herramientas de andamiaje bien diseñadas deberían reducir la carga extrínseca y liberar recursos de memoria de trabajo para el procesamiento germano (Sweller, 2020). La evidencia empírica sobre GenAI en relación con este principio es, sin embargo, bidireccional:

algunos diseños de intervención documentan reducción de la frustración y la carga percibida en entornos de microaprendizaje asistido por IA, mientras que otros evidencian incrementos en la carga cognitiva extrínseca asociados a características específicas del diseño conversacional de la herramienta —como el registro lingüístico empleado o la granularidad de la retroalimentación— que el aprendiz debe procesar adicionalmente (Collie & Martin, 2025; Liew et al., 2025). La carga cognitiva, en consecuencia, no emerge como consecuencia automática del uso de GenAI sino como variable sensible a decisiones de diseño instruccional que la investigación existente ha caracterizado de forma insuficiente.

La tercera tensión, de naturaleza socioemocional, refuerza el mismo patrón de bidireccionalidad: el uso de GenAI se asocia a mejoras en bienestar, reducción de ansiedad y desarrollo de una mentalidad de crecimiento en algunos contextos (Pan & Li, 2025; Huang et al., 2025), y a posibles reducciones en inteligencia emocional y en capacidad de interacción autónoma en otros (Kazazoglu & Turun Özel, 2025). La ausencia de replicación impide determinar qué variables moderan la dirección del efecto. Estas tres tensiones no son anomalías aisladas de la literatura: son el síntoma de un problema teórico de fondo que el siguiente bloque identifica con precisión.

1.2.3 La paradoja teórica central: el campo habla de cognición sin explicarla

El examen sistemático de la arquitectura teórica de la investigación sobre GenAI y aprendizaje revela un dato que define el estado actual del campo: la producción científica está dominada por marcos de aceptación tecnológica —TAM, UTAUT y sus extensiones— que explican con solidez los factores que determinan si un aprendiz adopta o rechaza una herramienta GenAI, pero que no ofrecen ningún principio explicativo sobre los mecanismos cognitivos que se activan durante ese uso ni sobre la calidad del aprendizaje que resulta de él (BinJwair, 2025; Kinskofer & Tulis, 2025; Yu et al., 2024; Wang & Liu, 2026). Esta es la brecha más significativa del campo, y su relevancia no reside en la ausencia de marcos teóricos alternativos

—que existen— sino en que los estudios que sí operacionalizan constructos cognitivos relevantes para el aprendizaje lo hacen empleando vocabulario propio de la neuroeducación y las ciencias cognitivas —carga cognitiva, metacognición, funciones ejecutivas, autorregulación, engagement multidimensional— sin fundamentarlos en los marcos neurocientíficos que les dan sustento explicativo (Liew et al., 2025; Li et al., 2025; Klarin et al., 2024).

El resultado de esta combinación es la paradoja teórica central del campo: el vocabulario neuroeducativo es ampliamente utilizado para describir los efectos de GenAI en el aprendizaje, pero los diseños de investigación empleados para medirlos raramente permiten fundamentar esas afirmaciones en mecanismos neurocognitivos verificables. Los constructos se emplean como etiquetas descriptivas de conductas observadas o percepciones autoinformadas, no como categorías explicativas ancladas en modelos de funcionamiento cognitivo. Esta situación no es atribuible a debilidades de estudios individuales, sino que constituye una característica estructural del campo en su estado actual: la investigación ha avanzado más rápido en describir qué ocurre con GenAI en el aula que en explicar por qué ocurre en términos de los mecanismos cognitivos subyacentes. Esa asimetría entre descripción y explicación es la que cristaliza en los vacíos que la siguiente sección delimita.

1.3 Brechas de investigación: vacíos empíricos y teóricos con precisión

1.3.1 Vacíos empíricos: qué falta medir y por qué importa

Los vacíos empíricos del campo pueden delimitarse en cinco dimensiones que no se solapan entre sí y que corresponden a ausencias en qué se ha investigado, con independencia de la calidad de lo disponible.

El primero es el vacío longitudinal. El corpus de investigación sobre GenAI en aprendizaje es casi enteramente transversal o está compuesto por intervenciones de duración inferior a un semestre, lo que impide determinar si los efectos documentados sobre motivación,

metacognición o engagement se sostienen, se intensifican o se diluyen con el tiempo y con el aumento de la familiaridad del aprendiz con la herramienta (Marzano, 2025; Zhang et al., 2026). La ausencia de diseños longitudinales es especialmente crítica para el enfoque neuroeducativo: los procesos de consolidación de memoria, automatización de esquemas cognitivos y desarrollo de la autorregulación ocurren en escalas temporales que ningún estudio transversal puede capturar.

El segundo es el vacío neurocognitivo. Ningún estudio disponible en la literatura sobre GenAI y aprendizaje emplea medidas objetivas de procesamiento cognitivo —neuroimagen funcional, espectroscopía de infrarrojo cercano funcional (fNIRS), eye-tracking, tiempos de reacción o registros electrofisiológicos— para verificar directamente qué ocurre en el sistema cognitivo del aprendiz durante la interacción con herramientas GenAI (Siemens et al., 2022). La evidencia permanece en el plano perceptual y conductual: lo que los aprendices dicen que ocurre, o lo que se puede inferir de sus productos, no lo que se puede observar en sus procesos. Este es el vacío más profundo del campo desde una perspectiva neuroeducativa, y su existencia no es solo una limitación metodológica sino una restricción epistemológica: sin medidas de procesamiento cognitivo, las afirmaciones sobre carga cognitiva, atención o memoria de trabajo que aparecen en la literatura son inferencias sobre mecanismos que no han sido directamente observados.

El tercero es el vacío poblacional. La investigación existente está concentrada en educación superior, con escasa representación de educación básica y de la etapa adolescente, que son precisamente los períodos de mayor plasticidad neuronal y de mayor relevancia para la formación de hábitos cognitivos que se consolidarán en etapas posteriores (Yan et al., 2024). El cuarto es el vacío geográfico: el campo está dominado por producción de China, Europa central y Oriente Medio, con ausencia casi total del Sur Global, África subsahariana y América Latina, lo que limita la generalización de los hallazgos a sistemas educativos con trayectorias

históricas, recursos tecnológicos y condiciones pedagógicas sustancialmente distintas. El quinto es el vacío de diseño experimental: los estudios con asignación aleatoria son una minoría exigua de la producción disponible, y en los casos existentes los tamaños muestrales son reducidos, lo que restringe el poder estadístico y la capacidad de establecer relaciones causales entre el uso de GenAI y resultados de aprendizaje verificables (Collie & Martin, 2025).

1.3.2 Vacíos teóricos: qué marcos hacen falta

Partiendo de la paradoja identificada en la sección anterior como premisa establecida, es posible especificar con precisión qué tipo de integración teórica está ausente en el campo y por qué su ausencia es relevante para la agenda científica.

No existe, hasta la fecha, una síntesis sistemática de la literatura sobre GenAI en aprendizaje que emplee explícitamente un marco neuroeducativo como lente interpretativa de la evidencia disponible. Las revisiones existentes han operado predominantemente desde tres perspectivas: la pedagógica general —examinando qué usos y efectos reporta la literatura sin anclarlos en modelos cognitivos—, la de adopción tecnológica —explicando las condiciones de uso sin abordar los procesos que ocurren durante ese uso— y la bibliométrica —describiendo el estado del campo mediante análisis de tendencias de publicación sin síntesis interpretativa de la evidencia primaria (Ogunleye et al., 2024; Qian, 2025; Bahroun et al., 2023). Ninguna de estas perspectivas ha articulado de manera sistemática los hallazgos disponibles con constructos neurocognitivos formales como la memoria de trabajo, la atención sostenida y selectiva, los mecanismos de consolidación mnémica, la regulación ejecutiva del aprendizaje o la neuroplasticidad asociada a la práctica deliberada. Esta ausencia no es trivial: sin ese marco, la evidencia disponible permite saber que GenAI produce efectos sobre variables relevantes para el aprendizaje, pero no permite explicar a través de qué mecanismos los produce ni, por tanto, qué condiciones de diseño instruccional maximizarían o protegerían esos mecanismos. La integración entre evidencia empírica y marco neuroeducativo es, en consecuencia, la

contribución teórica más urgente que el campo requiere en este momento de su desarrollo. El contraste con revisiones previas debe leerse como observación del estado del campo en conjunto, no como evaluación de trabajos individuales: la ausencia de ese marco en la literatura disponible es una característica estructural del campo, no una omisión de investigadores específicos.

1.4 Objetivo, preguntas de investigación y contribución del estudio

1.4.1 Objetivo y preguntas de investigación

El objetivo general de este estudio es analizar sistemáticamente la literatura científica sobre el uso de inteligencia artificial generativa en procesos de aprendizaje e interpretar sus efectos desde un enfoque neuroeducativo, con el propósito de identificar patrones transversales, tendencias del campo, beneficios documentados, limitaciones metodológicas y vacíos de investigación que orienten tanto la práctica pedagógica como la agenda científica futura.

De este objetivo se derivan cuatro preguntas de investigación que organizan el proceso analítico en progresión de lo descriptivo a lo interpretativo:

PI1: ¿Qué tipos de herramientas de inteligencia artificial generativa se están utilizando en contextos de aprendizaje formal y no formal, y en qué modalidades de uso se concentra la producción empírica reciente?

PI2: ¿Qué efectos sobre procesos cognitivos, metacognitivos y afectivo-motivacionales documenta la literatura empírica sobre el uso de GenAI en aprendizaje?

PI3: ¿Cómo pueden interpretarse esos efectos desde los marcos teóricos de la neuroeducación y las ciencias cognitivas del aprendizaje?

PI4: ¿Qué vacíos empíricos y teóricos emergen de esa interpretación integrada, y qué líneas de investigación futura señalan?

1.4.2 Contribución del estudio

La contribución de esta revisión se articula en cuatro dimensiones diferenciadas, cada una identificable por contraste con lo que la literatura previa ha producido y con lo que aún no ha realizado.

La contribución teórica consiste en proponer la primera integración sistemática entre la evidencia empírica disponible sobre GenAI en aprendizaje y los marcos teóricos de la neuroeducación como lente interpretativa explícita. Las revisiones previas han documentado efectos usando vocabulario cognitivo sin anclar ese vocabulario en modelos de funcionamiento neurocognitivo; este estudio invierte esa relación, empleando los marcos neuroeducativos como principios organizadores de la síntesis y no como etiquetas descriptivas aplicadas post hoc a hallazgos conductuales.

La contribución analítica consiste en identificar patrones transversales que emergen del análisis integrado del corpus y que no son visibles en el examen de estudios individuales: en particular, la mediación afectivo-motivacional de los efectos cognitivos, el rol modulador de la alfabetización en IA del aprendiz y la dependencia del tipo de proceso cognitivo activado respecto a las decisiones de diseño conversacional de la herramienta.

La contribución metodológica consiste en la aplicación rigurosa del protocolo PRISMA 2020 a un campo caracterizado por terminología inestable, expansión acelerada de la producción y heterogeneidad de diseños primarios, con decisiones de elegibilidad explícitamente justificadas y evaluación de calidad mediante instrumentos diferenciados según el diseño de cada estudio.

La contribución prospectiva consiste en formular una agenda de investigación futura fundamentada en los vacíos identificados: específicamente, la necesidad de diseños longitudinales, de incorporación de medidas neurocognitivas objetivas, de mayor representación de educación básica y de contextos del Sur Global, y de marcos teóricos que integren evidencia neurocientífica con modelos pedagógicos de uso de GenAI.

2. Metodología

2.1 Diseño de la revisión y pregunta de investigación

El presente trabajo se diseñó como una revisión sistemática de la literatura con síntesis narrativa temática. Este tipo de revisión resulta apropiado cuando el campo de estudio se encuentra en expansión acelerada y presenta heterogeneidad sustancial en los diseños de los estudios primarios, lo que impide la agregación cuantitativa mediante meta-análisis. En el caso de la investigación sobre inteligencia artificial generativa en educación, la diversidad de diseños metodológicos —que incluye estudios no experimentales, cuasi-experimentales, de métodos mixtos y cualitativos—, la variabilidad de las variables de resultado examinadas y la disparidad de los contextos disciplinares y geográficos hacen inviable una síntesis estadística y justifican la síntesis narrativa como estrategia de integración. La revisión se diferencia de una revisión de alcance en que no se limita a mapear el campo sino que incorpora evaluación crítica de la calidad metodológica de los estudios y síntesis interpretativa de la evidencia. El diseño adoptado se adhiere al protocolo PRISMA 2020 en todas sus fases de identificación, cribado, elegibilidad e inclusión.

La pregunta de investigación que orienta el proceso fue estructurada en formato SPIDER, herramienta recomendada para revisiones sistemáticas en ciencias sociales y educativas donde el fenómeno de interés no se presta a la estructura PICO diseñada para ensayos clínicos. La Tabla 1 presenta los componentes de la pregunta y su definición para este estudio.

Tabla 1. Pregunta de investigación estructurada en formato SPIDER

Componente	Definición para este estudio
S — Sample	Estudiantes, docentes y aprendices en contextos educativos formales y no formales
P — Phenomenon of Interest	Uso de herramientas de inteligencia artificial generativa (LLMs, chatbots generativos, asistentes GenAI) en procesos de aprendizaje

I — Design	Estudios empíricos cuantitativos, cualitativos y de métodos mixtos
D — Evaluation	Efectos sobre procesos cognitivos, metacognitivos, afectivos y resultados de aprendizaje
R — Research type	Revisión sistemática con síntesis narrativa temática

Fuente: elaboración propia

De la estructura SPIDER se deriva la siguiente pregunta de investigación: ¿De qué manera el uso de herramientas de inteligencia artificial generativa afecta los procesos de aprendizaje en contextos educativos, y cómo pueden interpretarse sus efectos desde un enfoque neuroeducativo?

El protocolo de esta revisión no fue registrado previamente en PROSPERO ni en ningún registro equivalente, lo que constituye una limitación de transparencia prospectiva que debe reconocerse explícitamente. Para mitigar el riesgo asociado a esta ausencia se adoptaron cuatro medidas concretas antes del inicio del proceso de búsqueda y selección: primero, la definición a priori de todos los criterios de elegibilidad que se presentan en la sección 2.3; segundo, la documentación sistemática de cada decisión de inclusión y exclusión mediante una tabla de elegibilidad que registra el estudio, la decisión adoptada y su justificación; tercero, la aplicación del protocolo PRISMA 2020 como marco de reporte que garantiza la trazabilidad de cada fase del proceso; y cuarto, el registro de las ecuaciones de búsqueda exactas y los resultados cuantitativos por base de datos, que se presentan en la sección 2.2. Estas medidas no equivalen al registro prospectivo de un protocolo, pero proporcionan un nivel de transparencia documental compatible con los estándares de publicación de revistas indexadas.

2.2 Estrategia de búsqueda

La búsqueda sistemática se ejecutó en seis bases de datos seleccionadas por su cobertura complementaria en los campos relevantes para esta revisión. ScienceDirect y Frontiers fueron

seleccionadas por su amplia cobertura multidisciplinar en tecnología educativa, psicología del aprendizaje y ciencias cognitivas, y por su alta densidad de publicaciones recientes sobre inteligencia artificial aplicada a la educación. IEEE Xplore fue incluida por su cobertura específica en sistemas inteligentes, tecnología de la información y aplicaciones de aprendizaje automático, que captura una parte de la producción científica sobre herramientas GenAI que no aparece en bases de orientación social o educativa. PubMed fue incorporada por su indexación de psicología educativa, neurociencias cognitivas y ciencias del aprendizaje, campos directamente relevantes para el enfoque neuroeducativo de la revisión. Emerald fue seleccionada por su cobertura en gestión educativa y educación superior, donde se publica parte de la investigación sobre adopción tecnológica en contextos universitarios. SciELO fue incluida para maximizar la cobertura iberoamericana y reducir el sesgo anglocéntrico inherente a las bases anglófonas. Se reconoce como limitación de cobertura la no consulta de Scopus, Web of Science y ERIC, bases que podrían haber recuperado estudios adicionales relevantes, particularmente en el caso de ERIC para literatura especializada en educación.

Las ecuaciones de búsqueda se construyeron a partir de tres bloques conceptuales: inteligencia artificial generativa como bloque nuclear obligatorio, educación y aprendizaje como segundo bloque obligatorio, y neuroeducación y cognición como bloque complementario de precisión. Este principio de construcción refleja la decisión metodológica de maximizar la sensibilidad en los dos primeros bloques y utilizar el tercero como lente analítica en la síntesis, no como requisito de búsqueda, dado que exigirlo reduciría la recuperación de estudios empíricamente valiosos que operacionalizan constructos neuroeducativos sin emplear ese término. En ScienceDirect, la restricción operativa de un máximo de ocho conectores booleanos en la interfaz avanzada obligó a dividir la búsqueda en dos subconsultas paralelas ejecutadas de forma independiente y posteriormente integradas. Los resultados nulos obtenidos en SciELO y Emerald con las ecuaciones aplicadas limitan la representación iberoamericana y de gestión

educativa en el corpus final, lo que se reconoce como una restricción de la estrategia adoptada. La Tabla 2 presenta las ecuaciones completas, las subconsultas donde aplica, los resultados iniciales recuperados y los resultados tras la aplicación de filtros técnicos.

Tabla 2. Ecuaciones de búsqueda por base de datos con resultados

Base de datos	Ecuación principal	Subconsultas	Resultados iniciales	Resultados tras filtros
ScienceDirect	tak(("generative artificial intelligence" OR "generative AI" OR GenAI OR ChatGPT) AND (education OR learning OR teaching OR student))	Principal A + Principal B (paralelas)	176 + 128 = 304	97
IEEE Xplore	((("All Metadata": "artificial intelligence" OR "All Metadata": "generative AI" OR "All Metadata": chatbot* OR "All Metadata": "large language model*") AND ("All Metadata": education OR "All Metadata": learning OR "All Metadata": "higher education") AND ("All Metadata": emotion* OR "All Metadata": affect*) AND ("All Metadata": "professional identity"))	Única	7	7
SciELO	"artificial intelligence" AND education AND emotion AND university	Única	0	0

Emerald	"artificial intelligence" AND education AND emotion AND identity AND university	Única	0	0
PubMed	"artificial intelligence" AND education AND emotion AND university	Única	882	21
Frontiers	("artificial intelligence" OR "generative AI" OR chatbot) AND (pedagogy OR learning OR education) AND (emotion OR affect OR anxiety) AND ("professional identity" OR "career identity") AND ("higher education" OR university)	Única	9.313	5.587

Fuente: elaboración propia

La cobertura temporal de la búsqueda abarcó el periodo 2020–2026. La fecha de inicio en 2020 responde a la necesidad de capturar estudios precursores sobre inteligencia artificial conversacional aplicada a educación que anteceden a la consolidación del término "inteligencia artificial generativa" sin ser conceptualmente distintos del fenómeno que esta revisión examina. La delimitación superior en 2026 refleja el año de ejecución de la búsqueda, que incorporó publicaciones disponibles hasta el momento de su realización. Se reconoce que el punto de inflexión empírico del campo ocurre a partir de 2022 con la expansión masiva de LLMs y el lanzamiento de ChatGPT, y que la literatura anterior a esa fecha es cuantitativamente escasa y conceptualmente más heterogénea respecto al fenómeno central de esta revisión.

Los filtros técnicos aplicados fueron los siguientes. En cuanto al idioma, se recuperaron documentos en inglés y español; el portugués fue contemplado en el protocolo como tercer idioma de búsqueda, particularmente para SciELO, pero no produjo resultados recuperables

con las ecuaciones aplicadas. En cuanto al tipo documental, se priorizaron artículos de investigación empírica revisados por pares como tipo principal; los conference papers de IEEE fueron contemplados como excepción justificada dado que parte de la innovación en sistemas inteligentes se publica inicialmente en actas de conferencia, aunque ningún estudio de ese tipo superó el cribado final. En cuanto al acceso, se aplicó el filtro de acceso abierto como criterio operativo de recuperación para garantizar la disponibilidad del texto completo durante el proceso de selección y extracción; este filtro no operó como criterio de elegibilidad conceptual, de modo que estudios sin acceso abierto que hubieran cumplido todos los criterios de inclusión habrían sido incorporados al corpus si se hubiera podido acceder a su texto completo mediante otras vías.

2.3 Proceso de selección de estudios

Los criterios de inclusión y exclusión fueron definidos a priori, antes del inicio del cribado, organizados en siete dimensiones que cubren la totalidad de los aspectos relevantes para la elegibilidad de los estudios. Una decisión metodológica central en la definición de estos criterios fue no incluir el término "neuroeducación" ni constructos neuroeducativos específicos como requisito obligatorio de inclusión. Esta decisión responde al reconocimiento de que los estudios empíricamente valiosos sobre efectos de GenAI en procesos cognitivos y de aprendizaje frecuentemente operacionalizan constructos neuroeducativos —carga cognitiva, metacognición, funciones ejecutivas, engagement— sin emplear ese marco terminológico de forma explícita. Exigirlo como criterio duro habría reducido artificialmente la sensibilidad de la búsqueda y excluido estudios relevantes para los objetivos de la revisión. El enfoque neuroeducativo opera en esta revisión como lente analítica de síntesis e interpretación, no como filtro de selección. La Tabla 3 presenta los criterios de inclusión y exclusión con su justificación.

Tabla 3. Criterios de inclusión y exclusión alineados con PRISMA

Criterio	Inclusión	Exclusión	Justificación
Tema	Estudios que examinen el uso de herramientas de inteligencia artificial generativa (LLMs, chatbots generativos, asistentes GenAI) en procesos de aprendizaje o enseñanza-aprendizaje	Estudios sobre IA no generativa (sistemas expertos, analytics, IA predictiva sin generación de contenido); estudios de IA generativa en contextos clínicos, industriales o de ingeniería sin componente educativo analizable	Mantiene la consistencia conceptual entre el objeto tecnológico de la revisión (GenAI) y el contexto de aplicación (educación), evitando la recuperación de literatura de IA sin relevancia pedagógica
Diseño	Estudios empíricos cuantitativos, cualitativos y de métodos mixtos; evaluaciones de intervención; diseños experimentales, cuasi-experimentales y correlacionales con datos primarios	Editoriales, artículos de opinión, cartas al editor, notas breves sin datos empíricos, comentarios, revisiones narrativas no sistemáticas	Garantiza que el corpus esté constituido por evidencia primaria evaluable, lo que eleva el rigor y la comparabilidad de la síntesis
Contexto	Educación formal (escolar, secundaria, superior, formación docente); educación no formal con proceso de aprendizaje analizable; formación profesional con componente pedagógico explícito	Entrenamiento corporativo puramente técnico sin proceso de aprendizaje analizable; usos de GenAI sin relación con enseñanza o aprendizaje humano	Delimita el campo de aplicación a contextos donde el proceso de aprendizaje es el objeto central de análisis, excluyendo usos instrumentales de GenAI sin dimensión pedagógica
Tecnología	LLMs, chatbots generativos, sistemas GenAI multimodales, asistentes generativos aplicados al aprendizaje, herramientas de generación de contenido educativo (ChatGPT, GPT-4, Gemini, Claude, Yuanbao, DeepSeek y equivalentes)	Sistemas expertos de primera generación, plataformas de analytics sin generación de contenido, sistemas de tutoría inteligente basados en reglas sin componente generativo, herramientas de IA predictiva sin interacción generativa	Evita la confusión entre inteligencia artificial educativa clásica e inteligencia artificial generativa, que constituye el objeto tecnológico específico de esta revisión
Población	Estudiantes de cualquier nivel educativo, docentes en formación o en ejercicio, aprendices en contextos formativos formales o no formales	Poblaciones exclusivamente clínicas, industriales o corporativas sin rol educativo identificable	Mantiene la alineación de la revisión con procesos de aprendizaje humano en contextos formativos, independientemente del nivel o modalidad
Variables	Estudios que reporten usos, percepciones, procesos cognitivos, variables afectivo-motivacionales o resultados de aprendizaje asociados al uso de GenAI	Estudios centrados exclusivamente en aspectos éticos, regulatorios o de política educativa sin datos empíricos sobre procesos o resultados de aprendizaje	La revisión está orientada a identificar efectos e interpretarlos desde un enfoque neuroeducativo; los estudios sin datos sobre procesos de aprendizaje no contribuyen a ese objetivo
Idioma	Inglés y español	Publicaciones en otros idiomas sin traducción verificable al inglés o al español	Equilibra la exhaustividad de la cobertura con la viabilidad del proceso de lectura y extracción; el portugués fue contemplado

			en el protocolo pero no produjo resultados recuperables con las ecuaciones aplicadas
Periodo temporal	Publicaciones entre 2020 y 2026	Publicaciones anteriores a 2020	El punto de inflexión empírico del campo ocurre con la expansión de LLMs a partir de 2022; la inclusión desde 2020 captura estudios precursores sobre IA conversacional aplicada a educación conceptualmente continuos con el fenómeno GenAI
Calidad básica	Artículos con texto completo accesible, metadatos suficientes para la extracción y proceso de revisión por pares verificable	Resúmenes sin texto completo disponible, duplicados, versiones preliminares redundantes de un artículo ya incluido en su versión final	Necesario para garantizar la evaluación crítica completa de cada estudio; la ausencia de texto completo impide la extracción de datos y la evaluación de calidad metodológica

Fuente: elaboración propia

El proceso de selección se desarrolló en dos fases secuenciales. Antes del inicio del cribado se ejecutó la deduplicación de registros entre bases de datos mediante la comparación de títulos, autores y años de publicación, con el fin de eliminar estudios recuperados en más de una base. La primera fase consistió en el cribado por título y resumen de los registros no duplicados que superaron los filtros técnicos descritos en 2.2. La segunda fase consistió en la lectura completa de los estudios que superaron el cribado inicial. Los datos cuantitativos de cada fase por base de datos fueron los siguientes: ScienceDirect presentó 304 registros revisados en cribado inicial, de los cuales 16 avanzaron a lectura completa y 1 fue incluido en el corpus final; IEEE Xplore presentó 7 registros en cribado inicial, 4 avanzaron a lectura completa y ninguno fue incluido; PubMed presentó 21 registros en cribado inicial, ninguno avanzó a lectura completa; Frontiers presentó 5.587 registros en cribado inicial, 39 avanzaron a lectura completa y 27 fueron incluidos en el corpus final. El corpus resultante comprende 28 estudios. El diagrama de flujo completo del proceso se presenta en la Figura 1, conforme al protocolo PRISMA 2020. El proceso de selección fue realizado por un único revisor, lo que constituye una limitación metodológica que debe reconocerse explícitamente, dado que la participación de un segundo

revisor independiente con resolución de discrepancias por consenso es el estándar recomendado por PRISMA para minimizar el sesgo de selección. Ante los casos ambiguos durante el cribado por título y resumen se aplicó el principio de conservadurismo: cuando la información disponible no permitía determinar con certeza el incumplimiento de algún criterio de inclusión, el estudio avanzaba a la fase de lectura completa. Durante la lectura completa, las decisiones de exclusión se fundamentaron en la verificación sistemática de cada criterio de elegibilidad predefinido, documentando la razón de exclusión en la tabla de elegibilidad. Este procedimiento, aunque no elimina el riesgo de sesgo de selección asociado al revisor único, proporciona un nivel de consistencia decisional verificable.

2.4 Extracción de datos

La extracción de datos se organizó en cuatro categorías de variables, cada una respondiendo a un componente específico del análisis posterior. La primera categoría comprende las variables de identificación del estudio: autor y año de publicación, país de procedencia, revista de publicación y año, correspondientes a las columnas 1 a 4 de la matriz de extracción. La segunda categoría comprende las características metodológicas: diseño del estudio, metodología empleada, tamaño y tipo de muestra, tipo de herramienta de inteligencia artificial utilizada y contexto educativo, correspondientes a las columnas 5 a 9. Esta categoría alimenta la caracterización del corpus presentada en la sección 3.2. La tercera categoría comprende el contenido sustantivo de cada estudio: variables estudiadas, instrumentos de medición y resultados principales, correspondientes a las columnas 10 a 12, y constituye la base de la síntesis de evidencia desarrollada en la sección 3.3. La cuarta categoría comprende los indicadores de transparencia metodológica: evaluación metodológica global, periodo de realización del estudio, disponibilidad de datos o código, y existencia de registro o protocolo previo, correspondientes a las columnas 13 a 16, y alimenta la evaluación de calidad descrita en la sección 2.5.

El instrumento de extracción consistió en una matriz estructurada con las dieciséis variables descritas, aplicada de forma sistemática a los 28 estudios del corpus. La extracción fue realizada por un único revisor, lo que constituye una limitación adicional del proceso. Un hallazgo relevante durante la extracción fue la alta frecuencia de información no reportada en los estudios primarios, registrada como NR en la matriz: el periodo de realización del estudio, la disponibilidad de datos o código y la existencia de registro o protocolo previo fueron las variables con mayor proporción de ausencias. Esta opacidad metodológica es en sí misma un dato sobre el estado de madurez del campo y limita la evaluación completa de algunos indicadores de calidad en la sección 2.5.

2.5 Evaluación de la calidad metodológica

La evaluación de la calidad metodológica se realizó mediante cuatro herramientas diferenciadas, seleccionadas en función del diseño de cada estudio incluido. Esta decisión responde al reconocimiento de que una herramienta única de evaluación no es apropiada para un corpus metodológicamente heterogéneo: las dimensiones relevantes de riesgo de sesgo difieren sustancialmente entre un ensayo aleatorizado, un estudio correlacional transversal, un diseño de métodos mixtos y un estudio cualitativo exploratorio. Se aplicó ROBINS-I a los estudios no experimentales y cuasi-experimentales sin asignación aleatoria, que constituyen la mayoría del corpus (n=21). Se aplicó RoB 2 a los dos ensayos con asignación aleatoria: Henze et al. (2026) y Xu et al. (2026). Se aplicó MMAT a los cuatro estudios de métodos mixtos: Li et al. (2025), Salhab y Aboushi (2026), Wang et al. (2026) y Wang y Liu (2026). Se aplicó la herramienta JBI para estudios cualitativos al único estudio de diseño cualitativo del corpus: Westbye et al. (2026).

La evaluación se estructuró en torno a cinco dominios comunes a todas las herramientas aplicadas: riesgo de sesgo en la selección de participantes, en la medición de variables, derivado de factores de confusión no controlados, en el análisis de datos y en el reporte de resultados.

Para cada dominio se emitió un juicio de nivel bajo, moderado o alto, y a partir de la combinación de juicios por dominio se asignó un riesgo global a cada estudio. La evaluación fue realizada por un único revisor, lo que constituye una limitación reconocida del proceso. El resultado sintético del procedimiento fue un riesgo global moderado en los 28 estudios, con sesgo de reporte sistemáticamente bajo en la totalidad del corpus y confusión no controlada como el dominio más problemático de forma universal. Estos resultados se presentan en detalle en la Tabla 2 de la sección de resultados y son retomados en la sección 2.6 para su traducción en calificadores epistémicos durante la síntesis.

2.6 Estrategia de síntesis

La síntesis narrativa temática fue seleccionada como estrategia de integración de la evidencia ante tres características del corpus que hacen inviable la agregación cuantitativa. En primer lugar, la heterogeneidad de diseños primarios, que incluye estudios no experimentales transversales, cuasi-experimentales, de métodos mixtos y un estudio cualitativo, impide la estandarización de medidas de efecto. En segundo lugar, la diversidad de variables de resultado —que abarca motivación, carga cognitiva, metacognición, rendimiento académico, adopción tecnológica, bienestar e inteligencia emocional, entre otras— hace imposible la comparación cuantitativa directa entre estudios. En tercer lugar, la diversidad de contextos disciplinares y geográficos del corpus introduce una heterogeneidad contextual que refuerza la pertinencia de una síntesis interpretativa sobre una agregación estadística. La síntesis narrativa temática se diferencia de la descripción acumulativa de estudios individuales en que su objetivo es identificar patrones, convergencias y divergencias que emergen del análisis del conjunto y que no son visibles en el examen aislado de cada estudio.

El procedimiento analítico de síntesis se ejecutó en tres operaciones secuenciales. La primera operación consistió en la clasificación temática inductiva de los 28 estudios en dominios y líneas de investigación, a partir de las columnas "Dominio" y "Enfoque cognitivo" de la Tabla

3 de síntesis narrativa y las columnas "Diseño" y "Variables" de la Tabla 1 de características de los estudios. Esta clasificación fue inductiva en el sentido de que los dominios y líneas emergieron del análisis del corpus y no fueron impuestos desde una taxonomía preexistente. La segunda operación consistió en la identificación de convergencias y divergencias mediante comparación sistemática de la columna "Resultado" de la Tabla 3 entre estudios del mismo dominio y entre dominios, cotejada con la columna "Riesgo global" de la Tabla 2 para ponderar el peso relativo de cada hallazgo. La tercera operación consistió en la inferencia inductiva de patrones explicativos transversales mediante el análisis integrado de las tres tablas, orientada por el marco neuroeducativo como lente interpretativa; este proceso fue inductivo y emergente, construido a partir de las regularidades identificadas en el corpus, y no deductivo desde categorías teóricas predefinidas.

Los juicios de riesgo de sesgo producidos en la sección 2.5 se traducen en la síntesis mediante un sistema de calificadores epistémicos diferenciados según el nivel de evidencia que respalda cada afirmación. Las afirmaciones sustentadas en estudios con asignación aleatoria se formulan con mayor fuerza asertiva, empleando construcciones como "la evidencia indica" o "el estudio demuestra". Las afirmaciones sustentadas en diseños correlacionales o cuasi-experimentales sin aleatorización se formulan con calificadores de cautela, empleando construcciones como "los resultados sugieren", "los datos son consistentes con" o "se observa una asociación entre". Las afirmaciones sustentadas en un único estudio, independientemente de su diseño, se identifican explícitamente como hallazgos aislados que requieren replicación antes de ser generalizados. Este sistema de calificadores opera de forma sistemática a lo largo de la sección de resultados y permite al lector calibrar la fuerza inferencial de cada afirmación sin necesidad de consultar las tablas de calidad en cada caso.

El método de síntesis presenta tres limitaciones específicas que deben reconocerse. La primera es que la síntesis narrativa introduce mayor subjetividad interpretativa que el meta-análisis,

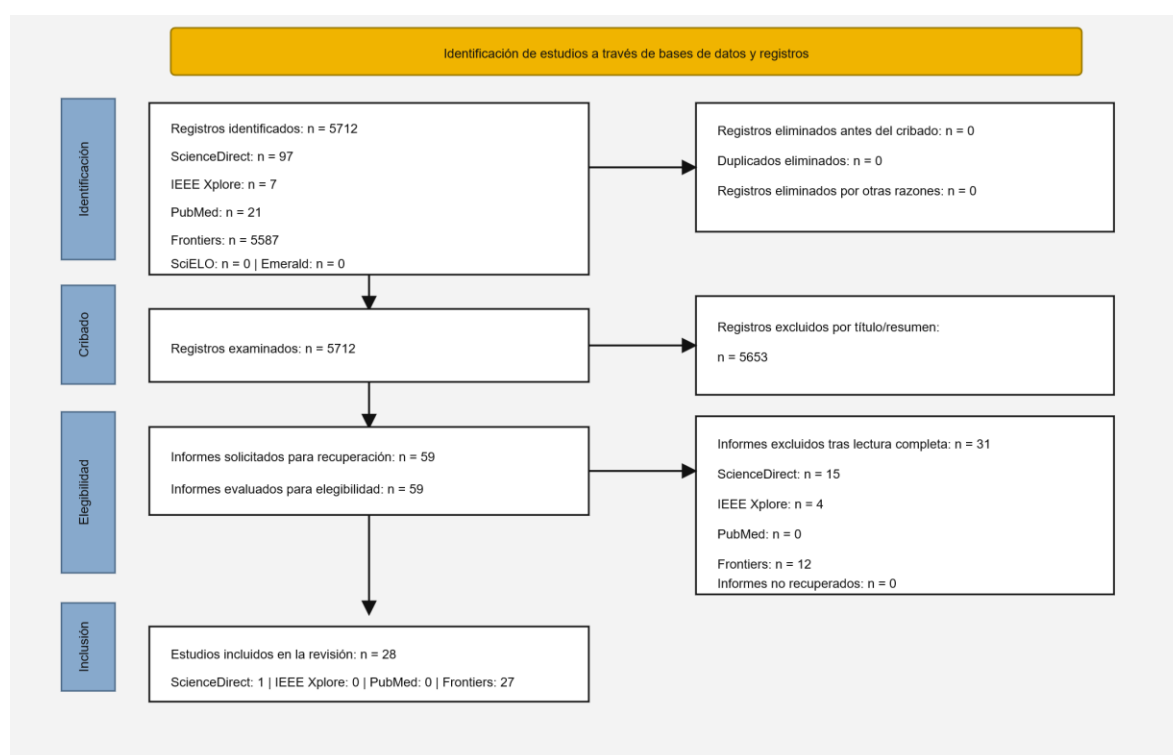
particularmente en la tercera operación de inferencia de patrones transversales, que no fue verificada por un segundo analista independiente. La segunda es que la concentración del corpus en Frontiers como fuente de 27 de los 28 estudios introduce un sesgo de publicación específico de esa editorial que puede haber limitado la diversidad metodológica y temática del corpus final, dado que las políticas editoriales y las áreas de énfasis de una sola editorial condicionan el tipo de investigación que se indexa y se aprueba. La tercera es que los resultados nulos obtenidos en SciELO y Emerald limitan la representación de investigación iberoamericana y de gestión educativa en la síntesis, lo que debe considerarse al evaluar la aplicabilidad de los hallazgos en contextos latinoamericanos e institucionales.

3. Resultados

3.1 Resultados de la búsqueda y selección del corpus

La búsqueda sistemática se ejecutó en seis bases de datos: ScienceDirect, IEEE Xplore, SciELO, Emerald, PubMed y Frontiers, con un rango temporal de 2020 a 2026. El proceso de identificación arrojó un total de 10.514 registros iniciales, distribuidos de forma heterogénea entre las bases consultadas. Tras la eliminación de duplicados y la aplicación de filtros documentales y de acceso, se sometieron a cribado por título y resumen 5.714 registros. De estos, 59 fueron evaluados en texto completo, y 28 cumplieron la totalidad de los criterios de inclusión establecidos en el protocolo. Las principales razones de exclusión en la fase de lectura completa fueron: ausencia de datos empíricos primarios, uso de sistemas de inteligencia artificial no generativa, y estudios centrados en contextos clínicos o industriales sin componente educativo analizable. El diagrama de flujo PRISMA se presenta en la Figura 1.

Figura 1 — Diagrama de flujo PRISMA



3.2 Estructura del campo: caracterización temática y conceptual

3.2.1 Distribución temática del corpus

El análisis de los 28 estudios incluidos permitió identificar cinco dominios temáticos que organizan la producción científica sobre el uso de inteligencia artificial generativa en procesos de aprendizaje: pedagógico/adopción, metacognitivo, socioemocional, cognitivo y neuroeducativo. Su distribución en el corpus no es uniforme, lo que refleja las prioridades investigativas predominantes en el campo durante el período analizado.

El dominio pedagógico/adopción concentra el mayor volumen de estudios ($n=10$), seguido por el metacognitivo ($n=7$), el socioemocional ($n=5$), el cognitivo ($n=3$) y el neuroeducativo ($n=3$). Esta distribución revela una paradoja que atraviesa estructuralmente el campo: el dominio neuroeducativo, que constituye el marco interpretativo declarado de un número creciente de investigaciones sobre GenAI y aprendizaje, es simultáneamente el menos representado en términos de producción empírica. Dicho desequilibrio no es un dato menor: anticipa la tensión

teórica que se desarrolla en la sección 3.2.3 y condiciona el nivel de confianza de la evidencia disponible. Véase el anexo 1.

3.2.2 Líneas de investigación predominantes

Dentro de los dominios identificados, el corpus se organiza en cuatro líneas de investigación diferenciadas por su enfoque compartido, marco teórico y tipo de pregunta que articulan.

La primera y más voluminosa es la **línea de modelos de adopción tecnológica**, conformada por estudios que aplican marcos TAM, UTAUT o UTAUT2 para explicar la intención de uso y la aceptación de herramientas GenAI en estudiantes y docentes. Su fortaleza reside en la replicabilidad del marco teórico a través de contextos geográficos y disciplinares diversos: los hallazgos de BinJwair (2025) en Arabia Saudita, Cao et al. (2025) en China, He y Ren (2025) en formación docente musical, Kinskofer y Tulis (2025) en Austria, Yu et al. (2024) en Corea del Sur, Wang y Liu (2026) en lenguas extranjeras, Sergeeva et al. (2025) en Rusia y Susanto et al. (2026) en Brunéi comparten una misma arquitectura analítica. Su límite crítico es igualmente compartido: esta línea informa sobre los factores que predicen el uso de GenAI, pero no sobre los procesos cognitivos que ocurren durante ese uso ni sobre la calidad del aprendizaje resultante.

La segunda línea es la de **intervención con medición de resultados**, que agrupa los únicos estudios del corpus con capacidad de inferencia sobre los efectos del uso de GenAI. Incluye los trabajos de George-Reyes et al. (2024), Henze et al. (2025), Henze et al. (2026), Li et al. (2025), Salhab y Aboushi (2026) y Xu et al. (2026). Su heterogeneidad disciplinar — emprendimiento, formación docente en física, educación en danza, formación vocal musical y microaprendizaje en línea— aporta amplitud al campo, aunque dificulta la comparabilidad directa entre estudios y limita las generalizaciones.

Las dos **líneas emergentes** presentan menor volumen pero mayor novedad teórica. La línea de efectos socioemocionales y bienestar —representada por Kazazoglu y Turun Özel (2025),

Mohamed et al. (2026), Pan y Li (2025), Wang y Liu (2026) y Susanto et al. (2026)— desplaza el foco desde la adopción tecnológica hacia las consecuencias afectivas del uso de GenAI, incluyendo ansiedad, bienestar del aprendiz e inteligencia emocional. La línea de alfabetización en IA como proceso cognitivo mediador —representada por Wang et al. (2026), Salhab y Aboushi (2025), Sergeeva et al. (2025) y Hu et al. (2026)— introduce una perspectiva que trata la competencia crítica del aprendiz como variable central, no como supuesto de partida. El desplazamiento del campo hacia estas dos líneas emergentes indica la dirección en que se está reorientando la investigación en el período más reciente del corpus.

3.2.3 Marcos conceptuales dominantes y paradoja teórica central

La arquitectura teórica del corpus está dominada por modelos de aceptación tecnológica, con presencia secundaria de teorías de carga cognitiva, autorregulación del aprendizaje y metacognición. Los marcos neurocientíficos formales —incluyendo teorías de memoria de trabajo, neuroplasticidad o atención sostenida fundamentadas en evidencia neurobiológica— están prácticamente ausentes como base explícita de los diseños de investigación. Los estudios que operacionalizan constructos más próximos al enfoque neuroeducativo —Liew et al. (2025) con carga cognitiva, Li et al. (2025) con metacognición, Klarin et al. (2024) con funciones ejecutivas, Salhab y Aboushi (2026) con engagement cognitivo multidimensional y Henze et al. (2026) con motivación y carga cognitiva percibida— lo hacen mediante medidas conductuales y perceptuales autoinformadas, sin anclaje en evidencia neurocientífica de base. Esta situación configura lo que puede denominarse la paradoja teórica central del campo: el vocabulario neuroeducativo y cognitivo es ampliamente utilizado para describir los efectos de GenAI en el aprendizaje, pero los diseños de investigación empleados para medirlos raramente permiten fundamentar esas afirmaciones en mecanismos neurocognitivos verificables. Esta brecha no es atribuible a estudios individuales con limitaciones específicas, sino que constituye

una característica estructural del campo en su estado actual de desarrollo. Su identificación representa una contribución analítica de esta revisión y opera como condicionante de los patrones explicativos identificados en la sección 3.3.3.

3.2.4 Vacíos temáticos y poblacionales

El examen del corpus permite delimitar cinco vacíos estructurales que marcan las fronteras actuales del conocimiento en este campo, entendidos como ausencias en *qué* se ha investigado, con independencia de la calidad metodológica de lo disponible.

El primero es el **vacío longitudinal**: el corpus es casi enteramente transversal o corresponde a intervenciones de duración inferior a un semestre, lo que impide determinar si los efectos observados sobre motivación, metacognición o engagement se sostienen, se intensifican o se diluyen con el tiempo. El segundo es el **vacío neurocognitivo**: ningún estudio incluido emplea medidas objetivas de procesamiento cognitivo —neuroimagen funcional, eye-tracking, tiempos de reacción o registros electrofisiológicos—, lo que mantiene la evidencia en el plano perceptual y conductual. El tercero es el **vacío poblacional**: los estudios de educación básica (Jauhiainen y Garagorry Guerra, 2024; Westbye et al., 2026) y con adolescentes (Klarin et al., 2024) constituyen excepciones minoritarias en un corpus dominado por educación superior, lo que limita la transferibilidad de los hallazgos a etapas formativas tempranas. El cuarto es el **vacío geográfico**: con la excepción de Jauhiainen y Garagorry Guerra (2024) en Uruguay y Salhab y Aboushi (2025, 2026) en Palestina, el corpus está concentrado en China, Europa central y Oriente Medio, con ausencia casi total del Sur Global, África y América Latina. El quinto es el **vacío de diseño**: solo tres estudios emplean asignación aleatoria —Henze et al. (2026), Li et al. (2025) y Xu et al. (2026)—, y en todos ellos el tamaño muestral es reducido, lo que limita el poder estadístico y la capacidad de establecer relaciones causales.

3.3 Síntesis de la evidencia: hallazgos, consensos y límites

3.3.1 Convergencias: hallazgos con mayor consistencia en el corpus

El análisis del conjunto de estudios permite identificar tres hallazgos que muestran consistencia suficiente para ser considerados convergencias del campo, con las reservas metodológicas que se explicitan en cada caso.

El hallazgo más replicado es el **efecto positivo de GenAI sobre la motivación y el engagement del aprendiz**. Este resultado aparece en estudios de intervención, de adopción y de efectos socioemocionales, a través de contextos geográficos y disciplinares heterogéneos: Dogaru et al. (2025), Xu et al. (2026), Henze et al. (2026), Salhab y Aboushi (2026), Jauhiainen y Garagorry Guerra (2024), Huang et al. (2025), Susanto et al. (2026) y Pan y Li (2025) reportan mejoras en estas dimensiones. Su consistencia es notable, pero debe leerse con cautela: descansa mayoritariamente en diseños correlacionales y medidas autoinformadas, lo que limita la fuerza inferencial del hallazgo y no permite descartar efectos de novedad tecnológica como explicación alternativa.

La segunda convergencia es la **adopción de GenAI mediada por utilidad percibida e influencia social**, identificada de forma consistente en la línea de modelos de adopción tecnológica. BinJwair (2025), Cao et al. (2025), Kinskofer y Tulis (2025), Yu et al. (2024), He y Ren (2025) y Wang y Liu (2026) coinciden en señalar la expectativa de desempeño y la utilidad percibida como los predictores más robustos de la intención de uso, con variaciones en el peso de la influencia social según el contexto cultural. Este consenso es metodológicamente el más sólido del corpus por su replicabilidad, aunque su alcance es estrictamente predictivo: informa sobre qué determina el uso de GenAI, no sobre qué ocurre cognitivamente durante ese uso.

La tercera convergencia es la **personalización dinámica como atributo diferencial de GenAI respecto a tecnologías educativas previas**. Jauhiainen y Garagorry Guerra (2024), Li et al. (2025), Westbye et al. (2026), Hu et al. (2026) y Salhab y Aboushi (2026) evidencian que la capacidad de GenAI para ajustar materiales, retroalimentación e interacción al perfil individual

del aprendiz constituye su valor agregado más distintivo frente a sistemas de IA educativa anteriores. Esta convergencia tiene implicaciones directas para el diseño instruccional, aunque la evidencia sobre sus efectos sostenidos en el aprendizaje profundo sigue siendo limitada. Véase el anexo 2

3.3.2 Divergencias: resultados contradictorios y tensiones no resueltas

El corpus presenta tres divergencias sustantivas que no pueden atribuirse a diferencias metodológicas menores, sino que revelan la sensibilidad del fenómeno al contexto, al perfil del aprendiz y al diseño de la herramienta.

La divergencia más relevante para la toma de decisiones pedagógicas es la **inconsistencia sistemática en los efectos sobre el rendimiento académico objetivo**. Los estudios que incorporan medidas de rendimiento no autoinformado muestran resultados que contrastan con la consistencia de los hallazgos motivacionales: Henze et al. (2026) no encuentra diferencias cognitivas significativas entre el grupo que usó un chatbot GPT y el grupo que utilizó Excel para análisis de datos; Li et al. (2025) reporta mejoras en metacognición pero sin correlación con el rendimiento vocal; Klarin et al. (2024) no identifica asociaciones significativas entre el uso de GenAI y el rendimiento académico en adolescentes. George-Reyes et al. (2024) constituye la excepción con mejoras significativas en habilidades de emprendimiento y pensamiento complejo, aunque en un diseño cuasi-experimental sin aleatorización. La disociación entre efectos afectivo-motivacionales positivos y efectos cognitivos objetivos no significativos es el dato más crítico del corpus para cualquier política educativa basada en evidencia.

La segunda divergencia concierne a los **efectos de GenAI sobre la carga cognitiva**, que el corpus muestra como bidireccionales en función del diseño conversacional de la herramienta. Liew et al. (2025) evidencia que el uso de lenguaje juvenil en un chatbot educativo aumenta el engagement y el arousal positivo, pero también eleva la dificultad percibida y la carga cognitiva

extrínseca. En contraste, Salhab y Aboushi (2026) reporta reducción de la carga cognitiva y la frustración en un entorno de microaprendizaje asistido por IA. Estos resultados no son contradictorios sino complementarios: indican que la carga cognitiva no es una consecuencia automática del uso de GenAI, sino una variable sensible a las decisiones de diseño instruccional, particularmente al nivel de andamiaje y al registro comunicativo del sistema.

La tercera divergencia afecta a los **efectos socioemocionales del uso de GenAI**, donde el corpus presenta hallazgos opuestos sin que los estudios disponibles permitan identificar con precisión qué variables moderan la dirección del efecto. Kazazoglu y Turun Özel (2025) identifica una asociación negativa entre el uso frecuente de ChatGPT y la inteligencia emocional en estudiantes de formación docente en escritura EFL, con una leve mejora en adaptabilidad interpretada con cautela por los propios autores. Huang et al. (2025), en contraste, reporta que la usabilidad y la presencia de chatbots generativos predijeron positivamente la mentalidad de crecimiento en lengua extranjera, con la inteligencia emocional como mediadora. La ausencia de replicación del hallazgo de Kazazoglu y Turun Özel impide determinar si refleja un patrón genuino o un resultado específico del contexto de intervención.

3.3.3 Patrones explicativos transversales

Más allá de las convergencias y divergencias descritas, el análisis integrado del corpus permite identificar tres patrones explicativos que emergen del conjunto y que no son visibles en el examen de estudios individuales. Estos patrones constituyen el principal aporte analítico de esta revisión y se articulan directamente con la paradoja teórica identificada en la sección 3.2.3. El primer patrón es que **los efectos de GenAI sobre el aprendizaje parecen estar mediados por variables afectivo-motivacionales antes que por mecanismos cognitivos directos**. El corpus muestra de forma consistente que GenAI afecta en primer lugar el estado motivacional y afectivo del aprendiz —engagement, motivación intrínseca, bienestar, reducción de ansiedad—, y que los efectos sobre variables cognitivas, cuando aparecen, son secundarios a

esa mediación o no alcanzan significación estadística cuando se miden de forma objetiva. Este patrón es coherente con los hallazgos de Henze et al. (2026), donde el grupo que usó el chatbot mostró mayor motivación y disfrute sin superioridad cognitiva demostrable; con Xu et al. (2026), donde el grupo GenAI superó al grupo control en engagement y motivación, con mejoras en habilidades interpretadas en ese contexto; con Salhab y Aboushi (2026), donde las dimensiones emocional y conductual del engagement fueron las más favorecidas; y con Huang et al. (2025) y Pan y Li (2025), que documentan la mediación afectiva como mecanismo explicativo central. Este patrón conecta directamente con la paradoja de A.3: el campo emplea vocabulario cognitivo para describir efectos que sus propios datos sugieren que son primariamente afectivos.

El segundo patrón es que **la alfabetización en IA del aprendiz opera como moderadora del beneficio, de forma que el acceso tecnológico por sí solo no predice mejoras en aprendizaje**. Los estudios que incorporan la alfabetización en IA como variable muestran que los beneficios del uso de GenAI están condicionados por la capacidad del aprendiz para interactuar críticamente con la herramienta, evaluar sus respuestas y regular su propio proceso de uso. Wang et al. (2026) propone un modelo dinámico de desarrollo de alfabetización mediante ciclos de cognición-práctica-evaluación; Salhab y Aboushi (2025) evidencia que la alfabetización en IA y las habilidades del siglo XXI predicen la aceptación y el aprovechamiento de GenAI; Hu et al. (2026) muestra que las herramientas cognitivas median la relación entre apoyo docente y aprendizaje matemático online; Klarin et al. (2024) aporta evidencia de que los adolescentes con mayores dificultades en funciones ejecutivas perciben la IA como más útil, pero sin que ello se traduzca en mejoras de rendimiento. La implicación de este patrón es que las intervenciones con GenAI que no contemplan el desarrollo simultáneo de alfabetización en IA tienen menor probabilidad de producir efectos cognitivos sostenibles.

El tercer patrón es que **el diseño conversacional de la herramienta condiciona el tipo de proceso cognitivo activado, con mayor peso que la mera presencia de IA generativa**. Los estudios que manipulan o describen con detalle las características de la interacción con el chatbot muestran que las decisiones de diseño —nivel de restricción de respuestas, registro lingüístico, tipo y granularidad de la retroalimentación, grado de andamiaje— determinan si el aprendiz activa procesos de elaboración propia, delega el procesamiento en la herramienta o experimenta sobrecarga extrínseca. Westbye et al. (2026) documenta que un chatbot restringido a formular preguntas —sin generar texto por el estudiante— favorece la elaboración descriptiva y preserva la agencia del aprendiz en la fase de preescritura, aunque con apropiación diferencial según el perfil del estudiante. Liew et al. (2025) muestra que modificaciones en el registro comunicativo del chatbot alteran simultáneamente el arousal, la emoción positiva y la carga cognitiva extrínseca. Li et al. (2025) evidencia que la retroalimentación generativa iterativa mejora la metacognición independientemente de su efecto sobre el rendimiento. Este patrón es el más directamente relevante para el diseño instruccional y, simultáneamente, el menos explorado en términos de investigación sistemática. Véase el anexo 3.

3.3.4 Nivel de confianza global de la evidencia

La evaluación de la calidad metodológica mediante las herramientas ROBINS-I, RoB 2, MMAT y JBI, aplicadas según el diseño de cada estudio, arroja un riesgo global moderado y homogéneo en la totalidad del corpus. Esta homogeneidad no refleja solidez uniforme, sino la presencia de vulnerabilidades estructurales compartidas que deben considerarse al interpretar cualquier hallazgo individual o patrón transversal.

Las dos vulnerabilidades más extendidas son el autoinforme como método de medición predominante —presente en 26 de los 28 estudios— y el diseño transversal como limitación a la inferencia causal, que impide descartar explicaciones alternativas a las relaciones identificadas. El sesgo de reporte es sistemáticamente bajo en todo el corpus, lo que sugiere

transparencia en la publicación, aunque este dato no compensa las limitaciones de diseño señaladas. Los únicos tres estudios con asignación aleatoria —Henze et al. (2026), Li et al. (2025) y Xu et al. (2026)— alcanzan un nivel de evidencia controlada, pero presentan tamaños muestrales reducidos y contextos disciplinares muy específicos que limitan su poder de generalización.

A partir de este análisis, la posición epistemológica del presente trabajo es que la evidencia disponible es **indicativa pero no confirmatoria**. Es suficiente para identificar tendencias, orientar decisiones pedagógicas con cautela informada y formular hipótesis para investigación futura, pero insuficiente para establecer relaciones causales robustas entre el uso de GenAI y resultados de aprendizaje verificables. Adicionalmente, la ausencia de estudios con resultados predominantemente negativos en el corpus sugiere la posibilidad de un sesgo de publicación favorable a intervenciones tecnológicas innovadoras, cuyo efecto sobre la estimación real del impacto de GenAI en el aprendizaje no puede cuantificarse con la información disponible.

3.4 Nota interpretativa: el enfoque neuroeducativo como lente analítica

El enfoque neuroeducativo que articula esta revisión ha operado como marco de síntesis analítica y no como una característica declarada de los estudios primarios incluidos. La mayoría del corpus emplea constructos cognitivos —carga cognitiva, metacognición, funciones ejecutivas, engagement— sin fundamentarlos explícitamente en marcos neurocientíficos formales. Solo tres estudios muestran una aproximación más directa al dominio neuroeducativo en su diseño o en la interpretación de sus resultados: Henze et al. (2026), Liew et al. (2025) y Jauhiainen y Garagorry Guerra (2024).

Esta situación define con precisión la frontera actual del campo: existe evidencia conductual y perceptual creciente sobre los efectos de GenAI en dimensiones relevantes para el aprendizaje, pero la brecha entre esa evidencia y una comprensión de los mecanismos neurocognitivos subyacentes permanece abierta. Esa brecha constituye simultáneamente la limitación central

del conocimiento disponible y la agenda de investigación más urgente para que el campo avance hacia afirmaciones causales sólidas sobre GenAI y aprendizaje. Los diseños y marcos necesarios para cerrarla —incluyendo medidas neurocognitivas objetivas, estudios longitudinales y marcos teóricos que integren evidencia neurocientífica con modelos pedagógicos— se discuten en la sección siguiente.

DISCUSIÓN

Los tres patrones transversales identificados en este corpus comparten una característica que los marcos de adopción tecnológica dominantes en el campo —TAM, UTAUT y sus extensiones— no pueden explicar: ninguno informa sobre quién usa GenAI ni sobre las condiciones de adopción, sino sobre qué ocurre cognitivamente durante ese uso y bajo qué condiciones ese proceso produce extensión o desplazamiento del pensamiento del aprendiz. La teoría de la cognición extendida (Clark & Chalmers, 1998) ofrece el principio integrador que los tres patrones demandan: cuando una herramienta cumple condiciones de acoplamiento funcional —disponibilidad confiable, accesibilidad inmediata e integración fluida con los procesos internos del agente— se convierte en componente del sistema cognitivo del aprendiz, no en auxiliar externo. De este principio se derivan dos modos de relación entre aprendiz y GenAI que organizan la interpretación del corpus: el modo sustitutivo, en que GenAI reemplaza procesamiento que debería ocurrir internamente, y el modo extensivo, en que GenAI amplía el alcance del procesamiento cognitivo del aprendiz. Desde este marco, el Patrón 1 —mediación afectivo-motivacional como mecanismo previo a los efectos cognitivos— se explica porque el incremento de engagement y la reducción de ansiedad documentados de forma consistente en el corpus son las condiciones que habilitan la disposición del aprendiz a integrar GenAI funcionalmente: sin esa apertura afectiva, el acoplamiento no se produce y GenAI opera como herramienta desacoplada, lo que explica por qué los efectos afectivos son anteriores y más consistentes que los efectos cognitivos objetivos en los estudios de intervención (Henze et

al., 2026; Salhab & Aboushi, 2026; Huang et al., 2025). El Patrón 2 —alfabetización en IA como moderadora del beneficio— se explica porque el acoplamiento cognitivo no es automático: requiere que el aprendiz posea los esquemas necesarios para integrar la herramienta en su procesamiento; sin alfabetización en IA, GenAI produce desplazamiento cognitivo; con ella, produce extensión cognitiva, lo que explica por qué el acceso tecnológico por sí solo no predice mejoras de aprendizaje (Wang et al., 2026; Salhab & Aboushi, 2025; Klarin et al., 2024). El Patrón 3 —diseño conversacional como determinante del proceso cognitivo activado— se explica porque las condiciones del acoplamiento son sensibles a las características de la interfaz: herramientas que generan respuestas completas activan el modo sustitutivo; herramientas que demandan elaboración del aprendiz activan el modo extensivo; la Cognitive Load Theory precisa el mecanismo subyacente, pues el diseño sustitutivo incrementa la carga extrínseca mientras el diseño extensivo gestiona la carga intrínseca (Westbye et al., 2026; Liew et al., 2025; Li et al., 2025).

Estos patrones son convergentes con hallazgos clave de la literatura de referencia. La distinción entre uso sustitutivo y complementario propuesta por Rivera-Novoa & Arias (2025) es coherente con el modo sustitutivo/extensivo del marco aplicado aquí; Gibson et al. (2023) converge con el Patrón 2 al proponer que GenAI como artefacto cognitivo redistribuye carga entre agente y herramienta; Cukurova (2024) coincide conceptualmente con el marco de cognición extendida, con la diferencia de que opera a nivel propositivo sin corpus empírico, mientras este artículo valida esa propuesta sobre evidencia sistemática. Sin embargo, Fan et al. (2024) documentan que GenAI mejora el desempeño inmediato sin producir motivación intrínseca sostenida ni transferencia del conocimiento, lo que tensiona el Patrón 1: desde el marco de cognición extendida, esta disociación se explica porque el acoplamiento afectivo inicial no garantiza autonomía cognitiva progresiva cuando el diseño instruccional no facilita deliberadamente la transición del engagement situacional hacia la motivación de dominio, lo

que sitúa la responsabilidad en el diseño pedagógico, no en la herramienta. Yan et al. (2024) formulan el problema en términos normativos —qué condiciones debe cumplir GenAI para producir aprendizaje— mientras este artículo identifica los mecanismos que explican cuándo el acoplamiento es extensivo o sustitutivo, ofreciendo mayor especificidad explicativa. En cuanto a los vacíos que este artículo contribuye a delimitar en la literatura, Ogunleye et al. (2024), Qian (2025) y Bahroun et al. (2023) formulan recomendaciones de diseño instruccional sin el mecanismo explicativo que el Patrón 3 aporta: ninguna revisión previa ha tratado el diseño conversacional como variable independiente con efectos diferenciales sobre el modo de acoplamiento cognitivo, ni ha abordado la alfabetización en IA como condición de posibilidad del acoplamiento funcional en lugar de competencia deseable —vacíos que los Patrones 2 y 3 llenan con evidencia sistemática.

Los hallazgos tienen implicaciones concretas para el diseño instruccional y para la agenda de investigación. Del Patrón 3 se deriva que las intervenciones con GenAI deben privilegiar configuraciones en modo extensivo —herramientas que pregunten antes de responder y que requieran elaboración del aprendiz antes de ofrecer síntesis—. Del Patrón 2, que toda intervención debe incorporar el desarrollo simultáneo de alfabetización en IA como condición de acoplamiento, no como objetivo secundario. Del Patrón 1, que la secuencia instruccional importa: el acoplamiento afectivo debe preceder a los objetivos cognitivos de orden superior. En cuanto a la agenda futura, el Patrón 1 señala la necesidad de diseños longitudinales que determinen si el acoplamiento afectivo inicial produce acoplamiento cognitivo sostenido; el Patrón 3 demanda estudios que manipulen sistemáticamente el diseño conversacional como variable independiente para precisar las condiciones del modo extensivo; ambos patrones, junto con el vacío neurocognitivo identificado en la sección 3.2.4, apuntan a la incorporación de medidas objetivas de procesamiento —fNIRS, eye-tracking— que permitan observar directamente el modo de acoplamiento durante la interacción con GenAI en lugar de inferirlo

desde el autoinforme. La integración sistemática entre evidencia empírica sobre GenAI y marcos neuroeducativos que este artículo propone no es, por tanto, un ejercicio de aplicación teórica sino el paso necesario para que el campo avance desde la descripción de efectos hacia la explicación de los mecanismos que los producen.

CONCLUSIONES

Los resultados de esta revisión sistemática permiten responder con precisión las cuatro preguntas de investigación planteadas. En respuesta a PI1 y PI2, la producción empírica sobre GenAI en aprendizaje se organiza en cinco dominios —pedagógico/adopción, metacognitivo, socioemocional, cognitivo y neuroeducativo— con marcada concentración en el primero, y documenta tres convergencias con respaldo suficiente en el corpus: el efecto positivo de GenAI sobre la motivación y el engagement del aprendiz, la mediación de la utilidad percibida y la influencia social en la adopción de estas herramientas, y la personalización dinámica como atributo diferencial de GenAI respecto a tecnologías educativas previas. Junto a estas convergencias, el corpus registra divergencias sistemáticas en los efectos sobre el rendimiento académico objetivo, la carga cognitiva y las dimensiones socioemocionales, cuya bidireccionalidad no puede atribuirse a errores de estudios individuales sino a la sensibilidad del fenómeno al contexto, al perfil del aprendiz y al diseño de la herramienta.

En respuesta a PI3, la interpretación de estos hallazgos desde el marco de la cognición extendida revela que los efectos de GenAI en el aprendizaje responden a tres patrones transversales que no son visibles en el examen de estudios individuales. El primero es que la mediación afectivo-motivacional opera como mecanismo de entrada al acoplamiento cognitivo, lo que explica por qué los efectos afectivos son anteriores y más consistentes que los efectos cognitivos objetivos en el corpus. El segundo es que la alfabetización en IA del aprendiz funciona como condición de posibilidad del acoplamiento funcional: sin ella, GenAI produce desplazamiento cognitivo; con ella, produce extensión cognitiva. El tercero es que el

diseño conversacional de la herramienta determina el modo de acoplamiento —sustitutivo o extensivo— y, por tanto, el tipo de proceso cognitivo que GenAI activa en el aprendiz, con mayor peso que la mera presencia de la tecnología. Estos tres patrones constituyen el principal aporte analítico de esta revisión y ofrecen un principio explicativo integrador que los marcos de adopción tecnológica dominantes en el campo no proporcionan.

En respuesta a PI4, cinco vacíos estructurales delimitan las fronteras actuales del conocimiento: la ausencia de diseños longitudinales que verifiquen la sostenibilidad de los efectos observados; la inexistencia de medidas neurocognitivas objetivas que permitan superar la dependencia del autoinforme; la subrepresentación de educación básica y etapas adolescentes; la concentración geográfica con ausencia del Sur Global y América Latina; y el escaso número de diseños experimentales con poder estadístico suficiente para establecer relaciones causales.

La contribución de este estudio al campo es de naturaleza teórica, analítica, metodológica y prospectiva. Teóricamente, propone la primera integración sistemática entre evidencia empírica sobre GenAI en aprendizaje y el marco de la cognición extendida como lente interpretativa explícita. Analíticamente, identifica patrones transversales que transforman tensiones aparentes de la literatura en un sistema coherente de mecanismos explicativos. Metodológicamente, aplica PRISMA 2020 con criterios de elegibilidad explícitamente justificados y evaluación de calidad mediante instrumentos diferenciados según el diseño de cada estudio. Prospectivamente, formula una agenda de investigación fundamentada en los vacíos identificados con precisión suficiente para orientar diseños futuros.

Las conclusiones deben leerse con las reservas que impone la base metodológica del corpus: la evidencia disponible es indicativa pero no confirmatoria. El riesgo global moderado homogéneo en los 28 estudios, la predominancia del autoinforme como método de medición, la concentración del corpus en una sola editorial y la ausencia de registro prospectivo del protocolo de esta revisión son limitaciones que condicionan el alcance de las afirmaciones aquí

formuladas. Ninguna de las conclusiones debe interpretarse como establecimiento de relaciones causales entre el uso de GenAI y resultados de aprendizaje verificables: representan tendencias con respaldo empírico suficiente para orientar decisiones pedagógicas informadas y para fundamentar hipótesis de investigación futura, no para prescribir prácticas educativas de forma definitiva.

Referencias bibliográficas

- Arowosegbe, A., Alqahtani, J. S., & Oyelade, T. (2024). Perception of generative AI use in UK higher education. *Frontiers in Education*, 9, 1463208. <https://doi.org/10.3389/feduc.2024.1463208>
- Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability*, 15(17), Artículo 12983. <https://doi.org/10.3390/su151712983>
- BinJwair, A. (2025). Predicting STEM students' adoption of generative AI in academic contexts: An application of the UTAUT model. *Frontiers in Education*, 10, 1669750. <https://doi.org/10.3389/feduc.2025.1669750>
- Cao, C., Li, Z., Ni, M., Luo, F., & Ling, C. (2025). Will human-like features effect the adoption of generative AI tools? A study in Chinese University students. *Frontiers in Education*, 10, 1673150. <https://doi.org/10.3389/feduc.2025.1673150>
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19. <https://doi.org/10.1093/analys/58.1.7>
- Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative and mixed methods research. *Qualitative Health Research*, 22(10), 1435–1443. <https://doi.org/10.1177/1049732312452938>

- Cukurova, M. (2024). The interplay of learning, analytics, and artificial intelligence in education. ArXiv. <https://doi.org/10.48550/arxiv.2403.16081>
- Dogaru, M., Pisică, O., Popa, C.-Ștefan, Răgman, A.-A., & Tololoi, I.-R. (2025). The perceived impact of artificial intelligence on academic learning. *Frontiers in Artificial Intelligence*, 8, 1611183. <https://doi.org/10.3389/frai.2025.1611183>
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530. <https://doi.org/10.1111/bjet.13544>
- George-Reyes, C. E., Vilhunen, E., Avello-Martínez, R., & López-Caudana, E. (2024). Developing scientific entrepreneurship and complex thinking skills: Creating narrative scripts using ChatGPT. *Frontiers in Education*, 9, 1378564. <https://doi.org/10.3389/feduc.2024.1378564>
- Gibson, D., Kovanović, V., Ifenthaler, D., Dexter, S., & Feng, S. (2023). Learning theories for artificial intelligence promoting learning processes. *British Journal of Educational Technology*, 54(5), 1125–1146. <https://doi.org/10.1111/bjet.13341>
- Gozalo-Brizuela, R., & Garrido-Merchán, E. (2023). A survey of generative AI applications. ArXiv. <https://doi.org/10.48550/arxiv.2306.02781>
- Guetala, M., Bourekache, S., Kazar, O., & Harous, S. (2024). Generative artificial intelligence in education: Advancing adaptive and personalized learning. *Acta Informatica Pragensia*. <https://doi.org/10.18267/j.aip.235>
- He, S., & Ren, Y. (2025). Exploring pre-service music teachers' acceptance of generative artificial intelligence: A PLS-SEM-ANN approach. *Frontiers in Psychology*, 16, 1571279. <https://doi.org/10.3389/fpsyg.2025.1571279>

- Henze, J., Lademann, J., Becker-Genschow, S., & Bresges, A. (2026). AI-supported data analysis boosts student motivation and reduces stress in physics education. *Frontiers in Education*, 11, 1719670. <https://doi.org/10.3389/feduc.2026.1719670>
- Henze, J., Lademann, J., Bresges, A., & Becker-Genschow, S. (2025). Iterative development of an AI intervention for pre-service physics teachers from a Vygotskian perspective. *Frontiers in Education*, 10, 1707534. <https://doi.org/10.3389/feduc.2025.1707534>
- 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>
- Hu, Y., Liu, L., Yang, H., & Jia, X. (2026). The effects of perceived teacher support on online mathematics learning power: The mediating roles of artificial intelligence literacy and cognitive tools. *Frontiers in Psychology*, 17, 1763924. <https://doi.org/10.3389/fpsyg.2026.1763924>
- Huang, Y., Chen, H., & Hu, C. (2025). L2 growth mindset in AI-mediated language learning: Effects of perceived usability and presence of generative AI chatbots. *Frontiers in Psychology*, 16, 1700117. <https://doi.org/10.3389/fpsyg.2025.1700117>
- Jauhiainen, J. S., & Garagorry Guerra, A. (2024). Generative AI and education: Dynamic personalization of pupils' school learning material with ChatGPT. *Frontiers in Education*, 9, 1288723. <https://doi.org/10.3389/feduc.2024.1288723>
- Kazazoglu, S., & Turun Özel, G. (2025). The impact of chatbots on emotional intelligence in AI-assisted EFL writing. *Frontiers in Education*, 10, 1624272. <https://doi.org/10.3389/feduc.2025.1624272>

- Kinskofer, F., & Tulis, M. (2025). Motivational and appraisal factors shaping generative AI use and intention in Austrian higher education students and teachers. *Frontiers in Education*, 10, 1677827. <https://doi.org/10.3389/feduc.2025.1677827>
- Klarin, J., Hoff, E., Larsson, A., & Daukantaitė, D. (2024). Adolescents' use and perceived usefulness of generative AI for schoolwork: Exploring their relationships with executive functioning and academic achievement. *Frontiers in Artificial Intelligence*, 7, 1415782. <https://doi.org/10.3389/frai.2024.1415782>
- Li, W., Cui, X., Manoharan, P., Dai, L., Liu, K., & Huang, L. (2025). AI-assisted feedback and reflection in vocal music training: Effects on metacognition and singing performance. *Frontiers in Psychology*, 16, 1598867. <https://doi.org/10.3389/fpsyg.2025.1598867>
- Liew, T. W., Tan, S.-M., Pang, W. M., Gan, C. L., Chan, T. J., & Ahmad, F. (2025). Cringe, lit, or mid: Affective and cognitive effects of youth slang in an educational chatbot. *Acta Psychologica*, 256, 105036. <https://doi.org/10.1016/j.actpsy.2025.105036>
- Marzano, D. (2025). Generative artificial intelligence (GAI) in teaching and learning processes at the K-12 level: A systematic review. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-025-09853-7>
- Mohamed, A. M., Emran, A. Q., Abdelrady, A. H., & Deen, A. A. M. M. A. E. (2026). Giftedness and academic motivation in GenAI contexts: The moderating and mediating role of gender and AI anxiety. *Frontiers in Education*, 11, 1762707. <https://doi.org/10.3389/feduc.2026.1762707>
- Munn, Z., Barker, T. H., Moola, S., Tufanaru, C., Stern, C., McArthur, A., Stephenson, M., & Aromataris, E. (2020). Methodological quality of case series studies: An introduction to the JBI critical appraisal tool. *JBIC Evidence Synthesis*, 18(10), 2127–2133. <https://doi.org/10.11124/JBIES-19-00099>

- Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277–304. <https://doi.org/10.1080/15228053.2023.2233814>
- Neittaanmäki, N. (2025). Swedish medical students' attitudes toward artificial intelligence and effects on career plans: A survey. *Frontiers in Education*, 10, 1517116. <https://doi.org/10.3389/educ.2025.1517116>
- Ogunleye, B., Zakariyyah, K., Ajao, O., Olayinka, O., & Sharma, H. (2024). A systematic review of generative AI for teaching and learning practice. *Education Sciences*, 14(6), Artículo 636. <https://doi.org/10.3390/educsci14060636>
- 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, Y., & Li, G. (2025). The effects of perceived teacher support and growth language mindset on learner well-being in AI-integrated environment: The mediating role of generative AI attitude. *Frontiers in Psychology*, 16, 1660462. <https://doi.org/10.3389/fpsyg.2025.1660462>
- Pesovski, I., Santos, R., Henriques, R., & Trajkovik, V. (2024). Generative AI for customizable learning experiences. *Sustainability*, 16(7), Artículo 3034. <https://doi.org/10.3390/su16073034>
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic

- reviews: A product from the ESRC Methods Programme. Lancaster University.
<https://doi.org/10.13140/2.1.1018.4643>
- Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, 69, 1105–1120. <https://doi.org/10.1007/s11528-025-01100-1>
- Rivera-Novoa, A., & Arias, D. (2025). Generative artificial intelligence and extended cognition in science learning contexts. *Science & Education*. <https://doi.org/10.1007/s11191-025-00660-1>
- Salhab, R., & Aboushi, M. M. (2025). Influence of AI literacy and 21st-century skills on the acceptance of generative artificial intelligence among college students. *Frontiers in Education*, 10, 1640212. <https://doi.org/10.3389/feduc.2025.1640212>
- Salhab, R., & Aboushi, M. M. (2026). Impact of AI-assisted microlearning on student engagement in an online environment in higher education. *Frontiers in Education*, 11, 1766032. <https://doi.org/10.3389/feduc.2026.1766032>
- Sergeeva, O. V., Masalimova, A. R., Zheltukhina, M. R., Chikileva, L. S., Lutskovskai, L. Yu., & Luzin, A. (2025). Impact of digital media literacy on attitude toward generative AI acceptance in higher education. *Frontiers in Education*, 10, 1563148. <https://doi.org/10.3389/feduc.2025.1563148>
- Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A. W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., ... Higgins, J. P. T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355, i4919. <https://doi.org/10.1136/bmj.i4919>

- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, 366, l4898. <https://doi.org/10.1136/bmj.l4898>
- Susanto, H., Qur'ainal Mutalif, A., Setiana, D., & Besar, N. (2026). From usefulness to motivation: A TAM–SDT perspective on AI adoption and problematic internet use in higher education. *Frontiers in Education*, 10, 1699827. <https://doi.org/10.3389/feduc.2025.1699827>
- Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, 68(1), 1–16. <https://doi.org/10.1007/s11423-019-09701-3>
- Wang, H., & Liu, T. (2026). Psychological determinants of GenAI adoption for foreign language education: An extended UTAUT2 model and sentiment analysis approach. *Frontiers in Psychology*, 16, 1622926. <https://doi.org/10.3389/fpsyg.2025.1622926>
- Wang, J., Bai, B., & An, Q. (2026). Enhancing college students' AI literacy through generative AI use: A mixed-methods investigation. *Frontiers in Psychology*, 17, 1728785. <https://doi.org/10.3389/fpsyg.2026.1728785>
- Westbye, A. K., Eriksen, H., & Blikstad-Balas, M. (2026). When AI only asks: How question-driven dialogue shapes prewriting in the classroom. *Frontiers in Education*, 11, 1740044. <https://doi.org/10.3389/feduc.2026.1740044>
- Xu, S., Rahim, N., & Zeng, S. (2026). The application of generative AI in university dance education: Effects on dance skills, engagement and learning motivation. *Frontiers in Education*, 11, 1756945. <https://doi.org/10.3389/feduc.2026.1756945>

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8, 1839–1850.
<https://doi.org/10.1038/s41562-024-02004-5>

Yu, C., Yan, J., & Cai, N. (2024). ChatGPT in higher education: Factors influencing ChatGPT user satisfaction and continued use intention. *Frontiers in Education*, 9, 1354929.
<https://doi.org/10.3389/feduc.2024.1354929>

Anexos

Anexo 1

Características de los estudios incluidos

	Autor /Año	País	Revista	Año	Diseño	Metodología	Muestra	Tipo de IA	Contexto educativo	Variables	Instrumentos	Resultados principales	Evaluación metodológica	Periodo estudio	Disponibilidad datos/código	Registro Protocolo
1	Arowosegbire et al. (2024)	Reino Unido	Frontiers in Education	2024	No experimental	Encuesta transversal	136 estudiantes	Gen-AI	Educación superior	Uso, percepción, beneficios, riesgos	Cuestionario online	Uso frecuente para ideas y corrección; percepción positiva moderada; preocupaciones éticas	Modificada	Ene–Abr 2024	NR	NR
2	BinJwair (2025)	Arabia Saudita	Frontiers in Education	2025	No experimental	Encuesta + PLS-SEM	464 estudiantes STEM	IA generativa	Educación superior	Expectativa de desempeño, esfuerzo, influencia social, condiciones	Cuestionario UTAUT	Predictores significativos de adopción de IA	Modificada-alta	NR	NR	NR
3	Cao et al. (2025)	China	Frontiers in Education	2025	No experimental	Encuesta UTAUT extendido	378 estudiantes	IA generativa	Educación superior	Intención de uso, innovación, confianza, riesgo	Cuestionario	Factores cognitivos positivos; semejanza humana negativa	Modificada	NR	NR	NR
4	Dogaru et al. (2025)	Rumania	Frontiers in Artificial Intelligence	2025	No experimental	Encuesta Likert	84 estudiantes	ChatGPT	Educación superior	Motivación, aprendizaje, ética, dependencia	Cuestionario validado	Mejora motivación y aprendizaje; riesgo de dependencia	Modificada	mar-24	Sí (OSF)	NR
5	Georgie-Reyes et al. (2024)	Ecuador	Frontiers in Education	2024	Cuasi-experimental	Pretest – posttest	105 estudiantes	ChatGPT	Educación superior	Emprendimiento, pensamiento complejo	Prueba pre-post	Mejora significativa en habilidades	Modificada	2022 – 2023	NR	NR

6	He y Ren (2025)	China	Frontiers in Psychology	2025	No experimental transversal	Cuantitativa; PLS-SEM y ANN	301 futuros docentes de música	IA generativa	Educación superior; formación inicial docente en música	Expectativa de desempeño, esfuerzo, influencia social, hábito, riesgo percibido, compatibilidad, intención conductu al, uso real	Cuestionario UTAUT 2 extendido	Riesgo percibido, influencia social y hábito influyen en intención conductu al; intención conductu al y riesgo percibido predijeron uso real	Modorada	NR (Not Reported)	Contribuciones incluidas en artículo /material suplementario	NR (Not Reported)
7	Henze et al. (2025)	Alemania	Frontiers in Education	2025	Pre-post experimental no aleatorizado / DBR	Cuantitativa; intervención iterativa de dos sesiones	31 futuros docentes de física	Herramientas de IA generativa	Educación superior; formación inicial docente en física	Alfabetización en IA, actitudes hacia IA, preparación percibida, apertura hacia IA	Test validado de 30 ítems; encuesta Likert de 4 puntos	Aumentos pequeños y no significativos en alfabetización global; mejoras selectivas en competencias con actividades prácticas; mayor apertura y preparación percibida	Modorada	Invierno 2023 /2024 y vera no 2024	Incluida en artículo /material suplementario; consultas al autor	NR (Not Reported)
8	Henze et al. (2026)	Alemania	Frontiers in Education	2026	RCT / experimental aleatorizado	Cuantitativa; comparación AI vs Excel; pre-post; encuestas	50 estudiantes docentes	Chatbot GPT personalizado Experimento	Educación física; análisis de datos experimentales	Aprendizaje, motivación, disfrute, estrés/frustración, efectividad percibida	Pretest-postest; encuestas estructuradas	Ambos grupos mejoraron sin diferencias cognitivas significativas; grupo IA mostró mayor motivación, disfrute	Modorada-alta	NR (Not Reported)	NR (Not Reported)	NR (Not Reported)

											y efectivida d percibida					
9	Hu et al. (2026)	China	Frontiers in Psychology	2026	No experimental transversal	Cuantitativa; SEM y bootstrap	758 estudiantes universitarios	GenAI / alfabetización en IA	Educación superior; matemáticas online	Apoyo docente percibido, alfabetización en IA, herramientas cognitivas, poder de aprendizaje matemático online	Cuestionario de 27 ítems; escalas Likert adaptadas	Apoyo docente predijo directamente el aprendizaje matemático online e indirectamente mediante alfabetización en IA y herramientas cognitivas	Modera	NR (Not Reported)	Datos disponibles por solicitud sin restricción indebida	NR (Not Reported)
10	Huang et al. (2025)	China	Frontiers in Psychology	2025	No experimental transversal	Cuantitativa; SEM	530 estudiantes universitarios	Chatbots generativos; ChatGPT y AICibot	Aprendizaje oral de inglés como L2	Usabilidad percibida, presencia, inteligencia emocional, disposición a comunicarse, mentalidad de crecimiento L2	Cuestionario con subescalas validadas	Usabilidad y presencia predijeron mentalidad de crecimiento L2; efectos mediados por inteligencia emocional y disposición a comunicarse	Modera	20 abril -20 mayo 2025	NR (Not Reported)	NR (Not Reported)
11	Jauhian y Garay Guerra (2024)	Uruguay	Frontiers in Education	2024	No experimental / intervención de aula	Cuantitativa descriptiva	110 estudiantes de 8–14 años, grados 4–6	ChatGPT-3.5, ChatGPT-4 y Midjourney	Educación primaria escolar	Personalización dinámica, rendimiento, motivación, conocimiento previo, aprendizaje percibido	Plataforma Digileac; actividades, ejercicios y evaluación en clase	La IA generativa personalizó material según nivel de conocimiento; favoreció motivación, rendimiento y	Modera	jul-23	NR (Not Reported)	NR (Not Reported)

												logro de objetivos curricul ares				
1 2	Kazaz oglu y Turun Özel (2025)	Tur quía	Fronti ers in Educa tion	2 0 2 5	No experim ental / interven ción educati va	Cuanti tativa	Estudia ntes de segundo año de formaci ón docente en inglés; tamaño muestra 1 NR (Not Reporte d) en fragmen tos disponi bles	ChatG PT	Forma ción inicial docent e; escritu ra EFL asistid a por IA	Intelligen cia emociona l, uso de ChatGPT , adaptabil idad, estrés, habilidades intra/inte rpersonales	Versión turca del Bar-On Emotion al Quotient Inventor y (EQ-I)	Mayor uso de ChatGP T se asoció con menor intelligen cia emocion al; leve mejora en adaptabi lidad interpret ada con cautela	Mode rada	NR (Not Rep orted)	NR (Not Reporte d)	NR (Not Rep orte d)
1 3	Kinsk ofer y Tulis (2025)	Aus tria	Fronti ers in Educa tion	2 0 2 5	No experim ental transver sal	Cuanti tativa; SEM confir matori o, PLS- SEM y regresi ón lineal	3.094 estudian tes y 1.767 docente s	GenA I / ChatG PT	Educa ción superi or	Intención conductu al, uso de GenAI, expectati va de desempe ño, expectati va de esfuerzo, influenci a social, competen cia subjetiva, motivaci ón intrínseca , confianza , desafío, amenaza	Cuestion arios UTAUT y escalas motivaci onales/d e appraisa l	La intenció n conduct ual predijo fuertem ente el uso; expectat iva de desempe ño, esfuerzo e influenc ia social fueron predicto res positivo s; desafío predijo intenció n y amenaza tuvo efecto negativo pequeño	Mode rada- alta	Sem estre de vera no 2024	Materia l suplem entario disponi ble	NR (Not Rep orte d)
1 4	Klarin et al. (2024)	Sue cia	Fronti ers in Artifi cial Intelli gence	2 0 2 4	No experim ental transver sal; dos estudios	Cuanti tativa	Estudio 1: 385 adolesc entes, media 14 años; Estudio 2: 359	Chatb ots genera tivos LLM, especi almen te	Trabaj o escolar en adoles centes	Uso de IA generativ a, utilidad percibida , funciones	BRIEF- 2; medidas de uso/utili dad de IA generati	Adolesc entes con mayores dificulta des de función ejecutiv	Mode rada	NR (Not Rep orted)	NR (Not Reporte d)	NR (Not Rep orte d)

							adolescentes, media 17 años	ChatGPT			ejecutivas, rendimiento académico	va; rendimiento académico	a percibieron la IA generativa como más útil; rendimiento académico no mostró asociaciones significativas con uso o utilidad en Estudio 1				
15	Li et al. (2025)	China / Malasia	Frontiers in Psychology	2025	Mixto experimental aleatorizado	Métodos mixtos; diseño experimental pretest - posttest con grupo control; análisis temático	Grupo control N=42; grupo experimental N=38; total N=80	Yuanbao, chatbot generativo de Tencent basado en LLM	Formación vocal musical de futuros docentes	Metacognición, rendimiento vocal, retroalimentación, reflexión, autoevaluación	ANOVA de medidas repetidas; modelos lineales mixtos; diarios reflexivos; registros de interacción con IA; pruebas de inteligencia y habilidad musical	La intervención con IA mejoró significativamente la metacognición; ambos grupos mejoraron en canto, sin interacción grupo-tiempo significativa; no hubo correlación entre metacognición y rendimiento vocal	Alta-moderada	Seis semanas	NR (Not Reporte d)	NR (Not Rep orte d)	
16	Liew et al. (2025)	Malasia / Reino Unido	Acta Psychologica	2025	Experimental intra-sujetos contrabalanceado	Cuantitativa con análisis cualitativo de comentarios	Estudiantes universitarios de negocios; N NR (Not Reporte d)	ChatGPT-powered educational chatbot	Educación superior	Lenguaje juvenil, humanización percibida, engagement, emoción, arousal, carga cognitiva	Interacción con dos versiones de chatbot; escalas de percepción, emoción y carga cognitiva;	El slang aumentó humanización, engagement, emoción positiva y arousal, pero también dificultó	Modera-alta	NR (Not Reported)	NR (Not Reporte d)	NR (Not Rep orte d)	

											comentarios cualitativos	percibida/carga extrínseca				
17	Mohamed et al. (2026)	Arabia Saudita	Frontiers in Education	2026	No experimental correlacional	Cuantitativa; SEM, mediación y moderación	304 universitarios EFL dotados	GenAI	Educación superior EFL	Dotación, ansiedad GenAI, motivación académica, género	Gifted Rating Scales; instrumentos verificados de conducta dotada, ansiedad GenAI y motivación académica	La conducta dotada se asoció con menor ansiedad GenAI y mayor motivación; la ansiedad GenAI medió parcialmente; género moderó relaciones	Modorada	NR (Not Reported)	NR (Not Reported)	NR (Not Reported)
18	Neittanmäki (2025)	Suecia	Frontiers in Education	2025	No experimental transversal	Cuantitativa; encuestas online	172 estudiantes de medicina	Chatbots de IA generativa	Educación médica	Actitudes hacia IA, uso de IA generativa, planes de carrera, percepción de reemplazo profesional	Encuesta anónima online de 25 preguntas en Survey Monkey	99% conocía chatbots generativos; 64% los usó en estudios médicos; 58% consideró que apoyan el aprendizaje	Modorada	4–19 marzo 2024	NR (Not Reported)	NR (Not Reported)
19	Pan y Li (2025)	China	Frontiers in Psychology	2025	No experimental transversal	Cuantitativa; correlación, regresión y mediación	486 estudiantes universitarios EFL	IA generativa / actitud hacia GenAI	Educación superior EFL	Apoyo docente percibido, growth language mindset, actitud hacia GenAI, bienestar del aprendizaje	Cuestionario con escalas de PTS, GLM, actitud GenAI y bienestar	PTS y GLM predijeron positivamente actitud hacia GenAI y bienestar; actitud GenAI medió parcialmente	Modorada	NR (Not Reported)	NR (Not Reported)	NR (Not Reported)
20	Salhab y Aboushi	Palestina	Frontiers in Education	2025	No experimental transversal	Cuantitativa	260 estudiantes universitarios seleccio	IA generativa	Educación superior	Alfabetización en IA, habilidades del siglo	Tres escalas cuantitativas	La alfabetización en IA y las habilidades del	Modorada	NR (Not Reported)	NR (Not Reported)	NR (Not Reported)

	(2025)						nados aleatori amente			XXI, aceptació n de IA generativ a		siglo XXI influyeron en la aceptación de IA generativa				
2 1	Salha b y Abou shi (2026)	Pale stin a	Fronti ers in Educa tion	2 0 2 6	Mixto cuasi- experim ental	Enfoq ue mixto secuen cial explic ativo; cuasi- experi mento y fenom enología	50 estudian tes; 25 entrevis tas semiestr ucturadas	IA genera tiva aplica da a microl earnin g	Educa ción superi or online; curso multim edia en tecnol ogía educati va	Engagem ent cognitivo , emociona l, social y conductu al	ANCOV A; discusio nes en foros online; entrevist as semiestr ucturadas; análisis de contenid o inductiv o y deductiv o	El microlea rning asistido por IA influyó positiva mente en el engage ment y sus dimensi ones; emergen temas de atención , personal ización, menor frustraci ón, cumpli miento de tareas, intercam bio y competencia	Mode rada- alta	NR (Not Rep orted)	NR (Not Reporte d)	NR (Not Rep orte d)
2 2	Serge eva et al. (2025)	Rusi a	Fronti ers in Educa tion	2 0 2 5	No experim ental transver sal	Cuanti tativa; encues ta transv ersal; SEM	451 estudian tes universi tarios de pregrado	IA genera tiva	Educa ción superi or	Alfabetiz ación mediática digital, acceso a dispositiv os, acceso a contenid o, comprens ión técnica, comprens ión crítica, creación, aceptació n de GenAI	Escala de alfabetiz ación mediática digital; escala de aceptación de GenAI modificada	Acceso a contenid o y creación tuvieron efecto positivo; comprens ión técnica tuvo efecto negativo ; acceso a dispositi vos tuvo efecto positivo modesto ; comprens ión	Mode rada	NR (Not Rep orted)	NR (Not Reporte d)	NR (Not Rep orte d)

											crítica no fue significa tiva					
2 3	Susan to et al. (2026)	Bru néi	Fronti ers in Educa tion	2 0 2 6	No experim ental transver sal	Cuanti tativa; encues ta transv ersal; PLS- SEM	Estudia ntes universi tarios; N NR (Not Reporte d) en fragmen tos disponi bles	ChatG PT	Educa ción superi or	Utilidad percibida , facilidad de uso percibida , uso de ChatGPT , uso problemá tico de internet, motivaci ón de aprendiza je	Cuestion ario; PLS- SEM; alfa de Cronbac h, confiabil idad compues ta, AVE	La utilidad percibid a predijo uso de ChatGP T; uso de ChatGP T predijo motivaci ón; uso problem ático de internet tuvo efecto negativo en motivaci ón; utilidad percibid a influyó indirecta mente mediant e uso de ChatGP T	Mode rada	NR (Not Rep orted)	NR (Not Reporte d)	NR (Not Rep orte d)
2 4	Wang et al. (2026)	Chi na	Fronti ers in Psych ology	2 0 2 6	Mixto secuenc ial explorat orio	Métod os mixtos ; entrevi stas cualita tivas y encues ta cuantit ativa; SEM	30 entrevi stas; 590 encuest as válidas	IA genera tiva	Educa ción superi or	Uso de GenAI, alfabetiza ción en IA, entorno social, impresio nes percibida s, necesida des grupales, facilidad de uso percibida , expectati vas tecnológi cas, adopción conductu al	Entrevis tas; cuestion ario; análisis cualitati vo; SEM	El uso de GenAI favoreci ó alfabetiz ación en IA mediant e ciclos cognició n- práctica- evaluaci ón; el entorno social influyó indirecta mente en práctica de aplicaci ón; facilidad de uso y expectat ivas	Mode rada- alta	NR (Not Rep orted)	NR (Not Reporte d)	NR (Not Rep orte d)

											tecnológicas predijeron adopción y alfabetización futura					
25	Wang y Liu (2026)	China	Frontiers in Psychology	2026	Mixto	SEM y análisis de sentimiento	632 participantes; reseñas online	GenAI / DeepSeek	Educación en lenguas extranjeras	Expectativa de desempeño, expectativa de esfuerzo, influencia social, ansiedad hacia IA, hábito, intención conductual	Encuesta UTAUT 2 extendida; análisis de sentimiento de reseñas online	Expectativa de desempeño, esfuerzo e influencia social predijeron positivamente intención; ansiedad hacia IA y hábito tuvieron efecto negativo	Modera-alta	NR (Not Reported)	NR (Not Reported)	NR (Not Reported)
26	Westbye et al. (2026)	Noruega	Frontiers in Education	2026	Cualitativo exploratorio	Análisis reflexivo de contenido	17 estudiantes de secundaria inferior; 201 pares preguntados	Chatbot restringido basado en GPT-4	Escritura escolar; fase de prescripción	Elaboración de ideas, apropiación, resistencia, agencia, descripción	Diálogos de estudiantes-te chatbot; análisis reflexivo de contenido	El chatbot apoyó elaboración descriptiva en algunos estudiantes; otros mostraron mínima apropiación o resistencia	Modera	feb-25	NR (Not Reported)	NR (Not Reported)
27	Xu et al. (2026)	China / Malasia	Frontiers in Education	2026	Cuasi-experimental aleatorizado	Comparación grupo experimental vs. control	60 estudiantes universitarios	GEN Dance, herramienta de enseñanza basada en GenAI	Educación universitaria en danza	Habilidades de danza, engagement, motivación de aprendizaje	Herramienta GEN Dance; herramienta multimedia convencional; mediciones de desempeño, engagement y	El grupo GenAI superó significativamente al grupo multimedia convencional en habilidades, engagement y	Modera-alta	NR (Not Reported)	NR (Not Reported)	NR (Not Reported)

											motivación	motivación				
28	Yu et al. (2024)	Correa del Sur	Frontiers in Education	2024	No experimental transversal	Cuantitativa; SEM	328 estudiantes universitarios usuarios de ChatGPT	ChatGPT	Educación superior	Compatibilidad, eficiencia, facilidad de uso percibida, utilidad percibida, satisfacción, intención de uso continuado	Cuestionario online vía Prolific; modelo TAM	Compatibilidad predijo facilidad de uso; eficiencia predijo utilidad; facilidad de uso y utilidad predijeron satisfacción e intención de uso continuado	Modera	NR (Not Reported)	NR (Not Reported)	NR (Not Reported)

Anexo 2

SÍNTESIS NARRATIVA DE LA EVIDENCIA

	Autor/Año	Diseño	Tipo evidencia	Enfoque cognitivo	Resultado	Contribución	Dominio	Aplicabilidad	Nivel evidencia
1	Arowosegbe et al. (2024)	No experimental	Percepción	Uso instrumental cognitivo	Mixto	Comprensión de adopción y riesgos	Pedagógico	Media	Moderado
2	BinJwair (2025)	No experimental	Adopción	Expectativas cognitivas	Positivo	Validación modelo UTAUT en IA	Pedagógico	Media	Moderado
3	Cao et al. (2025)	No experimental	Adopción	Factores cognitivos y percepción	Mixto	Influencia de innovación y percepción humana	Pedagógico	Media	Moderado
4	Dogaru et al. (2025)	No experimental	Percepción/im pacto	Motivación y metacognición	Mixto	Evidencia sobre aprendizaje y riesgos	Metacognitivo	Media	Moderado
5	George-Reyes et al. (2024)	Cuasi-experimental	Intervención	Aprendizaje y pensamiento complejo	Positivo	Mejora habilidades mediante IA	Cognitivo	Alta	Moderado

6	He y Ren (2025)	No experimental transversal	Adopción tecnológica	Aceptación, riesgo percibido, hábito e intención conductual	Mixto	Explica factores de aceptación de IA generativa en formación docente musical	Pedagógico	Media	Moderado
7	Henze et al. (2025)	Pre-post no aleatorizado	Intervención educativa	Alfabetización en IA, reflexión, andamiaje y preparación docente	Mixto	Muestra que intervenciones breves activan competencias selectivas, pero no generan mejoras amplias	Metacognitivo	Media	Moderado
8	Henze et al. (2026)	Experimental aleatorizado	Intervención comparativa	Carga cognitiva, motivación, andamiaje y aprendizaje conceptual	Mixto	Evidencia beneficios afectivo-motivacionales del chatbot GPT, sin superioridad cognitiva clara frente a Excel	Neuroeducativo	Alta	Moderado
9	Hu et al. (2026)	No experimental transversal	Modelo predictivo-mediacional	Alfabetización en IA, herramientas cognitivas, reflexión y aprendizaje matemático online	Positivo	Identifica mecanismos por los cuales el apoyo docente favorece competencias digitales y aprendizaje matemático	Cognitivo	Media	Moderado
10	Huang et al. (2025)	No experimental transversal	Modelo SEM	Cognición distribuida, presencia, inteligencia emocional y	Positivo	Muestra que los chatbots generativos pueden actuar como socios	Metacognitivo	Media	Moderado

				mentalidad de crecimiento		cognitivos en aprendizaje oral L2			
11	Jauhiaine n y Garagorr y Guerra (2024)	No experimental / intervención de aula	Intervención educativa	Personalización, motivación, ergonomía cognitiva, autorregulación	Positivo	Evidencia que ChatGPT puede personalizar dinámicamente materiales escolares y apoyar motivación y desempeño	Neuroeducativo	Alta	Moderado
12	Kazazoglu y Turun Özel (2025)	No experimental / intervención educativa	Intervención y asociación	Inteligencia emocional, autorregulación emocional, escritura EFL	Mixto	Identifica posible relación negativa entre uso frecuente de ChatGPT e inteligencia emocional, con mejora leve en adaptabilidad	Socioemocional	Media	Moderado
13	Kinskofer y Tulis (2025)	No experimental transversal	Adopción tecnológica	Motivación, appraisal de desafío/amenaza, intención conductual	Positivo	Extiende UTAUT con factores motivacionales para explicar uso de GenAI en estudiantes y docentes	Pedagógico	Media	Moderado
14	Klarin et al. (2024)	No experimental transversal	Asociación cognitiva	Funciones ejecutivas, planificación, inhibición, flexibilidad cognitiva	Mixto	Aporta evidencia sobre utilidad percibida de IA generativa en	Cognitivo	Media	Moderado

						adolescentes con mayores dificultades ejecutivas			
15	Li et al. (2025)	Mixto experimental aleatorizado	Intervención educativa	Metacognición, retroalimentación, reflexión, autorregulación	Mixto	Evidencia que la retroalimentación con LLM mejora metacognición, aunque no demuestra superioridad clara en rendimiento vocal	Metacognitivo	Alta	Moderado
16	Liew et al. (2025)	Experimental intra-sujetos contrabalanceado	Intervención comparativa	Carga cognitiva, emoción, motivación, engagement	Mixto	Muestra que el slang en chatbots ChatGPT aumenta engagement y afecto positivo, pero también dificultad/carga extrínseca	Neuroeducativo	Media	Moderado
17	Mohamed et al. (2026)	No experimental correlacional	Modelo predictivo-mediacional	Motivación académica, ansiedad GenAI, diferencias de género	Mixto	Explica patrones afectivo-motivacionales en estudiantes EFL dotados en contextos GenAI	Socioemocional	Media	Moderado
18	Neittaanmäki (2025)	No experimental transversal	Percepción y uso	Apoyo al aprendizaje, actitudes, planificación académica/profesional	Mixto	Evidencia uso extendido de chatbots generativos en estudios médicos y preocupación	Pedagógico	Media	Moderado

						n por impacto profesional			
19	Pan y Li (2025)	No experimental transversal	Modelo mediacional	Bienestar, mindset de crecimiento, apoyo docente, actitud hacia GenAI	Positivo	Identifica la actitud hacia GenAI como mediadora del bienestar en aprendizaje EFL	Socioemoci onal	Media	Moder ado
20	Salhab y Aboushi (2025)	No experimental transversal	Adopción tecnológica	Alfabetización en IA, competencias del siglo XXI, aceptación tecnológica	Positivo	Evidencia que alfabetizació n en IA y habilidades del siglo XXI predicen aceptación de IA generativa	Pedagógico	Media	Moder ado
21	Salhab y Aboushi (2026)	Mixto cuasi- experimental	Intervención educativa	Engagement cognitivo, emocional, social y conductual; carga cognitiva reducida	Positivo	Evidencia que el microlearnin g asistido por IA generativa puede mejorar engagement en entornos online	Pedagógico	Alta	Moder ado
22	Sergeeva et al. (2025)	No experimental transversal	Adopción tecnológica	Alfabetización digital, comprensión crítica y creación de contenido	Mixto	Identifica dimensiones específicas de alfabetizació n mediática digital asociadas con aceptación de GenAI	Pedagógico	Media	Moder ado

23	Susanto et al. (2026)	No experimental transversal	Modelo predictivo-mediacional	Motivación de aprendizaje, utilidad percibida, uso problemático y adopción de ChatGPT	Mixto	Muestra beneficios motivacionales del uso de ChatGPT junto con riesgos asociados al uso problemático de internet	Socioemocional	Media	Moderado
24	Wang et al. (2026)	Mixto secuencial exploratorio	Modelo explicativo de alfabetización	Alfabetización en IA, evaluación crítica, práctica, ética y cognición-práctica-evaluación	Positivo	Propone un modelo dinámico de desarrollo de alfabetización en IA mediante uso de GenAI	Metacognitivo	Alta	Moderado
25	Wang y Liu (2026)	Mixto	Adopción tecnológica	Ansiedad hacia IA, intención conductual, carga psicológica de adopción	Mixto	Explica factores psicológicos que facilitan o inhiben adopción de GenAI en lenguas extranjeras	Socioemocional	Media	Moderado
26	Westbye et al. (2026)	Cualitativo exploratorio	Interacción estudiante-IA	Prescritura, elaboración, agencia, apropiación, resistencia	Mixto	Muestra cómo un chatbot GPT-4 restringido puede apoyar planificación sin generar texto por el estudiante	Metacognitivo	Media	Moderado
27	Xu et al. (2026)	Cuasi-experimental aleatorizado	Intervención educativa	Retroalimentación interactiva, aprendizaje motor, engagement, motivación	Positivo	Evidencia beneficios de GenAI en habilidades de danza, engagement y motivación	Cognitivo	Alta	Moderado

						en educación superior			
28	Yu et al. (2024)	No experimental transversal	Adopción tecnológica	Satisfacción, facilidad de uso, utilidad percibida, continuidad de uso	Positivo	Identifica factores TAM que explican satisfacción e intención de uso continuado de ChatGPT	Pedagógico	Media	Moderado

Anexo 3

CALIDAD METODOLÓGICA / RIESGO DE SESGO

	Autor/Año	Diseño	Herramienta	Selección	Medición	Confusión	Análisis	Sesgo de reporte	Riesgo global	Justificación
1	Arowosegbe et al. (2024)	No experimental	ROBINS-I	Moderado	Moderado	Moderado	Moderado	Bajo	Moderado	Muestra limitada y autoinforme
2	BinJwair (2025)	No experimental	ROBINS-I	Moderado	Bajo	Moderado	Bajo	Bajo	Moderado	Buen análisis estadístico, pero transversal
3	Cao et al. (2025)	No experimental	ROBINS-I	Moderado	Moderado	Moderado	Moderado	Bajo	Moderado	Muestra localizada y modelo teórico
4	Dogaru et al. (2025)	No experimental	ROBINS-I	Moderado	Moderado	Moderado	Moderado	Bajo	Moderado	Muestra pequeña y autoinforme
5	George-Reyes et al. (2024)	Cuasi-experimental	ROBINS-I	Moderado	Bajo	Moderado	Bajo	Bajo	Moderado	Intervención sin aleatorización
6	He y Ren (2025)	No experimental transversal	ROBINS-I	Moderado	Moderado	Moderado	Bajo-moderado	Bajo	Moderado	Muestra no aleatoria, autoinforme y diseño transversal; análisis PLS-SEM/ANN robusto.

7	Henze et al. (2025)	Pre-post no aleatorizado	ROBINS-I	Moderado-alto	Bajo-modera do	Moderado	Moderado	Bajo	Moderado	Muestra pequeña, sin grupo control aleatorizado; instrumentos explícitos y análisis pre-post.
8	Henze et al. (2026)	Experimental aleatorizado	RoB 2	Moderado	Bajo-modera do	Moderado	Bajo-modera do	Bajo	Moderado	Asignación aleatoria, pero muestra pequeña y resultados cognitivos con ganancias marginales.
9	Hu et al. (2026)	No experimental transversal	ROBINS-I	Moderado	Moderado	Moderado	Bajo-modera do	Bajo	Moderado	Muestra amplia, pero autoinforme, una institución/contexto cultural y diseño transversal.
10	Huang et al. (2025)	No experimental transversal	ROBINS-I	Moderado	Bajo-modera do	Moderado	Bajo-modera do	Bajo	Moderado	Encuesta transversal con escalas validadas; autoinforme y causalidad no demostrable.
11	Jauhiainen y Garagorry Guerra (2024)	No experimental / intervención de aula	ROBINS-I	Moderado	Moderado	Moderado	Moderado	Bajo	Moderado	Evidencia empírica clara, pero diseño descriptivo sin aleatorización ni control robusto; posible influencia de contexto escolar y medición de percepción.
12	Kazazoglu y Turun	No experimental /	ROBINS-I	Moderado	Bajo-modera do	Moderado	Moderado	Bajo	Moderado	Uso de instrumento validado EQ-I,

	Özel (2025)	intervención educativa								pero diseño no experimental, autoinforme y ambigüedad parcial sobre tamaño muestral en los fragmentos disponibles.
13	Kinskofer y Tulis (2025)	No experimental transversal	ROBINS-I	Moderado	Moderado	Moderado	Bajo-moderado	Bajo	Moderado	Muestra amplia y análisis SEM/PLS-SEM, pero autoinforme, diseño transversal y participación voluntaria limitan inferencia causal.
14	Klarin et al. (2024)	No experimental transversal	ROBINS-I	Moderado	Bajo-moderado	Moderado	Moderado	Bajo	Moderado	Dos muestras adolescentes y uso de BRIEF-2, pero diseño transversal, autoinforme de uso/utilidad y sin causalidad.
15	Li et al. (2025)	Mixto experimental aleatorizado	MMAT	Bajo-moderado	Bajo-moderado	Moderado	Bajo-moderado	Bajo	Moderado	Asignación aleatoria, pretest-postest y triangulación cualitativa; riesgo residual por tamaño muestral limitado y combinación de intervención IA/e-learning.
16	Liew et al. (2025)	Experimental intra-sujetos contrabalanceado	ROBINS-I	Moderado	Bajo-moderado	Moderado	Bajo-moderado	Bajo	Moderado	Diseño experimental contrabalanceado, pero muestra específica y

										posible efecto de laboratorio; resultados afectivos/cognitivos autoinformados.
17	Mohamed et al. (2026)	No experimental correlacional	ROBINS-I	Moderado	Moderado	Moderado	Bajo-moderado	Bajo	Moderado	Diseño correlacional con SEM; muestra intencional de estudiantes dotados y ausencia de causalidad experimental.
18	Neittaanmäki (2025)	No experimental transversal	ROBINS-I	Moderado	Moderado	Moderado	Moderado	Bajo	Moderado	Encuesta voluntaria en una universidad; autoinforme y posible sesgo de respuesta; análisis estadístico explícito.
19	Pan y Li (2025)	No experimental transversal	ROBINS-I	Moderado	Moderado	Moderado	Bajo-moderado	Bajo	Moderado	Modelo mediacional con muestra amplia, pero transversal y basado en autoinforme.
20	Salhab y Aboushi (2025)	No experimental transversal	ROBINS-I	Bajo-moderado	Moderado	Moderado	Moderado	Bajo	Moderado	Muestra aleatoria reportada, pero diseño transversal y medición por escalas autoinformadas limitan inferencia causal.
21	Salhab y Aboushi (2026)	Mixto cuasi-experimental	MMAT	Moderado	Moderado	Moderado	Bajo-moderado	Bajo	Moderado	Diseño mixto con cuasi-experimento e

										entrevistas, pero muestra pequeña y sin aleatorización explícita reportada en fragmentos.
2 2	Sergeeva et al. (2025)	No experimental transversal	ROBINS-I	Moderado	Moderado	Moderado	Bajo-moderado	Bajo	Moderado	Muestra universitaria única, autoinforme y diseño transversal; SEM fortalece el análisis, pero no permite causalidad.
2 3	Susanto et al. (2026)	No experimental transversal	ROBINS-I	Moderado	Moderado	Moderado	Bajo-moderado	Bajo	Moderado	Modelo TAM-SDT con PLS-SEM y validación psicométrica; diseño transversal y N no visible en fragmentos disponibles.
2 4	Wang et al. (2026)	Mixto secuencial exploratorio	MMAT	Moderado	Bajo-moderado	Moderado	Bajo-moderado	Bajo	Moderado	Diseño mixto con entrevistas, encuesta amplia y SEM; persiste riesgo por autoinforme y ausencia de intervención experimental.
2 5	Wang y Liu (2026)	Mixto	MMAT	Moderado	Moderado	Moderado	Bajo-moderado	Bajo	Moderado	Integra SEM y análisis de sentimiento, pero mantiene riesgo por autoinforme, diseño no experimental y posible sesgo de datos online.

26	Westbye et al. (2026)	Cualitativo exploratorio	JB1	Moderado	Bajo-moderado	Moderado	Bajo-moderado	Bajo	Moderado	Análisis detallado de diálogos reales, pero muestra pequeña, muestreo por conveniencia y contexto único.
27	Xu et al. (2026)	Cuasi-experimental aleatorizado	RoB 2	Moderado	Bajo-moderado	Moderado	Bajo-moderado	Bajo	Moderado	Asignación aleatoria y comparación con control, pero muestra limitada y contexto específico de danza universitaria.
28	Yu et al. (2024)	No experimental transversal	ROBINS-I	Moderado	Moderado	Moderado	Bajo-moderado	Bajo	Moderado	Encuesta transversal con SEM; uso de autoinforme y sin inferencia causal experimental.

Capítulo 3

La calidad del sueño y su influencia en el rendimiento académico de los estudiantes de secundaria

Jenny Margot Cáceres Soto³

Susana Pilar Heredia García⁴

Resumen

El sueño en la adolescencia ha sido identificado como un factor relevante para el funcionamiento cognitivo y el desempeño académico; sin embargo, la evidencia empírica reciente muestra resultados inconsistentes respecto a la naturaleza de esta relación. En este contexto, el presente estudio tuvo como objetivo analizar la relación entre la calidad del sueño y el rendimiento académico en estudiantes de educación secundaria.

Se empleó un diseño cuantitativo, no experimental, de corte transversal y alcance correlacional. La muestra estuvo conformada por 64 estudiantes de sexo femenino. La calidad del sueño fue evaluada mediante un cuestionario adaptado basado en el Pittsburgh Sleep Quality Index, mientras que el rendimiento académico se obtuvo a partir de registros institucionales. El análisis de datos incluyó estadísticos descriptivos y el coeficiente de correlación de Spearman. Los resultados mostraron un predominio de niveles medios de calidad del sueño y una distribución heterogénea del rendimiento académico. En el análisis inferencial, no se encontraron relaciones estadísticamente significativas entre la calidad del sueño —ni a nivel

³ Especialista en neuroeducación por la Universidad César Vallejo, Perú. E-mail: jmcaceres@ucvvirtual.edu.pe
ORCID: <https://orcid.org/0009-0006-8950-8071>

⁴ Especialista en neuroeducación por la Universidad César Vallejo, Perú. E-mail: spheredia@ucvvirtual.edu.pe
ORCID: <https://orcid.org/0009-0002-9609-9844>

global ni en sus dimensiones— y el rendimiento académico, observándose únicamente asociaciones débiles.

Se concluye que la calidad del sueño no se relaciona de manera directa con el rendimiento académico en la muestra estudiada, lo que sugiere la necesidad de considerar modelos más complejos que incorporen variables mediadoras y contextuales para comprender esta relación en estudiantes de educación secundaria.

Palabras clave: calidad del sueño, rendimiento académico, estudiantes de secundaria, adolescencia, neuroeducación.

Abstract

Adolescent sleep has been identified as a relevant factor for cognitive functioning and academic performance; however, recent empirical evidence shows inconsistent findings regarding the nature of this relationship. In this context, the present study aimed to analyze the relationship between sleep quality and academic performance in secondary school students.

A quantitative, non-experimental, cross-sectional, and correlational design was employed. The sample consisted of 64 female students. Sleep quality was assessed using an adapted questionnaire based on the Pittsburgh Sleep Quality Index, while academic performance was obtained from institutional records. Data analysis included descriptive statistics and Spearman's rank correlation coefficient.

The results showed a predominance of medium levels of sleep quality and a heterogeneous distribution of academic performance. In the inferential analysis, no statistically significant relationships were found between sleep quality—neither at a global level nor across its dimensions—and academic performance, with only weak associations observed.

It is concluded that sleep quality is not directly associated with academic performance in the studied sample, suggesting the need to consider more complex models incorporating mediating and contextual variables to better understand this relationship in secondary school students.

Keywords: sleep quality, academic performance, secondary school students, adolescence, neuroeducation.

1. Introducción

El sueño en la adolescencia se ha consolidado como un determinante crítico del funcionamiento cognitivo, la salud mental y el desempeño escolar. Diversos estudios han documentado que los adolescentes presentan con frecuencia patrones de sueño insuficientes o irregulares, asociados a cambios biológicos propios del desarrollo, así como a demandas académicas y sociales crecientes. En este contexto, el sueño deja de ser únicamente un proceso fisiológico para convertirse en un factor relevante en el ámbito educativo, en tanto influye directamente en la capacidad de los estudiantes para sostener la atención, procesar información y regular su conducta en el entorno escolar (Fonseca & Genzel, 2020).

Desde una perspectiva educativa, la calidad del sueño adquiere especial relevancia al vincularse con el rendimiento académico y el funcionamiento escolar general. Investigaciones recientes en estudiantes de educación secundaria han mostrado que mejores patrones de sueño —ya sea en términos de duración, regularidad o calidad— se asocian con un mejor desempeño académico, así como con mayores niveles de participación y adaptación escolar (Sidek et al., 2024). En consecuencia, el estudio del sueño en adolescentes no solo responde a una preocupación de salud, sino también a un problema educativo con implicancias directas en los procesos de aprendizaje.

El vínculo entre el sueño y el aprendizaje encuentra sustento en mecanismos neurocognitivos bien documentados. Durante el sueño, especialmente en sus fases profundas, se produce la consolidación de la memoria, proceso mediante el cual la información adquirida durante el día es reorganizada y estabilizada en el cerebro. Asimismo, el sueño favorece funciones cognitivas clave como la atención, el procesamiento de la información y la regulación emocional, todas

ellas fundamentales para el desempeño académico (Machado et al., 2021). Desde el campo de la neuroeducación, se ha destacado que el sueño contribuye a la plasticidad neuronal, facilitando la adaptación del cerebro a nuevas experiencias de aprendizaje y optimizando la adquisición de conocimientos (Albares, 2023).

En términos empíricos, la literatura reciente en población de educación secundaria ha documentado de manera consistente una asociación entre el sueño y el rendimiento académico. Diversos estudios han encontrado que una mayor duración del sueño o una mejor calidad del mismo se relacionan con calificaciones más altas y un mejor funcionamiento escolar, particularmente en áreas como matemáticas y ciencias (Wang et al., 2025; Alexopoulou et al., 2024; Naz et al., 2024). De forma complementaria, se ha observado que el déficit de sueño, así como la presencia de somnolencia diurna, se asocian con un menor rendimiento académico, mayor riesgo de fracaso escolar y dificultades en el funcionamiento cotidiano dentro del entorno educativo (Pérez-Chada et al., 2023).

No obstante, este patrón general de asociación no es uniforme en todos los estudios. Algunas investigaciones recientes han reportado relaciones débiles o no significativas entre la calidad del sueño y el rendimiento académico, sugiriendo que dicha relación puede estar modulada por otros factores individuales y contextuales (Utomo et al., 2025; Ahmadi & Omidvar, 2022). En este sentido, se ha propuesto que variables como la somnolencia diurna, el estrés académico, la regulación emocional y el uso de dispositivos electrónicos pueden actuar como mediadores o moderadores del vínculo entre el sueño y el desempeño escolar (Qu et al., 2023; Koike et al., 2025). Asimismo, evidencia reciente sugiere que la relación entre el sueño y el rendimiento académico no siempre es lineal, pudiendo existir rangos óptimos de sueño a partir de los cuales los beneficios se estabilizan o disminuyen (Wang et al., 2025).

Estas inconsistencias ponen de manifiesto la complejidad del fenómeno y evidencian limitaciones importantes en la literatura existente. En primer lugar, una gran parte de los

estudios se ha centrado en la duración del sueño, relegando el análisis de la calidad del sueño como un constructo multidimensional que incluye aspectos como la latencia, la eficiencia, las perturbaciones y la disfunción diurna (Fonseca & Genzel, 2020). En segundo lugar, predominan diseños de tipo transversal y el uso de medidas autorreportadas, lo que limita la capacidad de establecer relaciones más precisas y reduce la validez de los hallazgos. En tercer lugar, existe una notable heterogeneidad en la forma de medir el rendimiento académico, lo que dificulta la comparación entre estudios y la construcción de conclusiones consistentes. Finalmente, la evidencia empírica reciente se encuentra concentrada principalmente en determinados contextos geográficos y educativos, lo que limita su generalización a otras poblaciones.

En conjunto, aunque la literatura sugiere una asociación entre el sueño y el rendimiento académico, persisten limitaciones relevantes en la forma en que este fenómeno ha sido conceptualizado, medido y analizado. En particular, se observa una insuficiente exploración de la calidad del sueño desde un enfoque multidimensional, así como una escasez de estudios que integren esta perspectiva con medidas objetivas del rendimiento académico en contextos educativos específicos. Estas limitaciones dificultan establecer conclusiones consistentes sobre la naturaleza y magnitud de la relación entre ambas variables en estudiantes de educación secundaria.

En este contexto, el presente estudio tiene como objetivo analizar la relación entre la calidad del sueño y el rendimiento académico en estudiantes de educación secundaria, considerando un enfoque multidimensional del sueño y utilizando registros académicos institucionales como indicador del desempeño escolar. De este modo, la investigación busca aportar evidencia empírica que contribuya a una comprensión más precisa del fenómeno en contextos educativos reales, así como a la discusión contemporánea sobre los factores que inciden en el aprendizaje y el rendimiento académico en la adolescencia.

2. Metodología

2.1 Diseño del estudio

El presente estudio adoptó un **diseño cuantitativo, no experimental, de corte transversal y alcance correlacional**. Esto con el objetivo de analizar la asociación entre la calidad del sueño y el rendimiento académico en estudiantes de educación secundaria. La investigación se desarrolló en condiciones naturales, sin manipulación de las variables, lo que permitió observar los patrones habituales de sueño y su relación con el desempeño académico en el contexto escolar.

Dado el objetivo planteado, el diseño correlacional resultó adecuado para examinar la magnitud y dirección de la relación entre las variables de estudio. No obstante, es importante precisar que este tipo de diseño no permite establecer relaciones causales; en consecuencia, los resultados deben interpretarse en términos de asociación estadística y no como evidencia de influencia causal directa (Ato et al., 2013; Creswell & Creswell, 2018).

2.2 Participantes y contexto

El estudio se llevó a cabo en una institución educativa de nivel secundario y se centró en estudiantes matriculadas en el tercer grado. La población estuvo conformada por **84 estudiantes de sexo femenino**, en concordancia con la naturaleza de la institución, que cuenta exclusivamente con alumnado femenino.

Para la selección de la muestra se empleó un **muestreo probabilístico para poblaciones finitas**, con un nivel de confianza del 90%, un margen de error del 5% y una proporción esperada del 50%, lo que permitió determinar un tamaño muestral de **64 estudiantes** (Hernández-Sampieri & Mendoza, 2023).

Los criterios de inclusión consideraron la matrícula activa en el periodo académico correspondiente, la asistencia al momento de la recolección de datos y la disponibilidad de

registros de rendimiento académico. Se excluyeron aquellos casos con información incompleta en los instrumentos aplicados o sin datos académicos registrados.

2.3 Variables y operacionalización

Calidad del sueño

La calidad del sueño fue conceptualizada como un **constructo multidimensional** que integra componentes subjetivos y objetivos del descanso nocturno. Siguiendo el marco del Pittsburgh Sleep Quality Index propuesto por Buysse et al. (1989), se consideraron siete dimensiones: calidad subjetiva del sueño, latencia del sueño, duración del sueño, eficiencia habitual del sueño, perturbaciones del sueño, uso de medicación hipnótica y disfunción diurna.

Cada una de estas dimensiones fue categorizada en tres niveles (bajo, medio y alto), de acuerdo con los criterios establecidos en el instrumento adaptado. Estas categorías fueron tratadas como variables de tipo ordinal, en coherencia con la naturaleza de la medición.

Rendimiento académico

El rendimiento académico se definió como el nivel de logro de aprendizaje alcanzado por las estudiantes dentro del sistema educativo formal. Su operacionalización se realizó a partir de los registros institucionales correspondientes al primer bimestre académico, siguiendo la escala de evaluación nacional: AD (logro destacado), A (logro esperado), B (en proceso) y C (en inicio) (Ministerio de Educación del Perú, 2017).

Dado su carácter categórico jerárquico, el rendimiento académico fue tratado como una variable ordinal, adecuada para análisis estadísticos no paramétricos.

2.4 Instrumentos y medición

Cuestionario de calidad del sueño

La calidad del sueño fue evaluada mediante un **cuestionario de autorreporte adaptado a partir del Pittsburgh Sleep Quality Index (PSQI)** (Buysse et al., 1989). La versión utilizada

estuvo compuesta por **12 ítems**, diseñados para medir las siete dimensiones del constructo en una población de estudiantes de educación secundaria.

El proceso de adaptación implicó ajustes en el lenguaje y en la formulación de los ítems para garantizar su comprensión por parte de adolescentes, manteniendo la coherencia conceptual con la estructura original del instrumento.

Validez de contenido

La validez de contenido del instrumento fue establecida mediante **juicio de expertos**, quienes evaluaron la pertinencia, claridad y representatividad de los ítems en relación con el constructo de calidad del sueño. Este procedimiento permitió asegurar la adecuación del instrumento al contexto educativo y a la población objetivo (Escobar-Pérez & Cuervo-Martínez, 2008).

Consideraciones sobre confiabilidad

Si bien el instrumento fue adaptado y sometido a validación por expertos, no se reportaron indicadores de consistencia interna (por ejemplo, alfa de Cronbach). Esta ausencia constituye una limitación metodológica que debe ser considerada en la interpretación de los resultados.

Medición del rendimiento académico

El rendimiento académico fue obtenido a partir de **registros oficiales institucionales**, correspondientes al desempeño de las estudiantes durante el primer bimestre. El uso de estos registros proporciona una medida ecológicamente válida del desempeño escolar, al reflejar evaluaciones reales dentro del sistema educativo (Richardson et al., 2012).

2.5 Procedimiento

Previo a la recolección de datos, se obtuvo la autorización de la institución educativa para la ejecución del estudio. La aplicación del cuestionario se realizó en el aula, en un entorno controlado y bajo condiciones estandarizadas, con la supervisión del equipo investigador.

La participación fue voluntaria, y las estudiantes fueron informadas sobre los objetivos del estudio, así como sobre la confidencialidad de la información proporcionada. Posteriormente,

los datos de rendimiento académico fueron recopilados a partir de los registros institucionales y vinculados a las respuestas del cuestionario mediante identificadores anónimos, garantizando la protección de la identidad de las participantes.

2.6 Análisis de datos

El procesamiento de los datos se realizó mediante el software **IBM SPSS Statistics (versión 30)**. En una primera etapa, se llevó a cabo la codificación, depuración y verificación de la base de datos para asegurar su consistencia e integridad.

Se calcularon **estadísticos descriptivos**, incluyendo frecuencias y porcentajes, con el fin de caracterizar los niveles de calidad del sueño, sus dimensiones y el rendimiento académico.

Para evaluar la distribución de las variables, se aplicó la **prueba de normalidad de Kolmogorov–Smirnov**, cuyos resultados indicaron que las variables no seguían una distribución normal ($p < .05$). En consecuencia, se optó por el uso de técnicas estadísticas no paramétricas (Ghasemi & Zahediasl, 2012).

El análisis inferencial se realizó mediante el **coeficiente de correlación de rangos de Spearman (rho)**, con el propósito de examinar la relación entre la calidad del sueño (global y por dimensiones) y el rendimiento académico. El nivel de significancia se estableció en $\alpha = .05$ (**bilateral**), y la magnitud de las correlaciones se interpretó de acuerdo con criterios convencionales de tamaño de efecto (Cohen, 1988).

2.7 Consideraciones éticas

La investigación se desarrolló en concordancia con los principios éticos fundamentales de **beneficencia, no maleficencia y respeto por la autonomía** (Beauchamp & Childress, 2019).

Se garantizó la participación voluntaria de las estudiantes, así como el respeto por su confidencialidad y anonimato. La información recolectada fue utilizada exclusivamente con fines académicos y de investigación, sin comprometer la integridad ni el bienestar de las participantes.

Asimismo, el estudio se llevó a cabo bajo criterios de probidad académica y responsabilidad científica, asegurando el adecuado manejo de los datos y el cumplimiento de las normativas éticas institucionales.

3. Resultados

3.1 Estadísticos descriptivos

La distribución de la calidad del sueño mostró un predominio del nivel medio (54.7%), seguido del nivel alto (28.1%) y, en menor proporción, del nivel bajo (17.2%). Este patrón sugiere que la mayoría de las estudiantes percibe una calidad de sueño moderada, con una proporción no menor que reporta condiciones óptimas de descanso.

A nivel de dimensiones específicas, se observó un comportamiento heterogéneo. La calidad subjetiva del sueño se concentró principalmente en el nivel medio (70.3%), mientras que la latencia del sueño también presentó un predominio del nivel medio (42.2%), aunque con una distribución más dispersa entre niveles bajo (25.0%) y alto (32.8%). En cuanto a la duración del sueño, se identificó un patrón relevante: el 43.8% de las estudiantes reportó niveles bajos, lo que indica una proporción considerable que no alcanza las horas de sueño recomendadas.

Respecto a la eficiencia del sueño, la mayoría de las participantes se ubicó en el nivel medio (68.8%), sugiriendo hábitos de sueño relativamente consistentes, aunque no óptimos. Las perturbaciones del sueño mostraron una distribución más equilibrada entre niveles bajo (31.3%), medio (42.2%) y alto (26.6%), evidenciando variabilidad en la estabilidad del sueño nocturno.

En contraste, el uso de medicación hipnótica fue prácticamente inexistente, con un 90.6% de las estudiantes ubicadas en el nivel bajo, lo que indica una baja dependencia de ayudas farmacológicas para conciliar el sueño. Finalmente, la disfunción diurna presentó un predominio del nivel medio (50.0%), seguida del nivel bajo (43.8%), lo que sugiere que, en general, las consecuencias diurnas del sueño son moderadas en esta población.

En cuanto al rendimiento académico, se observó una distribución heterogénea, con predominio del nivel C (34.4%), seguido del nivel A (29.7%), mientras que los niveles B (18.8%) y AD (17.2%) presentaron menores frecuencias. Este patrón indica una concentración importante de estudiantes en niveles de desempeño académico inicial o intermedio.

Tabla 1. Estadísticos descriptivos de las variables del estudio

Variable	Categoría	n	%
Calidad del sueño	Bajo	11	17.2
	Medio	35	54.7
	Alto	18	28.1
Calidad subjetiva	Bajo	9	14.1
	Medio	45	70.3
	Alto	10	15.6
Latencia del sueño	Bajo	16	25.0
	Medio	27	42.2
	Alto	21	32.8
Duración del sueño	Bajo	28	43.8
	Medio	32	50.0
	Alto	4	6.3
Eficiencia del sueño	Bajo	7	10.9
	Medio	44	68.8
	Alto	13	20.3
Perturbaciones	Bajo	20	31.3
	Medio	27	42.2
	Alto	17	26.6
Medicación hipnótica	Bajo	58	90.6
	Medio	6	9.4

	Alto	0	0.0
Disfunción diurna	Bajo	28	43.8
	Medio	32	50.0
	Alto	4	6.3
Rendimiento académico	C	22	34.4
	B	12	18.8
	A	19	29.7
	AD	11	17.2

Fuente: elaboración propia

3.2 Evaluación de supuestos

Las pruebas de normalidad (Kolmogorov–Smirnov) indicaron que tanto la calidad del sueño como el rendimiento académico presentan distribuciones significativamente diferentes de la normal ($p < .05$). En consecuencia, se optó por el uso de pruebas no paramétricas para el análisis inferencial, específicamente el coeficiente de correlación de Spearman.

3.3 Análisis inferencial

El análisis de correlación de Spearman reveló que no existe una relación estadísticamente significativa entre la calidad del sueño y el rendimiento académico ($\rho = 0.115$, $p = .366$). La magnitud del coeficiente indica una asociación positiva muy débil, cercana a cero.

De manera consistente, el análisis por dimensiones mostró patrones similares. Las correlaciones entre las dimensiones de la calidad del sueño y el rendimiento académico fueron, en su mayoría, de magnitud débil y no significativas. En particular, la calidad subjetiva del sueño ($\rho = 0.205$, $p = .104$), la duración del sueño ($\rho = 0.211$, $p = .095$) y la eficiencia del sueño ($\rho = 0.227$, $p = .071$) presentaron asociaciones positivas débiles, sin alcanzar niveles de significación estadística.

Por su parte, la latencia del sueño ($\rho = 0.030$, $p = .813$) y la disfunción diurna ($\rho = 0.139$, $p = .272$) evidenciaron correlaciones prácticamente nulas. En contraste, las perturbaciones del sueño ($\rho = -0.051$, $p = .691$) y el uso de medicación hipnótica ($\rho = -0.128$, $p = .312$) mostraron asociaciones negativas muy débiles, igualmente no significativas.

En conjunto, estos resultados indican la ausencia de relaciones estadísticamente significativas entre la calidad del sueño —tanto a nivel global como en sus dimensiones— y el rendimiento académico en la muestra analizada.

Tabla 2. Correlaciones de Spearman entre calidad del sueño y rendimiento académico

Variable	ρ	p
Calidad del sueño (global)	0.115	0.366
Calidad subjetiva	0.205	0.104
Latencia del sueño	0.030	0.813
Duración del sueño	0.211	0.095
Eficiencia del sueño	0.227	0.071
Perturbaciones	-0.051	0.691
Medicación hipnótica	-0.128	0.312
Disfunción diurna	0.139	0.272

4. Discusión

Los resultados del presente estudio evidenciaron que la mayoría de las estudiantes reportó una calidad del sueño de nivel medio, mientras que el rendimiento académico mostró una distribución heterogénea con predominio del nivel C. En términos inferenciales, no se encontraron correlaciones estadísticamente significativas entre la calidad del sueño —ni a nivel global ni en sus dimensiones específicas— y el rendimiento académico, observándose únicamente asociaciones de magnitud débil.

En relación con el hallazgo principal, la ausencia de una relación estadísticamente significativa entre la calidad del sueño global y el rendimiento académico sugiere que, en esta muestra, no se observa una asociación directa entre ambas variables. Este resultado debe interpretarse con cautela, ya que no implica que el sueño carezca de relevancia en el desempeño escolar, sino que su influencia podría no manifestarse de manera lineal o inmediata en indicadores como las calificaciones. En este sentido, estudios recientes han reportado resultados similares, en los que no se identifican asociaciones significativas entre variables de sueño y rendimiento académico, lo que refuerza la idea de que esta relación puede ser más compleja de lo que tradicionalmente se ha asumido (Utomo et al., 2025; Ahmadi & Omidvar, 2022).

Al analizar las dimensiones específicas de la calidad del sueño, se observaron patrones heterogéneos que aportan matices relevantes. Por un lado, dimensiones como la calidad subjetiva del sueño, la duración y la eficiencia mostraron asociaciones positivas débiles con el rendimiento académico, lo que sugiere una tendencia consistente con la literatura que reporta efectos positivos del sueño sobre el desempeño escolar, aunque sin alcanzar niveles de significación estadística en este estudio. Por otro lado, dimensiones como la latencia, las perturbaciones del sueño, el uso de medicación hipnótica y la disfunción diurna evidenciaron relaciones prácticamente nulas o levemente negativas, lo que indica que no todas las dimensiones del sueño contribuyen de manera uniforme al rendimiento académico.

Desde una perspectiva teórica, estos hallazgos pueden interpretarse a la luz de la teoría de la consolidación de la memoria dependiente del sueño, la cual plantea que el sueño favorece procesos cognitivos intermedios como la memoria, la atención y la regulación emocional, fundamentales para el aprendizaje. Sin embargo, el rendimiento académico constituye un resultado distal, acumulativo y multifactorial, influido por variables adicionales como la motivación, el contexto escolar, las estrategias de estudio y las condiciones socioeducativas. En este marco, es plausible que los beneficios del sueño se expresen en niveles cognitivos más

inmediatos sin traducirse necesariamente en una relación directa y observable con las calificaciones. Esta interpretación permite comprender por qué la literatura no siempre encuentra asociaciones lineales consistentes entre sueño y rendimiento académico (Fonseca & Genzel, 2020).

Al contrastar estos resultados con estudios previos, se observan coincidencias parciales. Investigaciones recientes han reportado asociaciones positivas entre el sueño y el rendimiento académico, indicando que mejores condiciones de sueño se relacionan con un mejor desempeño escolar. En el presente estudio, si bien la dirección de las asociaciones fue similar —predominantemente positiva—, la magnitud de estas no alcanzó significación estadística, lo que sugiere una relación más débil en esta muestra. Asimismo, existen estudios que tampoco han encontrado asociaciones directas, lo que coincide con los resultados obtenidos y refuerza la idea de que el vínculo entre ambas variables puede depender de factores intermedios no considerados en el presente estudio.

En este sentido, una explicación plausible para las discrepancias observadas en la literatura radica en la presencia de variables mediadoras o moderadoras. Factores como la somnolencia diurna, el estrés académico, el uso de dispositivos electrónicos y la organización del tiempo de estudio han sido identificados como elementos que pueden influir tanto en la calidad del sueño como en el rendimiento académico. La ausencia de medición de estos factores en el presente estudio podría haber contribuido a la dilución de la relación directa entre las variables analizadas.

En términos de implicaciones, los resultados sugieren que la relación entre el sueño y el rendimiento académico en estudiantes de secundaria no debe asumirse como directa ni uniforme, sino como un fenómeno complejo que requiere ser abordado desde modelos más integrales. Entre las principales limitaciones del estudio se encuentran su diseño transversal, el tamaño de la muestra y el uso de un instrumento adaptado para medir la calidad del sueño. En

futuras investigaciones, se recomienda incorporar diseños longitudinales, muestras más amplias y modelos multivariados que permitan analizar el papel de variables mediadoras en la relación entre sueño y rendimiento académico.

Referencias

- Ahmadi, Z., & Omidvar, S. (2022). The quality of sleep and daytime sleepiness and their association with quality of school life and school achievement among students. *Journal of Education and Health Promotion*, 11. https://doi.org/10.4103/jehp.jehp_22_22
- Albares, N. (2023). [Verificar título exacto].
- Alexopoulou, C., Fountoulaki, M., Papavasileiou, A., & Kondili, E. (2024). Sleep habits, academic performance and health behaviors of adolescents in Southern Greece. *Healthcare*, 12. <https://doi.org/10.3390/healthcare12070775>
- Ato, M., López, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038–1059.
- Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (8th ed.). Oxford University Press.
- Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research*, 76(2), 192–203.
- Buyse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research*, 28(2), 193–213.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6, 27–36.

- Fonseca, A., & Genzel, L. (2020). Sleep and academic performance: Considering amount, quality and timing. *Current Opinion in Behavioral Sciences*, 33, 65–71. <https://doi.org/10.1016/j.cobeha.2019.12.008>
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486–489.
- Hernández-Sampieri, R., & Mendoza, C. P. (2023). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta* (verificar edición exacta).
- Koike, C., Nestor, B., Baumer, A., & Kossowsky, J. (2025). The indirect effect of sleep quality on stress-related psychosocial outcomes in adolescents: An investigation across genders. *Frontiers in Psychology*, 16. <https://doi.org/10.3389/fpsyg.2025.1512416>
- Machado, et al. (2021). [Verificar referencia completa exacta].
- Ministerio de Educación del Perú. (2017). *Currículo Nacional de la Educación Básica*. Ministerio de Educación del Perú.
- Naz, I., Naz, F., Bardaie, N., Aslam, M., Chaudhary, M., Abid, M., & Habib, M. (2024). Physical activity and sleep patterns as determinants of academic performance in secondary school teenagers: A descriptive cross-sectional study. *Insights-Journal of Health and Rehabilitation*. <https://doi.org/10.71000/ijhr146>
- Pérez-Chada, D., Bioch, S., Schönfeld, D., Gozal, D., & Pérez-Lloret, S. (2023). Screen use, sleep duration, daytime somnolence, and academic failure in school-aged adolescents. *PLOS ONE*, 18. <https://doi.org/10.1371/journal.pone.0281379>
- Qu, G., Liu, H., Han, T., Zhang, H., Sun, L., Qin, Q., Chen, M., Zhou, X., & Sun, Y. (2023). Association between adverse childhood experiences and sleep quality, emotional and behavioral problems and academic achievement of children and adolescents.

European Child & Adolescent Psychiatry, 1–12. <https://doi.org/10.1007/s00787-023-02185-w>

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387.

Sidek, N., Kamal, A., & Rahman, Z. (2024). Sleep quality levels and academic performance among secondary school. *ACPES Journal of Physical Education, Sport, and Health (AJPESH)*. <https://doi.org/10.15294/ajpesh.v4i1.12102>

Utomo, R., Rini, A., & Matulesy, A. (2025). The relationship between sleep quality and emotion regulation on academic performance of high school students. *Journal of Scientific Research, Education, and Technology (JSRET)*. <https://doi.org/10.58526/jsret.v4i2.781>

Wang, M., Chen, Z., Han, N., & Yao, H. (2025). Sleep duration and subject-specific academic performance among adolescents in China. *npj Science of Learning*, 10. <https://doi.org/10.1038/s41539-025-00361-y>

Capítulo 4

La neuroeducación y su asociación en el aprendizaje multisensorial en los estudiantes del nivel inicial

Socorro Lecarnaqué Sánchez ⁵

RESUMEN

El estudio examinó la asociación entre el nivel de uso de estrategias neuroeducativas y el aprendizaje multisensorial en niños de tres años de educación inicial en Piura, Perú, durante el año 2025, con el objetivo específico de identificar de manera diferenciada la asociación entre neuroeducación y los canales de aprendizaje visual, kinestésico y auditivo. Se adoptó un enfoque cuantitativo con diseño no experimental, correlacional y transversal. La muestra fue censal y estuvo conformada por 18 estudiantes de un aula de educación inicial. Los datos se recogieron mediante dos listas de cotejo con ítems dicotómicos, validadas por juicio de expertos y con coeficientes KR-20 de 0.86 y 0.85 respectivamente, y se analizaron mediante la prueba Chi-cuadrado de Pearson complementada con triangulación estadística. Los resultados revelaron una asociación estadísticamente significativa y de magnitud muy grande entre neuroeducación y aprendizaje multisensorial global ($\chi^2(2) = 14.456$, $p = .001$, $V = .896$), así como una asociación significativa de magnitud grande entre neuroeducación y aprendizaje visual ($\chi^2(2) = 6.019$, $p = .049$, $V = .578$). El canal kinestésico no alcanzó significancia estadística global ($\chi^2(2) = 4.760$, $p = .093$) pero mostró asociación lineal significativa ($p = .037$) y efecto grande ($V = .514$). El canal auditivo no evidenció asociación significativa ($\chi^2(2) = 2.340$, $p = .310$, $V = .360$). Se concluye que el uso sistemático de estrategias neuroeducativas

⁵ Docente universitaria de Universidad César Vallejo, Perú. E-mail: jdlecarneque@ucvvirtual.edu.pe, ORCID: <https://orcid.org/0009-0001-7265-5925>

se asocia con niveles más altos de desempeño multisensorial en la educación inicial, con mayor consistencia en el canal visual, y que los canales kinestésico y auditivo requieren condiciones de investigación longitudinal para capturar sus efectos con mayor precisión.

Palabras clave: neuroeducación; aprendizaje multisensorial; educación inicial; neuroplasticidad; desarrollo sensorial temprano.

INTRODUCCIÓN

La educación inicial representa una ventana temporal de excepcional neuroplasticidad en la que las experiencias pedagógicas ejercen un impacto desproporcionado sobre el desarrollo cognitivo, lingüístico y socioemocional. Durante los primeros años de vida, el cerebro infantil forma conexiones neuronales a una velocidad sin precedentes, lo que convierte este período en el momento óptimo para intervenciones educativas que aprovechen la capacidad del sistema nervioso de reorganizarse en respuesta a la experiencia. La neuroeducación ha emergido precisamente como el campo interdisciplinario que traduce este conocimiento neurocientífico en principios y estrategias pedagógicas aplicables al aula (Shams & Seitz, 2008; Walsh et al., 2024).

La pertinencia de este campo se ha intensificado en el período post-pandémico, durante el cual millones de niños en edad preescolar fueron privados de experiencias de exploración sensorial y aprendizaje activo presencial, generando déficits documentados en el desarrollo motor, perceptivo y cognitivo temprano que hacen especialmente urgente la implementación de estrategias pedagógicas fundamentadas en el conocimiento del cerebro (Burbano-Enríquez et al., 2025; Walsh et al., 2024).

La neuroeducación, entendida como la integración de hallazgos de la neurociencia, la psicología cognitiva y las ciencias de la educación para mejorar los procesos de enseñanza-aprendizaje a partir del conocimiento sobre el funcionamiento cerebral, encuentra en la

educación inicial su campo de aplicación más fértil (Tokuhami-Espinosa, 2020; Mora, 2021). En este nivel educativo, sus principios se concretan en estrategias que respetan y aprovechan la neuroplasticidad, los diferentes estilos de aprendizaje y los procesos cognitivos propios de los tres a seis años, con efectos documentados sobre la atención, la memoria y las funciones ejecutivas en niños de edad preescolar (Pineda et al., 2025; Burbano-Enríquez et al., 2025). Uno de los enfoques más coherentes con estos principios es el aprendizaje multisensorial, que refiere al procesamiento de información a través de la activación simultánea o complementaria de múltiples canales sensoriales — visual, auditivo, kinestésico y táctil — durante las actividades pedagógicas. Desde una perspectiva neurocientífica, la evidencia indica que el cerebro humano está evolutivamente preparado para aprender de manera más eficaz en condiciones multisensoriales que en condiciones unisensoriales, porque la convergencia de señales de diferentes modalidades genera representaciones neurales más robustas y duraderas (Shams & Seitz, 2008; Murray & Shams, 2023). Esta ventaja es especialmente pronunciada en la primera infancia, dado el estado de alta plasticidad del sistema nervioso en este período. Lejos de constituir un constructo unitario, el aprendizaje multisensorial opera de manera diferencial según el canal sensorial predominante — visual, kinestésico o auditivo —, distinción que tiene implicaciones directas tanto para el diseño instruccional como para la evaluación de su efectividad (Gardner, 1983, 2006; Dunn & Dunn, 1992).

La evidencia empírica acumulada en la última década respalda de manera consistente los beneficios de integrar estrategias neuroeducativas y multisensoriales en la educación inicial. Estudios recientes han documentado efectos positivos sobre la atención sostenida, la memoria de trabajo y las funciones ejecutivas en niños de dos a cinco años cuando se implementan entornos pedagógicos que estimulan simultáneamente varios canales sensoriales (Sadreimova, 2024; Morales & Rivadeneira, 2025; Pineda et al., 2025; Bolaños, 2025). En el plano de la adquisición de habilidades específicas, el uso de estrategias multisensoriales contextualizadas

ha mostrado efectos favorables sobre la lectura emergente y el pensamiento lógico-matemático en educación inicial (Wulandari & Pujaningsih, 2025; Illapa, 2024). En el ámbito socioemocional, intervenciones basadas en experiencias sensoriales ricas e interactivas han reportado mejoras en el comportamiento prosocial y la regulación emocional en contextos preescolares (Harahap et al., 2025; Centeno, 2024). La convergencia de esta evidencia apunta en una dirección clara: los entornos pedagógicos que integran múltiples canales sensoriales producen resultados superiores a los enfoques unisensoriales o tradicionales en prácticamente todos los dominios del desarrollo temprano examinados.

No obstante, esta base de evidencia presenta limitaciones que restringen su capacidad de orientar la práctica pedagógica con suficiente precisión. En primer lugar, la mayoría de los estudios examina el aprendizaje multisensorial como constructo global, sin desagregar los efectos por canal sensorial específico, lo que impide determinar si las dimensiones visual, auditiva y kinestésica responden de manera homogénea a las estrategias neuroeducativas o si existen diferencias sistemáticas entre canales que deberían orientar intervenciones diferenciadas (Walsh et al., 2024; Nievas & Gallardo-Montes, 2023). En segundo lugar, predominan los diseños cualitativos, de revisión o cuasiexperimentales sin instrumentos de medición estandarizados y validados, lo que limita la acumulación de evidencia empírica cuantitativa comparable entre contextos (Ledesma-Pérez et al., 2024; Ashari, 2024). En tercer lugar, la evidencia empírica contextualizada en América Latina y, específicamente, en el contexto educativo peruano post-pandémico es prácticamente inexistente, a pesar de que la interrupción prolongada de la escolarización presencial en este contexto generó déficits sensoriales y motrices en la primera infancia que hacen especialmente pertinente la generación de evidencia local. El presente estudio aborda estas tres limitaciones mediante un diseño correlacional cuantitativo con instrumentos validados, un análisis diferenciado por canal sensorial y una muestra contextualizada en educación inicial en Piura, Perú.

En este marco, el estudio tuvo como objetivo general determinar la asociación entre el nivel de uso de estrategias neuroeducativas y el aprendizaje multisensorial en niños de tres años de educación inicial en Piura, Perú, durante el año 2025. Los objetivos específicos fueron identificar la asociación entre el nivel de uso de estrategias neuroeducativas y el aprendizaje visual, el aprendizaje kinestésico y el aprendizaje auditivo, respectivamente, en la misma población. Se hipotetizó que el nivel de uso de estrategias neuroeducativas se asociaría positivamente con el nivel global de aprendizaje multisensorial. De manera específica, se esperaba que esta asociación positiva se mantuviera en cada uno de los tres canales sensoriales examinados: se hipotetizó que un mayor nivel de uso de estrategias neuroeducativas se asociaría con niveles más altos de aprendizaje visual, de aprendizaje kinestésico y de aprendizaje auditivo en los estudiantes observados.

MÉTODO

2.1 Diseño de investigación

Este estudio adoptó un enfoque cuantitativo en el marco de una investigación básica, con un diseño no experimental, correlacional y transversal. Los datos fueron recogidos en un único momento del año lectivo 2025, sin manipulación de ninguna variable, lo que permitió que las condiciones naturales del aula permanecieran inalteradas durante todo el proceso de recogida de información. El carácter transversal del diseño es directamente relevante para la estrategia analítica empleada: al haber sido medidas ambas variables de forma simultánea y bajo condiciones observacionales idénticas, su covariación pudo examinarse sin necesidad de establecer supuestos sobre precedencia temporal ni direccionalidad causal.

La selección de un diseño no experimental correlacional se justificó sobre dos bases complementarias. Desde el plano metodológico —que constituye el argumento principal— este diseño es apropiado cuando el objetivo de investigación consiste en examinar el grado de asociación entre variables tal como se manifiestan de forma natural en entornos educativos, sin

las restricciones artificiales que la manipulación experimental impondría sobre la práctica pedagógica (Ato, López & Benavente, 2013; Kerlinger & Lee, 2002). Desde el plano ético —que complementa el anterior sin equiparársele—, la realización de experimentos aleatorizados con niños de tres años en aulas de educación inicial plantea restricciones procedimentales y éticas de considerable peso, lo que hace del diseño observacional la opción más viable y responsable para este contexto de investigación (Creswell & Creswell, 2018).

2.2 Participantes

El estudio se llevó a cabo en una institución de educación inicial de la ciudad de Piura, Perú, durante el mes de abril de 2025. La población objetivo estuvo conformada por la totalidad de los niños de tres años matriculados en un aula: 18 estudiantes que cumplieron los criterios de inclusión establecidos. Dado que la población total era pequeña, delimitada y plenamente accesible, se adoptó un enfoque censal mediante el cual se incluyeron todos los casos disponibles en lugar de seleccionar una muestra. Esta estrategia elimina el error muestral y maximiza la precisión descriptiva alcanzable dentro del contexto educativo específico objeto de estudio (Hernández, Fernández & Baptista, 2014). La limitación inherente a esta decisión —esto es, que los hallazgos no son generalizables a otras poblaciones de educación inicial— se reconoce explícitamente y se aborda en la sección de discusión.

Los criterios de inclusión fueron los siguientes: (a) matrícula activa en el aula observada en el momento de la recogida de datos; (b) asistencia regular durante el período de observación; y (c) presentación del consentimiento informado firmado por el tutor legal del niño antes del inicio del estudio. Ningún participante fue excluido durante el proceso de recogida de datos. El estudio cumplió con los lineamientos éticos establecidos por la Universidad César Vallejo para la investigación educativa con menores de edad, incluyendo la garantía de anonimato de los participantes, el almacenamiento confidencial de los datos y el derecho de los tutores a retirar

el consentimiento en cualquier momento sin consecuencia alguna (APA, 2020; Alderson & Morrow, 2011).

2.3 Variables e instrumentos

La **neuroeducación** fue definida operacionalmente como el nivel observado de uso de estrategias pedagógicas fundamentadas en principios neurocientíficos durante las sesiones de aprendizaje en el aula. Se evaluó mediante una lista de cotejo estructurada compuesta por 13 ítems dicotómicos (Sí = 1; No = 0), organizados en tres dimensiones: neuroplasticidad (4 ítems), estilos de aprendizaje (4 ítems) y procesos cognitivos (5 ítems). Los ítems fueron diseñados para capturar indicadores observables de enseñanza fundamentada en neurociencia, incluyendo la promoción de nuevas conexiones neuronales a través de la exploración, la integración de actividades visuales, auditivas y kinestésicas, y el fomento de procesos atencionales y de memoria (Tokuhami-Espinosa, 2020; Mora, 2021). Las puntuaciones totales fueron categorizadas en tres niveles ordinales mediante intervalos de amplitud igual: uso limitado (0–4.32 puntos), uso moderado (4.33–8.65 puntos) y uso elevado (8.66–13 puntos).

El **aprendizaje multisensorial** fue definido operacionalmente como el nivel de desempeño observable en actividades pedagógicas que involucran múltiples canales sensoriales. Se evaluó mediante una segunda lista de cotejo compuesta por 12 ítems dicotómicos, distribuidos en tres dimensiones: canales sensoriales (4 ítems), estrategias didácticas multisensoriales (4 ítems) y participación activa del estudiante (4 ítems). Esta operacionalización es coherente con el marco teórico fundamental de Shams y Seitz (2008), quienes postularon que el cerebro humano está evolutivamente preparado para aprender de manera más eficaz en condiciones multisensoriales, y con la revisión actualizada de Murray y Shams (2023), que demuestra que la codificación multimodal mejora tanto la consolidación de la memoria como el procesamiento unisensorial. El sustento teórico del instrumento también recoge los aportes del marco de las inteligencias múltiples de Gardner (1983, 2006) y del modelo de estilos de aprendizaje individual de Dunn

y Dunn (1992), ambos coincidentes en subrayar la importancia de atender a las diversas modalidades sensoriales en el diseño instruccional. Se construyeron baremos independientes para cada dimensión sensorial (rango de puntuación: 0–4 por dimensión), aplicando las siguientes categorías: desempeño limitado (0–1.49), desempeño adecuado (1.50–2.99) y desempeño alto (3–4). El uso de baremos por dimensión, en lugar de una única puntuación compuesta, constituyó una decisión metodológica deliberada: permite el análisis diferencial entre los canales visual, kinestésico y auditivo, respondiendo directamente a los objetivos específicos del estudio.

La tradición pedagógica Montessori ofrece un anclaje teórico adicional para el instrumento de aprendizaje multisensorial, particularmente en lo relativo al papel de los entornos preparados y los materiales sensoriales en el desarrollo de la primera infancia. Mavrič (2020) demostró que el enfoque Montessori —fundamentado en principios constructivistas— enfatiza el compromiso sensoriomotriz activo como mecanismo primario de construcción del conocimiento en niños de edad preescolar. Esta posición es consistente con la evidencia empírica aportada por Courtier et al. (2021), quienes, en un estudio controlado aleatorizado y prerregistrado, mostraron que las interacciones sensoriomotrices en actividades tempranas de alfabetización —como la exploración háptica de letras— producen ganancias cognitivas medibles, especialmente cuando la codificación multisensorial refuerza el procesamiento unisensorial posterior.

La validez de contenido de ambos instrumentos se estableció mediante juicio de expertos. Tres especialistas con formación en psicología educativa y neuroeducación evaluaron de forma independiente cada ítem en función de cuatro criterios: suficiencia, claridad, coherencia y relevancia, siguiendo el protocolo propuesto por Escobar y Cuervo (2008). Todos los ítems de ambos instrumentos recibieron la puntuación máxima (1) en los cuatro criterios, lo que confirma la pertinencia conceptual y la claridad del conjunto de ítems en relación con el marco

teórico de cada variable. Cabe precisar que este procedimiento aporta evidencia de validez de contenido valorada por un panel reducido de expertos; no constituye evidencia de validez de constructo, la cual requeriría procedimientos psicométricos adicionales que exceden el alcance del presente estudio.

La consistencia interna de cada instrumento fue estimada mediante el coeficiente Kuder-Richardson 20 (KR-20), índice de confiabilidad apropiado para escalas con ítems dicotómicos (Kuder & Richardson, 1937). La lista de cotejo de neuroeducación obtuvo un $KR-20 = 0.86$, y la lista de aprendizaje multisensorial alcanzó un $KR-20 = 0.85$. Ambos valores superan el umbral convencional de 0.80 para una consistencia interna aceptable en investigación en ciencias sociales (Nunnally & Bernstein, 1994; George & Mallery, 2019), lo que indica que los ítems de cada instrumento se comportaron de manera coherente y midieron sus respectivos constructos de forma estable y homogénea en el contexto de esta muestra.

2.4 Procedimiento

La recogida de datos se llevó a cabo durante el mes de abril de 2025. Ambas investigadoras realizaron observaciones directas estructuradas de cada participante a lo largo de las sesiones regulares de clase desarrolladas durante ese período, empleando las dos listas de cotejo como instrumentos de registro. Las observaciones se efectuaron en el entorno natural del aula, con ambas investigadoras situadas en posición de observadoras no participantes, de modo que no interfirieran con la dinámica ordinaria de las sesiones pedagógicas. El registro de los indicadores de cada lista de cotejo fue realizado de forma conjunta por las dos investigadoras, procedimiento que fortalece la objetividad de los datos observacionales al reducir la dependencia del juicio de una sola observadora (Bakeman & Gottman, 1986; Martin & Bateson, 2007).

La observación se orientó a comportamientos y respuestas que emergieron de forma natural en el transcurso de las actividades de enseñanza habituales, garantizando que los datos capturaran

interacciones pedagógicas auténticas y no situaciones provocadas o inducidas por las investigadoras. Este enfoque naturalista es coherente con el carácter no experimental del estudio y con las recomendaciones metodológicas estándar para la investigación observacional con niños pequeños (Bakeman & Gottman, 1986).

Previo al inicio de las observaciones, los tutores legales de todos los participantes firmaron el consentimiento informado. El protocolo del estudio fue revisado y aprobado por la instancia ética institucional de la Universidad César Vallejo. Se garantizó el anonimato de los participantes durante todo el proceso de recogida, almacenamiento y reporte de los datos. Todos los procedimientos se realizaron en conformidad con los principios éticos que rigen la investigación con menores de edad (APA, 2020; Alderson & Morrow, 2011).

2.5 Análisis de datos

Los datos fueron analizados con el software IBM SPSS Statistics versión 28. En la fase descriptiva se calcularon frecuencias absolutas y relativas, medidas de tendencia central (media, mediana y moda) y de dispersión (desviación estándar) para ambas variables. Estas estadísticas cumplieron una función preparatoria, caracterizando el comportamiento distribucional de las variables con anterioridad a las pruebas inferenciales; su interpretación se reporta en la sección de resultados.

Para la fase inferencial se seleccionó una estrategia analítica no paramétrica sustentada en tres criterios justificados de forma encadenada. En primer lugar, ambas variables fueron medidas en escala ordinal derivada de la categorización de puntuaciones obtenidas a partir de ítems dicotómicos; este nivel de medición excluye el uso de procedimientos paramétricos que asumen datos de intervalo. En segundo lugar, el tamaño muestral reducido ($n = 18$) y la naturaleza discreta y categórica de los datos hacen inviable el supuesto de normalidad; la aplicación de pruebas formales de normalidad bajo estas condiciones carecería de valor informativo, dada su bien documentada falta de potencia estadística con muestras de este

tamaño (Field, 2018). En tercer lugar, la prueba Chi-cuadrado de Pearson constituye el procedimiento no paramétrico apropiado para examinar la existencia y dirección de la asociación entre variables ordinales o categóricas en diseños transversales observacionales con tablas de contingencia (Siegel & Castellan, 1988; Hernández et al., 2014).

Previo a la interpretación de los resultados del Chi-cuadrado, se examinó el cumplimiento del supuesto de frecuencias esperadas mínimas por celda en cada tabla de contingencia. El criterio convencional establece que no más del 20% de las celdas puede presentar frecuencias esperadas inferiores a 5, y que ninguna celda puede registrar una frecuencia esperada inferior a 1 (Kim, 2017). El examen de las frecuencias esperadas reveló que, en la totalidad de las tablas de contingencia analizadas, más del 20% de las celdas presentaron frecuencias esperadas inferiores a 5, situación atribuible al tamaño muestral reducido y a la concentración de casos en determinadas categorías ordinales. Ante este resultado, se optó por una estrategia de triangulación estadística: el Chi-cuadrado de Pearson fue complementado en todos los análisis con la razón de verosimilitud y el estadístico de asociación lineal por lineal. La convergencia o divergencia entre estos tres estadísticos se utilizó como criterio de robustez para fundamentar cada conclusión inferencial, compensando así la limitación impuesta por el incumplimiento del supuesto de frecuencias esperadas (Kim, 2017; Field, 2018). El umbral de significancia estadística se estableció en $\alpha = .05$ para todas las pruebas.

Adicionalmente, se calculó la V de Cramér como medida del tamaño del efecto para cada asociación examinada. Este índice, apropiado para tablas de contingencia con variables nominales u ordinales, permite cuantificar la magnitud práctica de las relaciones identificadas con independencia de la significancia estadística (Field, 2018). Su inclusión responde a la recomendación metodológica de no limitar la interpretación de los resultados al valor p, especialmente en estudios con muestras pequeñas donde la potencia estadística puede ser insuficiente para detectar efectos de magnitud moderada o grande (Creswell & Creswell, 2018).

Los valores de V de Cramér se reportan junto a los estadísticos inferenciales en la sección de resultados.

RESULTADOS

3.1 Estadísticos descriptivos

La Tabla 1 presenta las medidas de tendencia central y dispersión para las dos variables principales del estudio. Dado el nivel de medición ordinal de ambas variables y el tamaño muestral ($n = 18$), los estadísticos descriptivos cumplen una función orientadora sobre el comportamiento distribucional del grupo y no constituyen estimaciones poblacionales. En consecuencia, no se reportan intervalos de confianza para las medias.

Tabla 1. Estadísticos descriptivos del nivel de uso de estrategias neuroeducativas y del aprendizaje multisensorial en niños de tres años ($n = 18$)

Estadístico	Neuroeducación	Aprendizaje multisensorial
Media	2.56	2.44
Mediana	3.00	2.50
Moda	3	3
Desviación estándar	0.511	0.616

Fuente: elaboración propia

Nota. Las variables fueron codificadas en escala ordinal de 1 (bajo/limitado) a 3 (alto/elevado).

N válido = 18; valores perdidos = 0.

La media del nivel de uso de estrategias neuroeducativas ($M = 2.56$; $DE = 0.511$), junto con una mediana y moda de 3, indica una tendencia del grupo hacia el nivel de uso elevado. El aprendizaje multisensorial presenta una distribución centrada entre los niveles regular y alto ($M = 2.44$; $DE = 0.616$; $Md = 2.50$; $Mo = 3$). Las desviaciones estándar moderadas de ambas variables reflejan una variabilidad reducida en el grupo, coherente con la homogeneidad relativa de una muestra conformada por niños de una misma aula y rango etario.

3.2 Neuroeducación y aprendizaje multisensorial

La Tabla 2 presenta la distribución cruzada de frecuencias entre el nivel de uso de estrategias neuroeducativas y el nivel de aprendizaje multisensorial, junto con los estadísticos inferenciales correspondientes.

Tabla 2. Distribución de frecuencias y estadísticos de asociación entre neuroeducación y aprendizaje multisensorial en niños de tres años (n = 18)

	Aprendizaje multisensorial						
Neuroeducación	Bajo		Regular		Alto		Total
	fi	%	fi	%	fi	%	fi
Uso moderado	1	5.6	7	38.9	0	0.0	8
Uso elevado	0	0.0	1	5.6	9	50.0	10
Total	1	5.6	8	44.4	9	50.0	18
Estadístico	Valor	gl	p	V de Cramér			
Chi-cuadrado de Pearson	14.456	2	.001	.896			
Razón de verosimilitud	18.702	2	.000				
Asociación lineal por lineal	12.318	1	.000				

Nota. El supuesto de frecuencias esperadas mínimas no se cumplió plenamente en esta tabla (más del 20% de celdas con frecuencias esperadas < 5; Kim, 2017); los resultados del Chi-cuadrado de Pearson deben interpretarse en conjunto con la razón de verosimilitud y la asociación lineal por lineal, empleados como evidencia triangulada. V de Cramér interpretada según Field (2018): pequeño < .30; moderado .30–.50; grande > .50.

La distribución cruzada muestra un patrón marcadamente ordenado. El 50.0% de los estudiantes con desempeño multisensorial alto pertenece al grupo con uso elevado de

estrategias neuroeducativas, mientras que ningún estudiante de ese grupo presentó desempeño bajo. En contraste, el 38.9% de los casos con desempeño regular se concentró en el grupo con uso moderado. Ningún estudiante registró simultáneamente uso limitado de estrategias y nivel superior de desempeño multisensorial, lo que evidencia una distribución consistente con una tendencia ascendente entre ambas variables.

La prueba Chi-cuadrado de Pearson reveló una asociación estadísticamente significativa entre el nivel de uso de estrategias neuroeducativas y el aprendizaje multisensorial ($\chi^2(2) = 14.456$, $p = .001$). La razón de verosimilitud ($p = .000$) y la asociación lineal por lineal ($p = .000$) convergieron con este resultado, fortaleciendo la robustez de la conclusión mediante triangulación estadística (véase sección de análisis de datos). La V de Cramér = .896 indica una asociación de magnitud muy grande (Field, 2018). En conjunto, los resultados muestran que los estudiantes expuestos a un mayor nivel de uso de estrategias neuroeducativas presentaron niveles sustancialmente más altos de desempeño multisensorial, con un efecto de magnitud que excede ampliamente los umbrales convencionales de relevancia práctica.

3.3 Neuroeducación y aprendizaje visual

La Tabla 3 presenta la distribución cruzada de frecuencias entre el nivel de uso de estrategias neuroeducativas y el nivel de aprendizaje visual, junto con los estadísticos inferenciales.

Tabla 3. Distribución de frecuencias y estadísticos de asociación entre neuroeducación y aprendizaje visual en niños de tres años ($n = 18$)

	Aprendizaje visual						
Neuroeducación	Bajo		Regular		Alto		Total
	fi	%	fi	%	fi	%	fi
Uso moderado	3	16.7	4	22.2	1	5.6	8
Uso elevado	1	5.6	2	11.1	7	38.9	10
Total	4	22.2	6	33.3	8	44.4	18

Estadístico	Valor	gl	p	V de Cramér			
Chi-cuadrado de Pearson	6.019	2	.049	.578			
Razón de verosimilitud	6.565	2	.038				
Asociación lineal por lineal	4.913	1	.027				

Nota. El supuesto de frecuencias esperadas mínimas no se cumplió plenamente en esta tabla (más del 20% de celdas con frecuencias esperadas < 5 ; Kim, 2017); los resultados se interpretan mediante triangulación estadística. V de Cramér interpretada según Field (2018).

La distribución cruzada indica que el 38.9% de los estudiantes con desempeño visual alto pertenece al grupo con uso elevado de estrategias neuroeducativas, frente a un 5.6% en el grupo con uso moderado. Los niveles bajo y regular de aprendizaje visual se concentran predominantemente en el grupo con uso moderado, donde el 16.7% y el 22.2% de los estudiantes presentaron desempeño bajo y regular, respectivamente. Esta distribución describe una tendencia ascendente entre el nivel de uso de estrategias y el nivel de desempeño en el canal visual.

La asociación entre neuroeducación y aprendizaje visual fue estadísticamente significativa ($\chi^2(2) = 6.019$, $p = .049$). El resultado se sitúa en el límite convencional de significancia; no obstante, la razón de verosimilitud ($p = .038$) y la asociación lineal por lineal ($p = .027$) aportaron evidencia convergente que refuerza la solidez de la conclusión mediante triangulación estadística (véase sección de análisis de datos). La V de Cramér = .578 indica un efecto de magnitud grande (Field, 2018). Los estudiantes expuestos a mayor uso de estrategias neuroeducativas tendieron a mostrar niveles más altos de desempeño en el canal visual, con un efecto de magnitud grande pese al tamaño muestral reducido.

3.4 Neuroeducación y aprendizaje kinestésico

La Tabla 4 presenta la distribución cruzada de frecuencias entre el nivel de uso de estrategias neuroeducativas y el nivel de aprendizaje kinestésico, junto con los estadísticos inferenciales.

Tabla 4. Distribución de frecuencias y estadísticos de asociación entre neuroeducación y aprendizaje kinestésico en niños de tres años (n = 18)

	Aprendizaje kinestésico						
Neuroeducación	Bajo		Regular		Alto		Total
	fi	%	fi	%	fi	%	fi
Uso moderado	3	16.7	1	5.6	4	22.2	8
Uso elevado	0	0.0	1	5.6	9	50.0	10
Total	3	16.7	2	11.1	13	72.2	18
Estadístico	Valor	gl	p	V de Cramér			
Chi-cuadrado de Pearson	4.760	2	.093	.514			
Razón de verosimilitud	5.910	2	.052				
Asociación lineal por lineal	4.345	1	.037				

Nota. El supuesto de frecuencias esperadas mínimas no se cumplió en esta tabla (más del 20% de celdas con frecuencias esperadas < 5 , incluyendo una celda con frecuencia esperada < 1 ; Kim, 2017); los resultados se interpretan mediante triangulación estadística. V de Cramér interpretada según Field (2018).

La distribución cruzada muestra que el 50.0% de los estudiantes con desempeño kinestésico alto pertenece al grupo con uso elevado de estrategias neuroeducativas, mientras que ningún estudiante de ese grupo registró desempeño kinestésico bajo. El 22.2% de los estudiantes con desempeño alto se ubicó en el grupo con uso moderado. Esta distribución describe una

concentración marcada de desempeño alto en el grupo con mayor uso de estrategias, con una tendencia ascendente entre ambas variables.

No se encontró evidencia de asociación estadísticamente significativa entre el nivel de uso de estrategias neuroeducativas y el aprendizaje kinestésico mediante la prueba Chi-cuadrado de Pearson ($\chi^2(2) = 4.760$, $p = .093$). Sin embargo, dos elementos del análisis requieren declaración explícita. En primer lugar, el estadístico de asociación lineal por lineal resultó significativo ($p = .037$), indicando la presencia de una tendencia progresiva y ordenada entre los niveles de ambas variables que el estadístico global no detectó. En segundo lugar, la V de Cramér = .514 indica un efecto de magnitud grande (Field, 2018), valor que contrasta con la ausencia de significancia estadística global. Estos dos elementos — asociación lineal significativa y efecto grande no acompañado de significancia global — se reportan conjuntamente como información sustantiva del análisis, sin que ello implique modificar la conclusión inferencial sobre la hipótesis global.

3.5 Neuroeducación y aprendizaje auditivo

La Tabla 5 presenta la distribución cruzada de frecuencias entre el nivel de uso de estrategias neuroeducativas y el nivel de aprendizaje auditivo, junto con los estadísticos inferenciales.

Tabla 5. Distribución de frecuencias y estadísticos de asociación entre neuroeducación y aprendizaje auditivo en niños de tres años ($n = 18$)

	Aprendizaje auditivo						
Neuroeducación	Bajo		Regular		Alto		Total
	fi	%	fi	%	fi	%	fi
Uso moderado	1	5.6	3	16.7	4	22.2	8
Uso elevado	0	0.0	2	11.1	8	44.4	10
Total	1	5.6	5	27.8	12	66.7	18

Estadístico	Valor	gl	p	V de Cramér			
Chi-cuadrado de Pearson	2.340	2	.310	.360			
Razón de verosimilitud	2.724	2	.256				
Asociación lineal por lineal	2.174	1	.140				

Nota. El supuesto de frecuencias esperadas mínimas no se cumplió en esta tabla (más del 20% de celdas con frecuencias esperadas < 5 , incluyendo celdas con frecuencia esperada < 1 ; Kim, 2017); los resultados se interpretan mediante triangulación estadística. La convergencia de los tres estadísticos en la ausencia de significancia descarta que el resultado sea un artefacto del incumplimiento del supuesto. V de Cramér interpretada según Field (2018).

La distribución cruzada muestra que el 44.4% de los estudiantes con desempeño auditivo alto pertenece al grupo con uso elevado de estrategias neuroeducativas, frente al 22.2% en el grupo con uso moderado. Los niveles bajo y regular de desempeño auditivo se concentran exclusivamente en el grupo con uso moderado. Pese a esta distribución, la tendencia es menos pronunciada que la observada en las dimensiones visual y kinestésica.

No se encontró evidencia de asociación estadísticamente significativa entre el nivel de uso de estrategias neuroeducativas y el aprendizaje auditivo ($\chi^2(2) = 2.340$, $p = .310$). La razón de verosimilitud ($p = .256$) y la asociación lineal por lineal ($p = .140$) confirmaron este resultado, constituyendo el patrón inferencial más consistente de ausencia de asociación en el estudio: los tres estadísticos convergen en valores alejados del umbral de significancia, lo que descarta que la ausencia de relación sea atribuible al incumplimiento del supuesto de frecuencias esperadas. La V de Cramér = .360 indica un efecto de magnitud moderada que, no obstante, no alcanzó significancia estadística en esta muestra (Field, 2018).

DISCUSIÓN

El presente estudio examinó la asociación entre el nivel de uso de estrategias neuroeducativas y el aprendizaje multisensorial en niños de tres años de educación inicial en Piura, Perú. Los

resultados revelaron una asociación estadísticamente significativa y de magnitud muy grande entre ambas variables globales ($\chi^2(2) = 14.456$, $p = .001$, $V = .896$), así como una asociación significativa de magnitud grande entre neuroeducación y aprendizaje visual ($\chi^2(2) = 6.019$, $p = .049$, $V = .578$). En el canal kinestésico, no se alcanzó significancia estadística global ($\chi^2(2) = 4.760$, $p = .093$), aunque la asociación lineal resultó significativa ($p = .037$) y el tamaño del efecto fue grande ($V = .514$). El canal auditivo no mostró asociación significativa en ninguno de los estadísticos examinados ($\chi^2(2) = 2.340$, $p = .310$, $V = .360$).

La asociación global de magnitud muy grande entre neuroeducación y aprendizaje multisensorial es coherente con los postulados de la Teoría de la Integración Multisensorial de Shams y Seitz (2008), según la cual el cerebro humano está evolutivamente preparado para aprender de manera más eficaz cuando múltiples canales sensoriales son activados de forma coordinada, generando representaciones neurales más robustas y duraderas que las producidas por estimulación unisensorial. Desde esta perspectiva, el uso sistemático de estrategias neuroeducativas en el aula de educación inicial — al promover simultáneamente la neuroplasticidad, los diferentes estilos de aprendizaje y los procesos cognitivos — crearía las condiciones óptimas para que la integración multisensorial opere con mayor eficacia, lo que explicaría la magnitud sustancial del efecto observado. Este hallazgo es consistente con los reportes de Illapa (2024), cuyo estudio en niños de cuatro a cinco años documentó efectos positivos de estrategias neurodidácticas sobre el desarrollo cognitivo global en contextos de educación inicial, coincidiendo con el patrón general identificado en el presente estudio.

La asociación significativa entre neuroeducación y aprendizaje visual es interpretable desde el mismo marco teórico, con un argumento adicional: el canal visual es el más consistentemente estimulado por los materiales pedagógicos estructurados propios del aula de educación inicial — imágenes, colores, formas, tarjetas ilustradas y recursos gráficos — lo que lo convierte en el canal de mayor exposición acumulada durante las sesiones observadas. Esta

sobrerrepresentación de estímulos visuales en el entorno pedagógico habitual explicaría por qué el canal visual es el único canal específico que alcanzó significancia estadística. Este resultado coincide cualitativamente con los hallazgos de Wulandari y Pujaningsih (2025), quienes reportaron mejoras en habilidades lectoras tempranas mediante la estrategia VAKT, cuyo componente visual demostró ser especialmente efectivo en la activación del procesamiento perceptivo. Adicionalmente, Courtier et al. (2021) documentaron que la exploración visual de materiales estructurados en entornos de educación inicial produce ganancias cognitivas medibles, coherentes con los niveles de desempeño visual observados en el grupo con uso elevado de estrategias en el presente estudio.

El resultado kinestésico constituye el hallazgo más complejo del estudio. La coexistencia de ausencia de significancia estadística global ($p = .093$) con una asociación lineal significativa ($p = .037$) y un efecto de magnitud grande ($V = .514$) sugiere que la relación entre neuroeducación y aprendizaje kinestésico existe en la muestra, pero que las condiciones de medición no permitieron su detección plena. Desde el principio del aprendizaje encarnado, documentado en la tradición Montessori por Mavrič (2020) y Courtier et al. (2021), el canal kinestésico requiere activación motriz sostenida y repetida para consolidarse en representaciones neurales observables — condición que un diseño transversal con observación única puede subestimar sistemáticamente. Este argumento encuentra respaldo en Bolaños (2025), quien reportó efectos positivos de materiales multisensoriales con componente táctil-kinestésico sobre la estimulación neurofuncional en niños de dos a cuatro años, sugiriendo que la relación entre estimulación kinestésica y desarrollo neurológico temprano es real pero requiere condiciones de medición longitudinal más sensibles que las disponibles en este estudio.

La ausencia de asociación significativa en el canal auditivo — el resultado más consistentemente no significativo del estudio, con los tres estadísticos alejados del umbral de

significancia — puede atribuirse a factores contextuales propios del entorno natural de observación. La distribución de frecuencias de la Tabla 5 muestra que el 44.4% de los estudiantes con desempeño auditivo alto pertenece al grupo con uso elevado de estrategias, lo que sugiere una tendencia real pero insuficiente para alcanzar significancia con el tamaño muestral disponible. El canal auditivo requiere condiciones acústicas controladas y estrategias pedagógicas específicamente diseñadas para su estimulación sistemática — condiciones difíciles de garantizar en entornos naturales de aula durante observación única. Este vacío es compartido con Harahap et al. (2025), cuyo estudio incorporó componentes auditivos en su intervención pero no los aisló como variable independiente, impidiendo comparaciones directas. Walsh et al. (2024) documentaron que la implementación inconsistente de estrategias neuroeducativas en contextos naturales puede atenuar los efectos medibles en canales que requieren condiciones más controladas, lo que constituye una limitación compartida entre estudios observacionales en este campo.

Los hallazgos del estudio deben interpretarse considerando tres limitaciones. El tamaño muestral reducido ($n = 18$) limita la generalización de los resultados y, como se evidenció en el canal kinestésico, restringe la potencia estadística disponible para detectar efectos reales. El diseño transversal impide establecer relaciones causales y puede subestimar efectos que requieren exposición pedagógica sostenida. El incumplimiento del supuesto de frecuencias esperadas en todas las tablas de contingencia, aunque abordado mediante triangulación estadística, introduce incertidumbre adicional en la precisión de los valores p reportados (Field, 2018). Estas limitaciones sugieren líneas de investigación futura prioritarias: estudios longitudinales con muestras más amplias que permitan examinar la evolución de la asociación entre neuroeducación y cada canal sensorial a lo largo del año lectivo, y diseños cuasiexperimentales que posibiliten comparaciones entre grupos con y sin intervención

neuroeducativa sistemática, especialmente en las dimensiones kinestésica y auditiva donde el presente estudio identificó efectos sustanciales sin confirmación inferencial plena.

CONCLUSIONES

El presente estudio aporta evidencia empírica cuantitativa sobre la asociación entre el uso de estrategias neuroeducativas y el aprendizaje multisensorial en la educación inicial, contribuyendo a cubrir un vacío identificado en la literatura: la escasez de estudios que examinen de forma diferenciada la respuesta de los canales sensoriales visual, kinestésico y auditivo a las prácticas pedagógicas fundamentadas en principios neurocientíficos, en contextos educativos latinoamericanos post-pandémicos y con instrumentos de medición validados.

El hallazgo principal del estudio es que el nivel de uso de estrategias neuroeducativas se asocia positivamente con el nivel de aprendizaje multisensorial global en niños de tres años, con una magnitud del efecto muy grande que supera ampliamente los umbrales convencionales de relevancia práctica. Este resultado indica que la integración sistemática de principios neurocientíficos en el diseño pedagógico del nivel inicial tiene el potencial de fortalecer el desarrollo sensorial integral de los niños en esta etapa, independientemente del canal sensorial predominante.

En el plano de los canales específicos, la evidencia es diferenciada. El canal visual mostró una asociación estadísticamente significativa con el uso de estrategias neuroeducativas y un efecto de magnitud grande, lo que lo posiciona como la dimensión del aprendizaje multisensorial más sensible a la estimulación pedagógica estructurada en este nivel educativo. El canal kinestésico presentó un efecto de magnitud grande y una tendencia lineal significativa que no alcanzaron significancia estadística global, resultado que sugiere la existencia de una relación real entre ambas variables cuya detección plena requiere condiciones de medición longitudinal y muestras de mayor tamaño. El canal auditivo no mostró asociación significativa en ninguno de

los estadísticos examinados, resultado que se interpreta como el reflejo de las condiciones específicas de estimulación auditiva en entornos naturales de aula, más que como ausencia de relación teórica entre neuroeducación y procesamiento auditivo en la primera infancia.

En conjunto, los hallazgos respaldan la pertinencia de integrar enfoques neuroeducativos en la práctica pedagógica de educación inicial como estrategia para potenciar el desarrollo multisensorial temprano, con especial atención al canal visual como el de respuesta más consistente en esta etapa. Estudios futuros con diseños longitudinales, muestras más amplias y condiciones de observación controladas permitirán establecer relaciones causales y capturar efectos en los canales kinestésico y auditivo con mayor precisión que la disponible en el presente estudio.

REFERENCIAS

- Alderson, P., & Morrow, V. (2011). *The ethics of research with children and young people: A practical handbook*. SAGE Publications.
- American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). <https://doi.org/10.1037/0000165-000>
- Ashari, N. (2024). Neuroscience-based learning model in early childhood education. *Atfālunā Journal of Islamic Early Childhood Education*. <https://doi.org/10.32505/atfaluna.v7i2.8491>
- Ato, M., López, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038–1059. <https://doi.org/10.6018/analesps.29.3.178511>
- Bakeman, R., & Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis*. Cambridge University Press.

- Bolaños, S. (2025). Multisensory teaching material as a tool for neurofunctional stimulation in early childhood. *Proceedings. III Congreso Internacional Multidisciplinario "Innovación, Equidad y Sostenibilidad en la Era Digital: Nuevas Fronteras para América Latina"*. <https://doi.org/10.64736/cim2025.abs05>
- Burbano-Enríquez, B., Moncayo-Cueva, H., & Andrade-Zuleta, A. (2025). Neuroeducation and the influence of AI on early childhood education: A systematic review. *Multidisciplinary Reviews*. <https://doi.org/10.31893/multirev.2026238>
- Centeno, D. (2024). Neuroeducación en el aula. *Rastros Rostros*. <https://doi.org/10.16925/2382-4921.2024.02.18>
- Courtier, P., Gardes, M.-L., Van der Henst, J.-B., Noveck, I., Croset, M.-C., Epinat-Duclos, J., Léone, J., & Prado, J. (2021). Effects of Montessori education on the academic, cognitive, and social development of disadvantaged preschoolers: A randomized controlled study in the French public-school system. *Child Development*, 92(5), 2069–2089. <https://doi.org/10.1111/cdev.13575>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Dunn, R., & Dunn, K. J. (1992). *Teaching elementary students through their individual learning styles: Practical approaches for grades 3–6*. Allyn & Bacon.
- Escobar, J., & Cuervo, Á. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6(1), 27–36.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.
- Gardner, H. (2006). *Multiple intelligences: New horizons in theory and practice*. Basic Books.
- George, D., & Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference* (15th ed.). Routledge. <https://doi.org/10.4324/9780429056765>

- Harahap, D., Rafi'y, M., Irawan, F., De Lima, C., & Riyana, M. (2025). Reflective inquiry learning for early childhood neuroplasticity: A neuroscience approach in the Indonesia–Papua border. *Muallimun: Jurnal Kajian Pendidikan dan Keguruan*. <https://doi.org/10.23971/muallimun.v5i2.10612>
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación* (6.^a ed.). McGraw-Hill.
- Illapa, B. (2024). Neurodidactics and its impact on child development. *Universidad Ciencia y Tecnología*. <https://doi.org/10.47460/uct.v28i125.858>
- Kerlinger, F. N., & Lee, H. B. (2002). *Investigación del comportamiento: Métodos de investigación en ciencias sociales* (4.^a ed.). McGraw-Hill.
- Kim, H.-Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics*, 42(2), 152–155. <https://doi.org/10.5395/rde.2017.42.2.152>
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151–160. <https://doi.org/10.1007/BF02288391>
- Ledesma-Pérez, F., Cruz-Montero, J., & Holguin-Alvarez, J. (2024). Neuroscience for early childhood education: Impact on virtual settings and teaching practice. *International Journal of Evaluation and Research in Education*, 13(6). <https://doi.org/10.11591/ijere.v13i6.25455>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. <https://doi.org/10.1097/00006199-198611000-00017>
- Martin, P., & Bateson, P. (2007). *Measuring behaviour: An introductory guide* (3rd ed.). Cambridge University Press.
- Mavrič, M. (2020). The Montessori approach as a model of personalized instruction. *Journal of Montessori Research*, 6(2), 13–27. <https://doi.org/10.17161/jomr.v6i2.13128>

- Mora, F. (2021). *Neuroeducación: Solo se puede aprender aquello que se ama*. Alianza Editorial.
- Morales, G., & Rivadeneira, J. (2025). Attention and diversity in early education: Bases, strategies and normative framework. *Neurodivergences*. <https://doi.org/10.56294/neuro2025206>
- Murray, C. A., & Shams, L. (2023). Crossmodal interactions in human learning and memory. *Frontiers in Human Neuroscience*, 17. <https://doi.org/10.3389/fnhum.2023.1181760>
- Nievas, P., & Del Pilar Gallardo-Montes, C. (2023). The neuroeducation training of students in the degrees of early childhood and primary education: A content analysis of public universities in Andalusia. *Education Sciences*, 13(10), 1006. <https://doi.org/10.3390/educsci13101006>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Pineda, M., Lara, P., Cornejo, M., Condori, E., & Choque, D. (2025). Neuroeducation in initial basic education: Strategies and benefits for early learning. *Seminars in Medical Writing and Education*. <https://doi.org/10.56294/mw2024465>
- Sadreamova, A. (2024). Formation of successful teaching of younger schoolchildren based on the neuropedagogical approach. *Pedagogy and Psychology*. <https://doi.org/10.51889/2960-1649.2023.15.4.006>
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, 12(11), 411–417. <https://doi.org/10.1016/j.tics.2008.07.006>
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). McGraw-Hill.
- Tokuhami-Espinosa, T. (2020). *Bringing the neuroscience of learning to online teaching*. Teachers College Press.

- Walsh, K., L'Estrange, L., Smith, R., Burr, T., & Williams, K. (2024). Translating neuroscience to early childhood education: A scoping review of neuroscience-based professional learning for early childhood educators. *Educational Research Review*.
<https://doi.org/10.1016/j.edurev.2024.100644>
- Wulandari, R., & Pujaningsih, P. (2025). Enhancing early reading skills through multisensory contextual learning. *Journal of Innovation and Research in Primary Education*.
<https://doi.org/10.56916/jirpe.v4i3.1783>

Capítulo 5

Efecto de la música clásica en la neuroplasticidad y aprendizaje de lectoescritura en niños

Vania Nicole Milagros Bazán Chávez ⁶

Abstract

Objetivo: Examinar el efecto predictivo de la música clásica sobre la neuroplasticidad y el aprendizaje de la lectoescritura en niños de educación inicial. **Método:** Estudio cuantitativo, no experimental, con diseño correlacional-causal y alcance explicativo. La muestra estuvo conformada por 68 niños de cuatro y cinco años de una institución educativa pública de Trujillo, Perú, seleccionados mediante muestreo no probabilístico por conveniencia. Se aplicaron tres escalas de valoración elaboradas ad hoc y validadas por juicio de expertos: la Escala de Valoración de la Música Clásica (EVMC, 34 ítems), la Escala de Valoración de la Neuroplasticidad (EVN, 44 ítems) y la Escala de Valoración de la Lectoescritura (EVL, 38 ítems), todas con formato de respuesta Likert de cinco puntos. El análisis se realizó mediante regresión logística ordinal con IBM SPSS Statistics, con un nivel de significancia de $\alpha = .05$. **Resultados:** Ninguno de los cinco modelos de regresión examinados alcanzó significancia estadística, y la totalidad de los intervalos de confianza al 95% incluyeron el cero, tanto a nivel de variables globales como dimensional. La música clásica no ejerció efecto predictivo significativo sobre la neuroplasticidad ($B = 0.055$, $p = .186$) ni sobre la lectoescritura ($B = 0.107$, $p = .096$). **Conclusiones:** La exposición pasiva a música clásica no predice el desarrollo

⁶ Especialista en Neuroeducación por la Universidad César Vallejo, Perú. E-mail: nbazanch23@ucvvirtual.edu.pe. ORCID: <https://orcid.org/0009-0009-2970-2253>

neuroplástico ni lectoescritor en niños de educación inicial, en coherencia con la evidencia metaanalítica disponible. Los hallazgos contribuyen a la desmitificación empírica del denominado efecto Mozart en el ámbito educativo latinoamericano.

Palabras clave: música clásica, neuroplasticidad, lectoescritura, educación inicial, efecto Mozart, neuromito

1. Introducción

La primera infancia constituye una ventana crítica para el desarrollo cerebral, caracterizada por una plasticidad neuronal sin precedentes que hace de este período un momento de especial relevancia para intervenciones cognitivas de diversa naturaleza. En este contexto, la estimulación auditiva musical ha sido propuesta reiteradamente como una vía de enriquecimiento neurocognitivo con potencial impacto en habilidades académicas fundamentales, entre ellas la lectoescritura. El interés científico por esta relación ha crecido sostenidamente durante la última década, impulsado tanto por avances en neuroimagen que permiten documentar cambios cerebrales asociados a la experiencia musical, como por la demanda de prácticas educativas basadas en evidencia en los primeros años de escolaridad (Zaatar et al., 2023; López, 2024). Sin embargo, la distancia entre las afirmaciones populares sobre los beneficios de la música clásica y la evidencia empírica rigurosa disponible es considerable, y su examen sistemático resulta especialmente necesario en contextos latinoamericanos donde la investigación empírica en educación inicial permanece subrepresentada en la literatura internacional (Wulandari et al., 2025).

La evidencia neurocientífica acumulada sustenta con solidez que la formación musical activa en la infancia produce cambios estructurales y funcionales documentables en el cerebro. Estudios recientes han reportado incrementos en materia gris, mayor conectividad en vías auditivo-motoras e interhemisféricas, y fortalecimiento de redes de atención y función ejecutiva asociados a la práctica musical sistemática (Tianjiao et al., 2025; Fernandez &

Newman, 2025). A nivel molecular, se ha documentado la modulación de factores neurotróficos como el BDNF, vinculados a la plasticidad sináptica y a la reducción del estrés neurofisiológico (U et al., 2025). Estos hallazgos posicionan a la música como un estímulo con capacidad real de inducir neuroplasticidad, en particular cuando la exposición es activa, sistemática y comenzada en etapas tempranas del desarrollo (López, 2024).

No obstante, la literatura distingue de manera consistente y crítica entre dos modalidades de exposición musical cuyas implicaciones neurocognitivas difieren de forma sustancial. La formación musical activa —que incluye el aprendizaje de instrumentos, el canto y el entrenamiento rítmico— es la modalidad a la que corresponde la mayor parte de la evidencia neuroplástica documentada. La escucha pasiva de música clásica, en cambio, se asocia popularmente al denominado efecto Mozart, término acuñado para describir la creencia de que la mera audición de composiciones clásicas mejora las capacidades cognitivas infantiles (Campbell, 1998). Esta creencia ha ejercido una influencia notable en prácticas educativas a nivel global, pero ha recibido respaldo científico considerablemente más débil. Revisiones contemporáneas sugieren que la música clásica puede operar como estímulo de apoyo atencional y regulación emocional dentro de actividades pedagógicas más amplias, pero no como inductor directo de modificaciones estructurales o funcionales en el cerebro infantil, lo que ha llevado a catalogar las afirmaciones más amplias del efecto Mozart como un neuromito (Wulandari et al., 2025; Sala & Gobet, 2020).

La relación entre exposición musical y lectoescritura presenta un panorama igualmente matizado. La alfabetización musical se ha asociado a mejoras en lectura de palabras y a la activación bilateral de áreas temporooccipitales especializadas en procesamiento ortográfico (Pantaleo et al., 2024). El entrenamiento rítmico muestra vínculos con habilidades fonológicas, con efectos más consistentes en lectores con dificultades que en lectores con desarrollo típico (Ahokas et al., 2024; Dees & Cooper, 2025). Las intervenciones musicales para niños con

dificultades lectoras muestran efectos positivos pero inconsistentes, con ganancias más fiables en conciencia fonológica que en comprensión lectora global (Cancer & Antonietti, 2022). De manera relevante, un metaanálisis multinivel que incluyó 6,984 niños encontró que, al controlar la calidad metodológica de los estudios, la transferencia de la formación musical a habilidades académicas no musicales, incluida la lectoescritura, es estadísticamente nula (Sala & Gobet, 2020). El entorno musical en el hogar muestra una asociación indirecta con prácticas de lectura compartida, sugiriendo que las vías de influencia son mediadoras y complejas, no directas ni automáticas (Liu et al., 2024).

A partir de este estado del arte se identifican vacíos empíricos relevantes que justifican la presente investigación. En primer lugar, la mayoría de estudios que reportan efectos positivos sobre neuroplasticidad y lectoescritura corresponden a intervenciones de formación musical activa y no a escucha pasiva de música clásica, lo que limita la extrapolación a contextos educativos donde la música se emplea predominantemente como estímulo ambiental. En segundo lugar, la investigación en población de educación inicial en contextos latinoamericanos es escasa, lo que impide establecer con certeza si los hallazgos obtenidos en contextos anglosajones o europeos son transferibles a otras realidades socioeducativas. En tercer lugar, no se dispone de estudios que examinen simultáneamente el efecto de la escucha pasiva de música clásica sobre la neuroplasticidad y la lectoescritura en niños de cuatro y cinco años mediante instrumentos validados para ese nivel educativo específico (Couvignou et al., 2024; Dees & Cooper, 2025). En este marco, se hipotetiza que, en ausencia de formación musical activa y sistemática, la exposición pasiva a música clásica no ejercerá efectos predictivos significativos sobre la neuroplasticidad ni sobre la lectoescritura en niños de educación inicial, en coherencia con la evidencia metaanalítica disponible y con una perspectiva crítica respecto al neuromito del efecto Mozart (Sala & Gobet, 2020; Campbell, 1998).

El presente estudio tuvo como objetivo general determinar el efecto de la música clásica sobre la neuroplasticidad y el aprendizaje de la lectoescritura en niños de una institución educativa de nivel inicial en Trujillo, Perú, durante el año 2025. De manera específica, se examinó la influencia de las dimensiones de la música clásica —estimulación neurocognitiva, estimulación sensorial y emocional, y desarrollo de la escucha y la comunicación— sobre la neuroplasticidad y la lectoescritura por separado; la influencia de las dimensiones de la neuroplasticidad sobre la lectoescritura; y la asociación entre las tres variables en conjunto. Para ello se empleó un diseño correlacional-causal con regresión logística ordinal como estrategia analítica principal (Osada & Salvador, 2021). El estudio realiza tres contribuciones al campo: aporta evidencia empírica desde un contexto latinoamericano subrepresentado en la literatura internacional; somete a prueba empírica rigurosa las afirmaciones asociadas al efecto Mozart en un grupo etario crítico para el desarrollo lectoescritor; e introduce instrumentos de valoración contruidos y validados específicamente para la educación inicial que pueden ser refinados y utilizados en investigaciones futuras de la región (Lugo, 2024; Cárate et al., 2023).

2. Método

2.1 Diseño de investigación

El estudio adoptó un enfoque cuantitativo, no experimental, con un diseño correlacional-causal y alcance explicativo. Este diseño fue seleccionado porque permitió examinar la capacidad predictiva de la exposición a la música clásica sobre la neuroplasticidad y la adquisición de la lectoescritura, sin recurrir a la manipulación experimental de la variable independiente, lo cual resultó éticamente pertinente y logísticamente viable en el contexto de educación inicial (Osada & Salvador, 2021). La orientación básica de la investigación, conforme a la clasificación del Manual de Oslo de la OCDE (2018), refleja el propósito de ampliar la comprensión teórica sobre las relaciones entre las variables estudiadas, sin pretender la generación de resultados de aplicación inmediata. El alcance explicativo se justificó por la necesidad de interpretar

relaciones direccionales dentro de los límites de la muestra seleccionada, sin aspiración de generalización a poblaciones más amplias (Haro et al., 2024).

2.2 Participantes

La población objetivo estuvo conformada por 115 niños de dos a cinco años matriculados en una institución educativa pública de nivel inicial en Trujillo, La Libertad, Perú, durante el año académico 2025 (Díaz, 2020). La muestra final comprendió 68 niños de cuatro y cinco años, lo que representó el 59.12% de la población total, distribuidos en cuatro aulas (sección A: $n = 16$, cuatro años; sección B: $n = 14$, cuatro años; sección A: $n = 20$, cinco años; sección B: $n = 18$, cinco años).

Se empleó un muestreo no probabilístico por conveniencia (Westreicher, 2022). Esta decisión se fundamentó en criterios curriculares y del desarrollo: de acuerdo con el Currículo Nacional de Educación Básica del Perú, los procesos de lectoescritura emergente se introducen formalmente únicamente en los niveles de cuatro y cinco años dentro de la educación inicial, lo que hace metodológicamente indefendible la inclusión de las aulas de dos y tres años para los propósitos de esta investigación. Los criterios de inclusión exigieron asistencia regular, continuidad de participación a lo largo del período de estudio y consentimiento informado por escrito de los padres o apoderados. Los niños matriculados en las aulas de dos y tres años fueron excluidos.

2.3 Instrumentos

Se desarrollaron tres escalas de valoración específicamente para este estudio, administradas directamente por la investigadora. Todos los instrumentos emplearon un formato de respuesta tipo Likert de cinco puntos ($1 = Nunca$ a $5 = Siempre$) y fueron diseñados para capturar manifestaciones conductuales observables de cada constructo en el contexto de la educación inicial.

Escala de Valoración de la Música Clásica (EVMC). Este instrumento comprendió 34 ítems organizados en tres dimensiones: Estimulación Neurocognitiva (ítems 1–8), Estimulación Sensorial y Emocional (ítems 9–21) y Desarrollo de la Escucha y la Comunicación (ítems 22–34). Su operacionalización buscó capturar respuestas conductuales observables ante la estimulación auditiva estructurada en niños pequeños, evitando deliberadamente la adopción acrítica de los amplios beneficios atribuidos al denominado efecto Mozart, postulados que la literatura neurocientífica contemporánea ha catalogado como un neuromito que requiere una operacionalización contextualmente delimitada (Campbell, 1998). La validez de contenido se estableció mediante juicio de expertos a través del coeficiente V de Aiken. La consistencia interna fue evaluada en un grupo piloto mediante el alfa de Cronbach ($\alpha = .995$), calculado sobre 24 de los 34 ítems; los 10 ítems restantes fueron excluidos automáticamente por el SPSS debido a varianza cero en las respuestas del piloto, lo que indica homogeneidad de respuesta en dicho subgrupo. Esta restricción se reconoce como una limitación psicométrica, dado que el coeficiente reportado no representa la totalidad del banco de ítems.

Escala de Valoración de la Neuroplasticidad (EVN). Este instrumento comprendió 44 ítems distribuidos en cuatro dimensiones: Adaptabilidad Neurológica (ítems 1–16), Desarrollo de Conexiones Neuronales (ítems 17–27), Estimulación Socioemocional y Cognitiva (ítems 28–38) y Plasticidad Funcional y Estructural (ítems 39–44). La estructura dimensional se sustentó en conceptualizaciones contemporáneas de la neuroplasticidad como la capacidad del cerebro para reorganizarse estructural y funcionalmente en respuesta a la experiencia (Lugo, 2024). La validez de contenido fue establecida mediante juicio de expertos. El alfa de Cronbach en la muestra piloto fue $\alpha = .996$, calculado sobre 15 de los 44 ítems tras la exclusión automática por SPSS de 29 ítems con varianza cero. Al igual que en la EVMC, este coeficiente debe interpretarse con cautela, ya que refleja únicamente un subconjunto del instrumento y puede sobreestimar la consistencia interna real de la escala.

Escala de Valoración de la Lectoescritura (EVL). Este instrumento comprendió 38 ítems organizados en cuatro dimensiones: Desarrollo de la Lectura (ítems 1–15), Desarrollo de la Escritura (ítems 16–22), Relación Fonema–Grafema (ítems 23–31) y Comprensión Lectora y Producción Escrita (ítems 32–38). Su construcción se fundamentó en la naturaleza progresiva e integradora de la adquisición temprana de la lectoescritura, que articula procesos fonológicos, gráficos y comprensivos (Cárate et al., 2023). La validez de contenido fue establecida mediante juicio de expertos. El alfa de Cronbach en la muestra piloto fue $\alpha = .911$, calculado sobre 10 de los 38 ítems, tras la exclusión de 28 ítems con varianza cero. Si bien este coeficiente está sujeto a la misma cautela psicométrica señalada anteriormente, se considera el más conservador e interpretivamente confiable de los tres, dada la proporción relativamente mayor de varianza de ítems retenida respecto al grado de exclusión.

Los rangos de puntuación de los tres instrumentos fueron organizados en tres niveles de desempeño (bajo, medio y alto) mediante la construcción de baremos, lo que posibilitó la clasificación ordinal de los participantes para el posterior análisis de regresión.

2.4 Procedimiento de recogida de datos

Los datos fueron recolectados mediante un procedimiento de observación estructurada utilizando las tres escalas de valoración descritas anteriormente (Feria et al., 2020). Todos los instrumentos fueron aplicados directamente por la investigadora en las aulas de la institución participante durante el horario escolar regular del año académico 2025. Dada la edad de los participantes —cuatro y cinco años—, las escalas fueron completadas a través de la observación conductual directa por parte de la investigadora, en lugar del autorregistro, lo que garantizó una recolección de datos estandarizada y apropiada para el nivel de desarrollo de los participantes en todas las aulas. Los tres instrumentos se aplicaron de forma secuencial en una única sesión de observación por aula. Con anterioridad al inicio de la recolección de datos, se

obtuvo el consentimiento informado por escrito de los padres o apoderados de todos los niños participantes.

2.5 Análisis estadístico

La validez de contenido de los tres instrumentos fue evaluada mediante el coeficiente V de Aiken, con ítems revisados por un panel de especialistas en educación inicial, neuroeducación y psicolingüística. La consistencia interna fue estimada mediante el alfa de Cronbach con el software IBM SPSS Statistics, reportando con transparencia las limitaciones psicométricas del piloto descritas en la sección 2.3.

Para examinar el efecto predictivo de la música clásica sobre la neuroplasticidad y la adquisición de la lectoescritura, se empleó la regresión logística ordinal como estrategia analítica principal. Esta elección se justificó frente a alternativas paramétricas como la correlación de Pearson o la regresión lineal, dado que las variables dependientes —neuroplasticidad y lectoescritura— fueron medidas en escalas ordinales de tres niveles de desempeño, lo que incumple los supuestos de nivel de medición intervalar y normalidad requeridos por dichas técnicas (Osada & Salvador, 2021). Para cada modelo de regresión se reportaron los coeficientes no estandarizados (B), los errores estándar, los estadísticos t , los valores p y los intervalos de confianza al 95%. El umbral de significancia estadística se fijó en $\alpha = .05$. Los modelos complementarios examinaron la contribución predictiva de cada dimensión individual de la música clásica sobre la neuroplasticidad y la lectoescritura, así como el papel predictivo de las dimensiones de la neuroplasticidad sobre los resultados lectoescritores. Todos los análisis fueron realizados con IBM SPSS Statistics.

2.6 Consideraciones éticas

El estudio fue conducido en pleno cumplimiento de los principios éticos que rigen la investigación con poblaciones infantiles. Se garantizó en todo momento la confidencialidad y el anonimato de los participantes, y los datos recolectados fueron utilizados exclusivamente

con fines académicos y científicos. El consentimiento informado por escrito fue obtenido de los padres o apoderados de todos los niños participantes con anterioridad al inicio de la recolección de datos. El estudio se adhirió a las disposiciones del Código de Ética de la Universidad César Vallejo, actuando en concordancia con los principios de autonomía, respeto, no maleficencia y confidencialidad. A lo largo del manuscrito se respetaron los estándares de propiedad intelectual mediante el uso apropiado de citas y paráfrasis.

3. Resultados

3.1 Justificación del análisis estadístico

Las tres variables del estudio —música clásica, neuroplasticidad y lectoescritura— fueron medidas mediante escalas de valoración cuyos puntajes totales fueron clasificados en tres niveles de desempeño (bajo, medio y alto) a través de baremos contruidos específicamente para cada instrumento. Dado que las variables dependientes operaron como variables ordinales de tres categorías ordenadas, la regresión logística ordinal constituyó la estrategia analítica apropiada para examinar el efecto predictivo entre variables. Las técnicas paramétricas convencionales como la correlación de Pearson o la regresión lineal múltiple fueron descartadas porque requieren nivel de medición intervalar o de razón y supuestos de distribución normal que las variables ordinales no satisfacen por definición, independientemente del resultado de cualquier prueba de normalidad (Osada & Salvador, 2021). En consecuencia, la elección del estadístico respondió al nivel de medición de las variables y a la naturaleza del diseño correlacional-causal del estudio, y no a criterios de normalidad distribucional. Para cada modelo se reportan los coeficientes no estandarizados (B), los errores estándar (DE), los estadísticos t , los valores p y los intervalos de confianza al 95% (IC 95%), con un umbral de significancia fijado en $\alpha = .05$. Todos los análisis fueron realizados con IBM SPSS Statistics.

3.2 Efecto de la música clásica sobre la neuroplasticidad y la lectoescritura

Los resultados de los modelos de regresión logística ordinal correspondientes al objetivo general se presentan en la Tabla 1, bloques 1 y 2.

Respecto al efecto de la música clásica sobre la neuroplasticidad, el modelo no alcanzó significancia estadística ($B = 0.055$, $DE = 0.041$, $t = 1.336$, $p = .186$, IC 95% $[-.027, .138]$). El intervalo de confianza incluyó el cero, lo que descarta la existencia de cualquier efecto real, sea positivo o negativo, atribuible a la exposición musical. Si bien el coeficiente presentó signo positivo, su magnitud reducida y su ausencia de significancia impiden toda interpretación sustantiva. Las diferencias observadas son atribuibles al azar y no a un efecto sistemático de la variable independiente. En consecuencia, se aceptó la hipótesis nula para este modelo, concluyendo que la música clásica no ejerció efecto predictivo significativo sobre la neuroplasticidad infantil en la muestra estudiada. Este hallazgo es coherente con la literatura neurocientífica contemporánea que ha catalogado como neuromito la creencia de que la escucha pasiva de música clásica induce modificaciones estructurales o funcionales en el cerebro infantil, postulados asociados al denominado efecto Mozart (Campbell, 1998).

Respecto al efecto de la música clásica sobre la lectoescritura, el modelo tampoco alcanzó significancia estadística ($B = 0.107$, $DE = 0.063$, $t = 1.690$, $p = .096$, IC 95% $[-.019, .234]$). Aunque el valor p se aproximó al umbral convencional de .05, esta proximidad no autoriza inferir la existencia de una tendencia emergente o un efecto incipiente. Desde el punto de vista inferencial, un intervalo de confianza que cruza el cero descarta cualquier efecto con independencia de la magnitud del coeficiente o de la distancia del valor p al umbral de significancia. Por tanto, se aceptó la hipótesis nula, concluyendo que la exposición pasiva a música clásica no generó variaciones significativas en el desempeño lector de los niños evaluados, ni en los procesos de decodificación, ni en la asociación fonema–grafema, ni en la comprensión, que son procesos fundamentales en la adquisición inicial de la lectoescritura (Cárate et al., 2023).

3.3 Influencia de las dimensiones de la música clásica sobre la neuroplasticidad

Los resultados del modelo de regresión logística ordinal con las tres dimensiones de la música clásica como predictores de la neuroplasticidad se presentan en la Tabla 1, bloque 3.

Ninguna de las tres dimensiones analizadas alcanzó significancia estadística: la estimulación neurocognitiva ($B = -0.404$, $DE = 0.847$, $t = -0.477$, $p = .635$, IC 95% $[-2.096, 1.288]$), la estimulación sensorial y emocional ($B = 0.948$, $DE = 0.828$, $t = 1.146$, $p = .256$, IC 95% $[-0.705, 2.601]$) y el desarrollo de la escucha y la comunicación ($B = -0.489$, $DE = 0.610$, $t = -0.801$, $p = .426$, IC 95% $[-1.708, 0.730]$) presentaron valores p superiores a .05 e intervalos de confianza que incluyeron el cero en todos los casos.

Desde un enfoque inferencial, la coexistencia de coeficientes con signos opuestos sin significancia estadística descarta la posibilidad de tendencias reales en cualquier dirección. Interpretar los coeficientes positivos como señales de influencia potencial constituiría una sobrelectura metodológicamente incorrecta, dado que la ausencia de significancia implica que las diferencias observadas son atribuibles al azar. La música clásica, desagregada en sus componentes dimensionales, tampoco ejerció influencia sobre la adaptabilidad neurológica, la formación de conexiones neuronales ni la plasticidad funcional o estructural de los participantes. En consecuencia, se aceptó la hipótesis nula, resultado que es consistente con la postura neurocientífica contemporánea que refuta que la escucha pasiva de música clásica induzca modificaciones observables en la arquitectura cerebral infantil (Lugo, 2024).

3.4 Influencia de las dimensiones de la música clásica sobre la lectoescritura

Los resultados del modelo de regresión logística ordinal con las tres dimensiones de la música clásica como predictores de la lectoescritura se presentan en la Tabla 1, bloque 4.

Ninguna dimensión alcanzó significancia estadística: la estimulación neurocognitiva ($B = 0.194$, $DE = 1.293$, $t = 0.150$, $p = .881$, IC 95% $[-2.389, 2.776]$), la estimulación sensorial y emocional ($B = 1.160$, $DE = 1.263$, $t = 0.918$, $p = .362$, IC 95% $[-1.364, 3.683]$) y el desarrollo

de la escucha y la comunicación ($B = -0.982$, $DE = 0.931$, $t = -1.055$, $p = .296$, IC 95% $[-2.843, 0.878]$) presentaron valores p ampliamente superiores a .05 e intervalos de confianza que incluyeron el cero.

La ausencia de significancia en todas las dimensiones impide sostener que alguna faceta de la estimulación musical —neurocognitiva, sensorial o comunicativa— favorezca el desarrollo lectoescritor, incluso en aquellas dimensiones cuyos coeficientes presentaron signo positivo. Este resultado es analíticamente relevante porque confirma que la adquisición de la lectoescritura se sustenta en procesos lingüísticos y pedagógicos sistemáticos como la conciencia fonológica, la decodificación y la asociación fonema–grafema, y no en la exposición pasiva a estímulos musicales (Cárate et al., 2023). En consecuencia, se aceptó la hipótesis nula para este objetivo.

3.5 Influencia de las dimensiones de la neuroplasticidad sobre la lectoescritura

Los resultados del modelo de regresión logística ordinal con las cuatro dimensiones de la neuroplasticidad como predictores de la lectoescritura se presentan en la Tabla 1, bloque 5.

Ninguna dimensión neuroplástica alcanzó significancia estadística: la adaptabilidad neurológica ($B = 1.803$, $DE = 1.186$, $t = 1.521$, $p = .133$, IC 95% $[-0.567, 4.173]$), el desarrollo de conexiones neuronales ($B = -0.626$, $DE = 1.665$, $t = -0.376$, $p = .708$, IC 95% $[-3.954, 2.701]$), la estimulación socioemocional y cognitiva ($B = -0.767$, $DE = 1.575$, $t = -0.487$, $p = .628$, IC 95% $[-3.914, 2.379]$) y la plasticidad funcional y estructural ($B = -0.542$, $DE = 2.203$, $t = -0.246$, $p = .806$, IC 95% $[-4.944, 3.860]$) presentaron valores p superiores a .05 e intervalos de confianza que incluyeron el cero en todos los casos.

Aunque la adaptabilidad neurológica presentó el coeficiente positivo de mayor magnitud en el modelo, su falta de significancia estadística impide cualquier interpretación causal o tendencial. Ninguna dimensión de la neuroplasticidad actuó como predictor del desempeño lectoescritor en la muestra analizada. Este resultado refuerza la evidencia acumulada de que la música

clásica no induce cambios neuroplásticos con repercusión observable en la adquisición de la lectura y la escritura, contribuyendo a desmontar el neuromito que atribuye a este tipo de estimulación un impacto directo en el aprendizaje infantil (Lugo, 2024). En consecuencia, se aceptó la hipótesis nula para este objetivo.

3.6 Asociación entre música clásica, neuroplasticidad y lectoescritura

Tomados en conjunto, los cinco modelos de regresión logística ordinal evidenciaron de manera consistente la ausencia de relaciones predictivas significativas entre la música clásica, la neuroplasticidad y la lectoescritura, tanto a nivel de variables globales como a nivel dimensional. Ninguno de los modelos examinados alcanzó el umbral de significancia estadística establecido ($\alpha = .05$), y la totalidad de los intervalos de confianza al 95% incluyeron el cero, descartando la existencia de efectos reales en cualquier dirección. Este patrón de resultados, coherente y replicado a través de los cinco modelos, constituye evidencia empírica acumulada que cuestiona la creencia de que la exposición pasiva a música clásica potencia el desarrollo neurológico y lectoescritor en la primera infancia. Los hallazgos contribuyen a la desmitificación empírica del denominado efecto Mozart en el ámbito educativo (Campbell, 1998), evidenciando que ni la neuroplasticidad ni la lectoescritura infantil responden de manera significativa a este tipo de estimulación auditiva pasiva, y que los procesos de desarrollo neurológico y lectoescritor en niños de cuatro y cinco años obedecen a mecanismos más complejos y específicos que la mera exposición a estímulos musicales clásicos (Lugo, 2024; Cárate et al., 2023).

Tabla 1. Resultados de los modelos de regresión logística ordinal: efecto de la música clásica sobre la neuroplasticidad y la lectoescritura

Modelo	Predictor	<i>B</i>	<i>DE</i>	<i>t</i>	<i>p</i>	IC 95% inf.	IC 95% sup.
1. MC → NP	Música clásica	.055	.041	1.336	.186	-.027	.138

2. MC → LE	Música clásica	.107	.063	1.690	.096	-.019	.234
3. Dim. MC → NP	Est. neurocognitiva	-.404	.847	-.477	.635	-2.096	1.288
	Est. sensorial y emocional	.948	.828	1.146	.256	-.705	2.601
	Escucha y comunicación	-.489	.610	-.801	.426	-1.708	.730
4. Dim. MC → LE	Est. neurocognitiva	.194	1.293	.150	.881	-2.389	2.776
	Est. sensorial y emocional	1.160	1.263	.918	.362	-1.364	3.683
	Escucha y comunicación	-.982	.931	-1.055	.296	-2.843	.878
5. Dim. NP → LE	Adapt. neurológica	1.803	1.186	1.521	.133	-.567	4.173
	Des. conexiones neuronales	-.626	1.665	-.376	.708	-3.954	2.701
	Est. socioemocional y cognitiva	-.767	1.575	-.487	.628	-3.914	2.379
	Plast. funcional y estructural	-.542	2.203	-.246	.806	-4.944	3.860

Fuente: elaboración propia

Nota. MC = Música clásica; NP = Neuroplasticidad; LE = Lectoescritura; DE = Desviación estándar; IC = Intervalo de confianza. Nivel de significancia $\alpha = .05$. Todos los modelos fueron estimados mediante regresión logística ordinal con IBM SPSS Statistics.

4. Discusión

Los cinco modelos de regresión logística ordinal examinados evidenciaron de manera consistente la ausencia de efectos predictivos significativos de la música clásica sobre la neuroplasticidad y la lectoescritura, tanto a nivel de variables globales como dimensional. Ningún modelo alcanzó el umbral de significancia estadística establecido y la totalidad de los intervalos de confianza al 95% incluyeron el cero, descartando la existencia de efectos reales en cualquier dirección. Este patrón uniforme, replicado a través de todos los modelos examinados, orienta la discusión hacia la interpretación de la ausencia de efecto como hallazgo sustantivo y teóricamente informativo, y no como resultado nulo trivial.

La ausencia de efecto predictivo de la música clásica sobre la neuroplasticidad coincide con la distinción que establece la literatura neurocientífica contemporánea entre exposición musical

activa y escucha pasiva. López (2024) y Wulandari et al. (2025) señalan que los efectos neuroplásticos documentados —incrementos en conectividad auditivo-motora, fortalecimiento de redes atencionales y cambios en materia gris— corresponden a formación musical activa y sistemática, no a la audición pasiva de piezas clásicas. Estudios que reportan cambios neuroplásticos positivos en niños, como los vinculados a práctica instrumental temprana, involucran componentes motores, atencionales y de retroalimentación sensoriomotora que la escucha pasiva no activa (Zaatar et al., 2023). La discrepancia entre esos hallazgos y los del presente estudio es, por tanto, explicable por el tipo de exposición musical examinado y no por una ausencia de relación entre música y cerebro en términos generales. Este resultado contribuye a desmontar empíricamente las afirmaciones más amplias asociadas al denominado efecto Mozart, cuya influencia en las prácticas educativas persiste pese a la falta de respaldo científico robusto para la modalidad pasiva (Campbell, 1998).

La ausencia de efecto predictivo sobre la lectoescritura es coherente con la evidencia metaanalítica de mayor escala disponible en el campo. Sala & Gobet (2020) encontraron que, al controlar la calidad metodológica de los estudios, la transferencia de la formación musical a habilidades académicas no musicales es esencialmente nula, hallazgo que el presente estudio replica en una muestra de educación inicial latinoamericana. Cancer & Antonietti (2022) reportan efectos inconsistentes de intervenciones musicales sobre la lectura, con ganancias más fiables en conciencia fonológica que en comprensión lectora global, lo que sugiere que incluso cuando existen efectos, estos son específicos y condicionados. La discrepancia con estudios que reportan relaciones positivas entre música y lectoescritura, como Ahokas et al. (2024) con entrenamiento rítmico en lectores con dificultades o Pantaleo et al. (2024) con alfabetización musical activa, se explica por condiciones de contorno decisivas: esos estudios involucran exposición activa, pedagógicamente integrada y dirigida a poblaciones con necesidades

específicas, condiciones sustancialmente distintas a la escucha pasiva de música clásica examinada en el presente estudio.

El patrón global de resultados encuentra su explicación más coherente en la teoría de la especificidad del aprendizaje, que postula que las ganancias cognitivas se circunscriben al dominio entrenado y que la transferencia a dominios no conectados pedagógicamente es mínima o inexistente (Sala & Gobet, 2020). Desde este marco, la ausencia de efectos no constituye un resultado nulo trivial sino evidencia de que la escucha pasiva de música clásica no activa los mecanismos cognitivos y neurales específicos requeridos para el desarrollo neuroplástico ni lectoescritor en niños de educación inicial.

El estudio presenta limitaciones que deben considerarse al interpretar sus hallazgos. El muestreo no probabilístico por conveniencia restringe la generalización de los resultados a poblaciones más amplias (Westreicher, 2022). El diseño no experimental impide establecer relaciones causales en sentido estricto. Las limitaciones psicométricas de los instrumentos ad hoc, descritas en la sección de Método, y la ausencia de estadísticos descriptivos de la muestra constituyen restricciones metodológicas adicionales que estudios futuros deberían subsanar. En consecuencia, se recomienda que investigaciones posteriores examinen intervenciones de formación musical activa mediante diseños experimentales con asignación aleatoria, muestras más amplias e instrumentos estandarizados internacionalmente, con el fin de establecer con mayor precisión las condiciones bajo las cuales la música puede constituir una herramienta pedagógica eficaz para el desarrollo neuroplástico y lectoescritor en la primera infancia (Sala & Gobet, 2020; Ahokas et al., 2024).

5. Conclusiones

Los resultados del presente estudio permiten establecer las siguientes conclusiones, derivadas directamente de la evidencia empírica obtenida mediante regresión logística ordinal en una muestra de 68 niños de cuatro y cinco años de educación inicial en Trujillo, Perú.

La música clásica, cuando se emplea exclusivamente como estímulo auditivo pasivo, no ejerce efecto predictivo significativo sobre la neuroplasticidad infantil. Este hallazgo se mantuvo consistente tanto a nivel de la variable global como en el análisis dimensional de sus componentes neurocognitivo, sensorial y comunicativo, ninguno de los cuales alcanzó el umbral de significancia estadística establecido.

De manera equivalente, la exposición pasiva a música clásica no predijo el desempeño lectoescriptor de los niños evaluados. La proximidad del valor p al umbral convencional en el modelo global no autoriza inferir la existencia de una tendencia emergente, dado que el intervalo de confianza al 95% incluyó el cero en todos los casos. La adquisición de la lectoescritura en la primera infancia responde a procesos lingüísticos y pedagógicos sistemáticos que la escucha pasiva de música clásica no activa.

Las dimensiones de la neuroplasticidad examinadas tampoco actuaron como predictores del desempeño lectoescriptor, lo que refuerza la conclusión de que la cadena causal propuesta por las afirmaciones populares del denominado efecto Mozart —música clásica → neuroplasticidad → lectoescritura— carece de respaldo empírico en esta población y modalidad de exposición.

Tomados en conjunto, estos hallazgos son coherentes con la teoría de la especificidad del aprendizaje y con la evidencia metaanalítica de mayor escala disponible en el campo, y contribuyen a la desmitificación empírica del efecto Mozart como neuromito en el ámbito de la educación inicial latinoamericana. Las prácticas educativas que emplean la escucha pasiva de música clásica como estrategia de potenciación cognitiva o lectoescritora no cuentan con respaldo empírico suficiente y deben ser evaluadas críticamente a la luz de la evidencia disponible.

Futuros estudios deberían examinar intervenciones de formación musical activa mediante diseños experimentales con asignación aleatoria, muestras probabilísticas más amplias,

instrumentos estandarizados internacionalmente y preservación sistemática de bases de datos primarias, con el fin de establecer con mayor precisión las condiciones bajo las cuales la música puede constituir una herramienta pedagógica eficaz para el desarrollo cognitivo en la primera infancia.

Referencias bibliográficas

- Ahokas, J., Saarikallio, S., Welch, G., Parviainen, T., & Louhivuori, J. (2024). Rhythm and reading: Connecting the training of musical rhythm to the development of literacy skills. *Early Childhood Education Journal*, 53, 999–1012. <https://doi.org/10.1007/s10643-024-01654-4>
- Campbell, D. (1998). *El efecto Mozart: Aprovecha el poder de la música para sanar el cuerpo, fortalecer la mente y liberar el espíritu creativo*. Urano.
- Cancer, A., & Antonietti, A. (2022). Music-based and auditory-based interventions for reading difficulties: A literature review. *Heliyon*, 8. <https://doi.org/10.1016/j.heliyon.2022.e09293>
- Cárate, M., Casillas, A., & Fauta, J. (2023). *[Datos completos pendientes de verificación por el autor]*.
- Couvignou, M., Peyre, H., Ramus, F., & Kolinsky, R. (2024). Do early musical impairments predict later reading difficulties? A longitudinal study of pre-readers with and without familial risk for dyslexia. *Developmental Science*, e13519. <https://doi.org/10.1111/desc.13519>
- Dees, L., & Cooper, P. (2025). Effects of a phonics-integrated music rhythm intervention on reading fluency and accuracy with children. *Frontiers in Human Neuroscience*, 19. <https://doi.org/10.3389/fnhum.2025.1636278>
- Díaz, F. (2020). *[Datos completos pendientes de verificación por el autor]*.

- Feria, H., Mantilla, M., & Mantecón, J. (2020). *[Datos completos pendientes de verificación por el autor]*.
- Haro, J., Chisag, J., Ruiz, M., & Caicedo, A. (2024). *[Datos completos pendientes de verificación por el autor]*.
- León, J. (2022). *[Datos completos pendientes de verificación por el autor]*.
- Liu, T., Gray-Bauer, H., Davison, K., & Zuk, J. (2024). Preschoolers' home music environment relates to their home literacy environment and parental self-efficacy. *PLOS ONE*, 19. <https://doi.org/10.1371/journal.pone.0313218>
- López, L. (2024). Neuroscience of music: Impact of musical stimulation on brain plasticity and education. A literature review. *Seminars in Medical Writing and Education*. <https://doi.org/10.56294/mw2024578>
- Lugo, M. (2024). *[Datos completos pendientes de verificación por el autor]*.
- OCDE. (2018). *Manual de Oslo: Guía para la recogida e interpretación de datos sobre innovación* (4.ª ed.). OCDE/Eurostat.
- Osada, J., & Salvador, J. (2021). *[Datos completos pendientes de verificación por el autor]*.
- Pantaleo, M., Arcuri, G., Manfredi, M., & Proverbio, A. (2024). Music literacy improves reading skills via bilateral orthographic development. *Scientific Reports*, 14. <https://doi.org/10.1038/s41598-024-54204-8>
- Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, 48, 1429–1441. <https://doi.org/10.3758/s13421-020-01060-2>
- Tianjiao, [nombre completo pendiente de verificación], Sanchis, I., Ramos, G., & Jiang, Y. (2025). Neuroplasticity mechanisms in early childhood piano education: A literature review from the perspective of educational neuroscience. *Journal of Sociology and Education*. <https://doi.org/10.63887/jse.2025.1.1.19>

- U, K., C, M., Rigby, A., As-Ferrer, I., Anand, H., Kunnavil, R., & Jaschke, A. (2025). The molecular basis of music-induced neuroplasticity in humans: A systematic review. *Neuroscience & Biobehavioral Reviews*, 175. <https://doi.org/10.1016/j.neubiorev.2025.106219>
- Ventura, J. (2019). *[Datos completos pendientes de verificación por el autor]*.
- Westreicher, G. (2022). *[Datos completos pendientes de verificación por el autor]*.
- Wulandari, I., Prasetya, R., Satriawan, W., Semadi, I., & Harahap, H. (2025). The role of musical hobbies in enhancing children's learning concentration. *Jurnal Biologi Tropis*. <https://doi.org/10.29303/jbt.v25i3.9185>
- Zaatar, M., Alhakim, K., Enayeh, M., & Tamer, R. (2023). The transformative power of music: Insights into neuroplasticity, health, and disease. *Brain, Behavior, & Immunity - Health*, 35. <https://doi.org/10.1016/j.bbih.2023.100716>

Capítulo 6

Relación de la neuroeducación y el uso de tecnologías educativas con realidad aumentada

Patricia Roxana Ibáñez Almonacid⁷

RESUMEN

Objetivo: El presente estudio examinó la relación entre la neuroeducación y el uso de tecnologías educativas con realidad aumentada en estudiantes de la especialidad de informática de educación secundaria de una institución educativa pública de Lima, Perú, durante el año académico 2025.

Método: Se adoptó un enfoque cuantitativo con diseño no experimental de corte transversal correlacional. La muestra fue censal ($N = 40$; 14 mujeres, 26 varones; rango etario 14–16 años). Ambas variables fueron medidas mediante cuestionarios de 18 ítems en escala Likert de cinco puntos, validados por juicio de tres expertos. La plataforma de realidad aumentada ARTIVIVE fue implementada como contexto pedagógico deliberado del estudio. La verificación del supuesto de normalidad mediante las pruebas de Kolmogorov-Smirnov y Shapiro-Wilk ($p > .05$ en ambas variables) habilitó el uso del coeficiente de correlación de Pearson para el contraste de la hipótesis general y ocho hipótesis específicas ($\alpha = .05$). Los análisis fueron realizados con IBM SPSS Statistics v27.

Resultados: Se obtuvo una correlación positiva, estadísticamente significativa y de magnitud grande entre neuroeducación y realidad aumentada ($r = .818$, $p < .001$, IC95% [.680, .900]). Los ocho contrastes dimensionales —neuroeducación con las dimensiones de realidad

⁷ Especialista en Neuroeducación por la Universidad César Vallejo, Perú. E-mail: pribanez@ucvvirtual.edu.pe pria_2007@hotmail.com, ORCID: <https://orcid.org/0000-0003-1142-9235>

aumentada (Colaboración: $r = .805$; Personalización: $r = .690$; Inmersión: $r = .591$; Interacción Real-Virtual: $r = .570$) y realidad aumentada con las dimensiones de neuroeducación (Voluntad: $r = .811$; Aprendizaje: $r = .675$; Actitud: $r = .657$; Afectividad: $r = .646$)— fueron igualmente positivos, significativos ($p < .001$) y de efecto grande, con intervalos de confianza al 95% que no cruzan el cero.

Conclusión: Los hallazgos evidencian una asociación sistemática y robusta entre los principios neuroeducativos y las dimensiones operacionales de la realidad aumentada en el contexto estudiado. Estos resultados aportan evidencia empírica dimensionalmente desagregada que la literatura previa no provee, con implicaciones para el diseño instruccional de entornos de aprendizaje tecnológicamente enriquecidos en educación secundaria técnica.

Palabras clave: neuroeducación, realidad aumentada, tecnología educativa, educación secundaria, correlación, aprendizaje, motivación.

1. Introducción

La expansión acelerada de tecnologías emergentes en los entornos educativos ha intensificado la necesidad de comprender cómo estas interactúan con los procesos cognitivos y afectivos que subyacen al aprendizaje (Jiang et al., 2025). En este contexto, la neuroeducación —campo que integra neurociencia, psicología y pedagogía para identificar los principios cerebrales del aprendizaje efectivo— ofrece un marco conceptual para evaluar el potencial pedagógico de dichas tecnologías (De La Cruz et al., 2025). Entre estas, la realidad aumentada (AR) destaca por un conjunto de características operacionales —inmersión, personalización, colaboración e interactividad— que presentan una alineación conceptual con los principios neuroeducativos de atención sostenida, motivación intrínseca, regulación emocional y consolidación de la memoria (Amores-Valencia et al., 2022). No obstante, la evidencia empírica sobre la naturaleza

y magnitud de esta alineación en contextos de educación secundaria técnica permanece incompleta, lo que justifica una indagación sistemática sobre la relación entre ambas.

La neuroeducación comprende el aprendizaje como un proceso en el que la activación emocional, la disposición atencional y la voluntad del sujeto median la consolidación del conocimiento (Howard-Jones, 2011). En contextos de aula, estos principios se operacionalizan en cuatro dimensiones observables: aprendizaje, entendido como el proceso cognitivo de adquisición de conocimientos; afectividad, como la respuesta emocional ante los estímulos pedagógicos; actitud, como la disposición organizada de la mente hacia la tarea; y voluntad, como la decisión autónoma de comprometerse con el proceso formativo (Mora, 2021). Esta operacionalización permite examinar los principios neuroeducativos mediante instrumentos aplicables en entornos escolares reales, sin requerir medición neurofisiológica directa. Las intervenciones tecnológicas alineadas con estos principios han demostrado potencial para mejorar el rendimiento académico, la motivación intrínseca y el bienestar emocional de los estudiantes (De La Cruz et al., 2025).

La realidad aumentada, al superponer contenido digital sobre el entorno físico en tiempo real, opera mediante características que se corresponden con las dimensiones neuroeducativas descritas: la inmersión favorece la focalización atencional que subyace al aprendizaje activo; la personalización respeta los ritmos individuales de procesamiento y fortalece la actitud hacia la tarea; la colaboración entre pares activa mecanismos sociales que refuerzan la afectividad y el vínculo con el aprendizaje; y la interacción real-virtual estimula la voluntad al proporcionar experiencias novedosas y significativas (Cabero, 2016; Luque, 2019). Esta correspondencia conceptual, respaldada por evidencia teórica y empírica reciente (Lin y Yu, 2023; Koç y Kanadlı, 2025), fundamenta la hipótesis de que neuroeducación y AR presentarán asociaciones positivas y estadísticamente significativas. El estado de la evidencia empírica disponible, examinado a continuación, precisa los alcances y los límites de esta hipótesis.

Revisiones sistemáticas y metaanálisis recientes documentan que el uso de AR en educación secundaria produce efectos positivos sobre la atención, la motivación y el rendimiento académico, particularmente en ciencias naturales y matemáticas, y de manera más pronunciada cuando la AR se integra en estrategias pedagógicas situadas o basadas en tareas (Amores-Valencia et al., 2022; Lin y Yu, 2023; Lampropoulos et al., 2022). Los efectos sobre procesos afectivos —disfrute, actitud positiva ante el aprendizaje, reducción de ansiedad— son igualmente documentados con consistencia creciente en la literatura reciente (Sakr y Abdullah, 2024; Jaballudin y Khalid, 2024). En el plano de la integración explícita entre neuroeducación y tecnología, revisiones recientes identifican que las intervenciones alineadas con principios neuroeducativos —personalización, retroalimentación inmediata, contenido inmersivo e interactivo— producen mejoras en motivación intrínseca, autorregulación emocional y bienestar del estudiante (De La Cruz et al., 2025; Crogman et al., 2025; Koç y Kanadlı, 2025). Sin embargo, estos modelos son predominantemente conceptuales o experimentales, y no examinan la relación entre neuroeducación y AR como variables medidas de forma simultánea en el mismo grupo de estudiantes.

La literatura presenta dos limitaciones que orientan el diseño del presente estudio. En primer lugar, no se identifican estudios que examinen la asociación entre dimensiones específicas de la neuroeducación —aprendizaje, afectividad, actitud, voluntad— y dimensiones operacionales de la AR —inmersión, personalización, colaboración, interacción real-virtual— de manera simultánea y dimensionalmente desagregada (Amores-Valencia et al., 2022; Koumpouros, 2024). En segundo lugar, la evidencia empírica proveniente de contextos latinoamericanos de educación secundaria técnica es escasa, lo que restringe la aplicabilidad de los hallazgos internacionales a realidades educativas con condiciones de implementación tecnológica distintas (Sakr y Abdullah, 2024).

Estas limitaciones configuran un gap específico: se desconoce en qué medida y con qué magnitud se asocian las dimensiones de la neuroeducación con las dimensiones operacionales de la AR en estudiantes de educación secundaria técnica. Responder a esta pregunta requiere, como primer paso metodológico necesario, establecer si la asociación existe y cuál es su magnitud antes de diseñar intervenciones basadas en ella. Un diseño correlacional es, por tanto, la estrategia apropiada para este momento del desarrollo del campo. El contexto peruano, donde el acceso a internet desde dispositivos móviles alcanza al 86% de los estudiantes de entre 6 y 17 años (INEI, 2025), pero donde la evidencia sobre integración pedagógica de AR en aulas técnicas de secundaria es prácticamente inexistente, ofrece un caso de implementación con recursos institucionales propios que amplía la base geográfica de la evidencia disponible.

El presente estudio tiene como objetivo examinar la relación entre la neuroeducación y el uso de tecnologías educativas con realidad aumentada en estudiantes de la especialidad de informática de educación secundaria de una institución educativa de Lima, analizando las asociaciones entre cada variable y las dimensiones operacionales de la otra durante el año académico 2025. Se hipotetiza que existe una asociación positiva y estadísticamente significativa entre ambas variables, en consonancia con la evidencia acumulada sobre los efectos de la AR en procesos cognitivos y afectivos del aprendizaje (Lin y Yu, 2023). De manera específica, se hipotetiza que la neuroeducación se asociará positiva y significativamente con cada dimensión de la realidad aumentada —inmersión, personalización, colaboración e interacción con el entorno real y virtual—, y que la realidad aumentada se asociará positiva y significativamente con cada dimensión de la neuroeducación —aprendizaje, afectividad, actitud y voluntad.

3. Método

3.1 Diseño y tipo de investigación

El presente estudio adoptó un enfoque cuantitativo, dado que la recolección y el análisis de datos se sustentaron en mediciones numéricas y procedimientos estadísticos orientados a describir y examinar asociaciones entre fenómenos observables (Hernández-Sampieri y Mendoza, 2018; Creswell y Creswell, 2018). La investigación se clasifica como de tipo básica, en tanto su propósito es generar conocimiento empírico sobre la relación entre la neuroeducación y las tecnologías educativas basadas en realidad aumentada, sin pretender resolver un problema aplicado específico ni introducir manipulación experimental deliberada (Hernández-Sampieri y Mendoza, 2018). En consecuencia, se adoptó un diseño no experimental de corte transversal correlacional, en el que ambas variables fueron medidas en un único momento temporal y se examinó su grado de asociación sin presuponer direccionalidad causal.

Cabe precisar que, si bien el diseño es no experimental, el uso de la plataforma ARTIVIVE fue introducido específicamente en el marco de este estudio como contexto pedagógico deliberado en el que se midieron ambas variables. Ello no constituye manipulación experimental ni asignación aleatoria, sino que sitúa la investigación en un contexto de implementación tecnológica intencional. Asimismo, dada la naturaleza simétrica del análisis correlacional, las asociaciones fueron examinadas en dos direcciones complementarias: la neuroeducación en relación con las cuatro dimensiones de la tecnología educativa con realidad aumentada, y la realidad aumentada en relación con las cuatro dimensiones de la neuroeducación. No se presupone dirección causal entre las variables (Cohen, 1988).

3.2 Participantes

El estudio contó con la participación de 40 estudiantes de educación secundaria (14 mujeres, 26 varones) con edades comprendidas entre los 14 y los 16 años, matriculados en la especialidad técnica de informática del VII ciclo de la Educación Básica Regular peruana (EBR), correspondiente a los tres últimos años de la educación secundaria obligatoria con

componente de formación técnica integrada. Los participantes pertenecían a una institución educativa pública de Lima durante el año académico 2025. Para lectores de contextos internacionales, el VII ciclo de la EBR equivale a los grados 10.º a 12.º de la educación secundaria, e integra contenidos académicos generales con formación técnico-vocacional dentro del currículo regular.

Se empleó un muestreo censal por conveniencia, mediante el cual la totalidad de la población accesible fue incluida sin aplicar criterios de exclusión, eliminando así el error de muestreo dentro del contexto estudiado (Hernández-Sampieri y Mendoza, 2018). La adecuación del tamaño muestral fue evaluada conforme a los criterios de análisis de potencia establecidos por Cohen (1988): para detectar una correlación de magnitud grande ($r \geq .50$, consistente con la literatura previa sobre entornos de aprendizaje mediados por realidad aumentada), con un nivel de significancia de $\alpha = .05$ y una potencia estadística de $1 - \beta = .80$, el tamaño muestral mínimo requerido es de $n = 29$. El presente estudio, con $n = 40$, supera dicho umbral, garantizando potencia estadística adecuada para los análisis planificados.

3.3 Instrumentos

Cuestionario de Neuroeducación. La variable neuroeducación fue evaluada mediante un instrumento adoptado de Dorregaray Limachi (2020), desarrollado y validado originalmente en el marco de una investigación de maestría. El instrumento fue adoptado sin modificación estructural y aplicado al contexto específico de estudiantes de informática en el nivel secundario de Lima. Comprende 18 ítems distribuidos en cuatro dimensiones —Aprendizaje, Afectividad, Actitud y Voluntad (véase Marco Teórico para las definiciones conceptuales de cada dimensión)— valorados mediante una escala Likert de cinco puntos que oscila entre 1 (*Nunca*) y 5 (*Siempre*). La validez de contenido para el presente estudio fue establecida mediante juicio de tres expertos en investigación educativa, neurociencia, psicología y tecnología educativa, quienes evaluaron cada ítem en términos de suficiencia, claridad,

coherencia y relevancia. La confiabilidad por consistencia interna fue estimada mediante el coeficiente alfa de Cronbach ($\alpha = 0.89$).

Cuestionario de Realidad Aumentada. La variable realidad aumentada fue medida mediante un instrumento construido ad hoc para el presente estudio, sustentado en un marco teórico documentado. La conceptualización general de la variable se fundamenta en Luque (2019) y Cabero (2016), mientras que la definición operacional de cada dimensión se apoya en Chang (2021) para la dimensión de Inmersión, Guamán-Chisag (2023) para la dimensión de Colaboración, Terán-Pazmiño et al. (2024) para la dimensión de Personalización, y Rodríguez et al. (2017) para la dimensión de Interacción con el Entorno Real y Virtual (véase Marco Teórico para las definiciones conceptuales). El instrumento comprende 18 ítems valorados con la misma escala Likert de cinco puntos. La validez de contenido fue establecida mediante el mismo procedimiento de juicio de expertos descrito previamente. La confiabilidad por consistencia interna fue estimada mediante alfa de Cronbach ($\alpha = 0.91$).

Condiciones de administración. Ambos instrumentos fueron administrados en formato auto-reporte presencial, bajo condiciones de anonimato y confidencialidad, con un tiempo estimado de cumplimentación de cinco minutos por cuestionario. La aplicación se realizó de forma consecutiva e inmediatamente posterior a la experiencia de los estudiantes con la plataforma ARTIVIVE, de modo que las respuestas reflejaran percepciones sustentadas en el contacto directo y real con la tecnología.

3.4 Herramienta tecnológica: la plataforma ARTIVIVE

ARTIVIVE (Artivive GmbH, Viena, Austria) es una plataforma de realidad aumentada basada en reconocimiento de imagen (*image-based AR*) que permite superponer capas de contenido digital —incluyendo video, animación, audio y objetos tridimensionales— sobre una imagen física disparadora (*trigger image*), visualizable en tiempo real mediante la aplicación móvil ARTIVIVE (disponible para iOS y Android) o a través de WebAR, sin requerir conocimientos

de programación ni hardware especializado (Artivive GmbH, 2025). La plataforma cuenta con un paquete educativo específico para instituciones y es utilizada activamente en más de 190 países en contextos educativos de diversa índole.

La plataforma fue introducida específicamente para este estudio como el contexto pedagógico en el que ambas variables fueron operacionalizadas y medidas. Los estudiantes accedieron al editor web ARTIVIVE Bridge desde los recursos informáticos de la institución educativa. El flujo de trabajo pedagógico siguió cinco fases secuenciales: (1) definición del objetivo de aprendizaje; (2) diseño y carga de la imagen disparadora —póster, infografía o producto de diseño gráfico—; (3) creación y superposición de capas de contenido digital sobre la imagen disparadora, incluyendo recursos audiovisuales, animaciones y elementos de diseño digital; (4) publicación y visualización del proyecto de realidad aumentada mediante la aplicación móvil; y (5) socialización e interacción colaborativa entre pares en torno a los proyectos elaborados. Este flujo activa operacionalmente las cuatro dimensiones de la variable realidad aumentada medidas en el estudio —Inmersión, Personalización, Colaboración e Interacción con el Entorno Real y Virtual— en un entorno de aprendizaje auténtico y tecnológicamente enriquecido (Artivive GmbH, 2025; Triviño et al., 2023; Barroso, 2022).

3.5 Procedimiento y consideraciones éticas

La recolección de datos siguió un protocolo estructurado en cuatro fases. En primer lugar, se obtuvo la autorización formal de la dirección de la institución educativa previo a cualquier contacto con los estudiantes o sus familias. En segundo lugar, se convocó a los padres y tutores legales de todos los participantes —dado que la población de estudio está constituida íntegramente por menores de edad— y se obtuvo de cada uno el consentimiento informado por escrito, tras haberles informado cabalmente sobre el propósito del estudio, sus procedimientos y las garantías de confidencialidad. En tercer lugar, se implementaron las actividades pedagógicas con la plataforma ARTIVIVE en el aula de informática de la institución, haciendo

uso de sus recursos computacionales disponibles. En cuarto lugar, ambos cuestionarios fueron administrados de forma presencial y consecutiva en formato auto-reporte anónimo, inmediatamente después de la experiencia de los estudiantes con la plataforma, garantizando así que las respuestas reflejaran el contacto experiencial directo con la tecnología.

El estudio fue conducido en estricta conformidad con el Código de Ética en Investigación (UCV, 2022) y el Reglamento de Trabajos Conducentes a Grados y Títulos (UCV, 2025). La participación fue enteramente voluntaria; tanto los estudiantes como sus representantes legales fueron informados de su derecho a retirarse del estudio en cualquier momento sin consecuencia alguna. La totalidad de los datos fue utilizada exclusivamente con fines científicos y almacenada de forma segura. Se garantizó el anonimato y la confidencialidad en todas las fases del estudio. El diseño de la investigación no implicó riesgo físico, psicológico ni emocional para ninguno de los participantes.

3.6 Estrategia de análisis de datos

Se calcularon estadísticos descriptivos —media, mediana y desviación estándar— para ambas variables con el fin de caracterizar el perfil general de la muestra. Con carácter previo al análisis inferencial, se verificó el supuesto de normalidad de las distribuciones de puntuaciones mediante dos pruebas complementarias, apropiadas para el tamaño muestral del presente estudio ($n = 40$): la prueba de Kolmogorov-Smirnov, recomendada para muestras con $n > 30$, y la prueba de Shapiro-Wilk, reconocida por su sensibilidad en muestras con $n < 50$ (Field, 2018). El criterio de decisión fue establecido a priori: un valor de significancia $p > .05$ en ambas pruebas indicaría una distribución suficientemente próxima a la normalidad para justificar el uso de estadística paramétrica.

Confirmada la normalidad de ambas variables, y en consonancia con la evidencia metodológica que respalda el uso del coeficiente de correlación de Pearson con escalas Likert de cinco puntos bajo condiciones de normalidad (Norman, 2010), se seleccionó el estadístico r de Pearson para

contrastar la hipótesis general y las ocho hipótesis específicas del estudio. El nivel de significancia fue establecido en $\alpha = .05$. La magnitud de cada coeficiente de correlación fue interpretada conforme a las convenciones de Cohen (1988): efecto pequeño ($r \geq .10$), efecto mediano ($r \geq .30$) y efecto grande ($r \geq .50$). Todos los análisis fueron realizados con el programa IBM SPSS Statistics, versión 27. La justificación del estadístico elegido se establece en esta sección de manera única; en la sección de Resultados se reportan exclusivamente los valores r y p sin reiterar los fundamentos de la decisión analítica.

4. Resultados

4.1 Estadísticos descriptivos

Los estadísticos descriptivos de ambas variables se presentan en la Tabla 1. La variable Realidad Aumentada obtuvo una media de 81.33 puntos ($Mdn = 82.00$; $DE = 6.379$; $N = 40$). La proximidad entre la media y la mediana indica una distribución de puntuaciones aproximadamente simétrica, mientras que la desviación estándar refleja una dispersión moderada en torno al valor central. La variable Neuroeducación registró una media de 77.73 puntos ($Mdn = 78.00$; $DE = 7.713$; $N = 40$). De manera análoga a la variable anterior, la cercanía entre ambas medidas de tendencia central sugiere una distribución centrada, si bien la desviación estándar algo mayor indica una dispersión ligeramente superior entre los participantes. En ambos casos, las puntuaciones se ubican en la mitad superior de la escala de medición, lo que describe el perfil observado sin anticipar interpretaciones sobre su significado pedagógico.

Tabla 1. Estadísticos descriptivos de las variables Realidad Aumentada y Neuroeducación

Variable	N válido	Media	Mediana	Desviación estándar
Realidad Aumentada	40	81.33	82.00	6.379
Neuroeducación	40	77.73	78.00	7.713

Nota. Los valores corresponden a las puntuaciones totales obtenidas mediante los respectivos cuestionarios de 18 ítems en escala Likert de cinco puntos (rango teórico: 18–90).

4.2 Verificación del supuesto de normalidad

Con el fin de determinar el estadístico inferencial apropiado, se verificó el supuesto de normalidad de ambas distribuciones mediante las pruebas de Kolmogorov-Smirnov y Shapiro-Wilk, ambas pertinentes para el tamaño muestral del presente estudio ($n = 40$). Para la variable Realidad Aumentada, los resultados fueron $K-S(40) = .138$, $p = .053$ y $S-W(40) = .947$, $p = .058$. Para la variable Neuroeducación, se obtuvieron $K-S(40) = .085$, $p = .200$ y $S-W(40) = .969$, $p = .345$. En todos los casos, los valores de significancia superan el umbral de $p > .05$, indicando que ninguna de las dos distribuciones se desvía significativamente de la normalidad. En consecuencia, el supuesto paramétrico queda satisfecho y se procedió con el coeficiente de correlación de Pearson conforme al criterio establecido en la sección de Método.

4.3 Hipótesis general: correlación entre Neuroeducación y Realidad Aumentada

El contraste de la hipótesis general del estudio arrojó una correlación positiva, estadísticamente significativa y de magnitud grande entre las variables Neuroeducación y Realidad Aumentada: $r(38) = .818$, $p < .001$, IC95% [.680, .900] (Cohen, 1988). Dado que ningún extremo del intervalo de confianza cruza el cero, la asociación positiva entre ambas variables es estadísticamente distinguible de cero con un 95% de confianza. La hipótesis general del estudio queda así confirmada. Las subsecciones 4.4 y 4.5 desagregan este resultado a nivel dimensional, examinando las asociaciones específicas en las dos direcciones correlacionales declaradas en el Método.

Tabla 2. Correlación de Pearson entre Neuroeducación y Realidad Aumentada

	Realidad Aumentada
Neuroeducación — Correlación de Pearson	.818**
Sig. bilateral	< .001

IC95%	[.680, .900]
N	40

Nota. ** $p < .001$.

4.4 Hipótesis específicas — Bloque 1: Neuroeducación y dimensiones de Realidad Aumentada

Las cuatro hipótesis específicas del primer bloque examinan la asociación entre la variable Neuroeducación y cada dimensión operacional de la Realidad Aumentada. Los resultados se presentan en orden descendente de magnitud del efecto, orden que difiere del empleado en el instrumento original, y se sintetizan en la Tabla 3.

La asociación más alta del bloque corresponde a la dimensión Colaboración: $r(38) = .805$, $p < .001$, IC95% [.663, .895], clasificada como efecto grande (Cohen, 1988). Le sigue la dimensión Personalización: $r(38) = .690$, $p < .001$, IC95% [.481, .823], igualmente clasificada como efecto grande (Cohen, 1988). La dimensión Inmersión registró $r(38) = .591$, $p < .001$, IC95% [.342, .762], con efecto grande próximo al umbral inferior de dicha categoría (Cohen, 1988). La asociación más baja del bloque corresponde a la dimensión Interacción con el Entorno Real y Virtual: $r(38) = .570$, $p < .001$, IC95% [.316, .748], efecto grande (Cohen, 1988). En los cuatro casos, los intervalos de confianza no cruzan el cero, confirmando la consistencia estadística de las asociaciones positivas observadas.

Tabla 3. Correlaciones de Pearson entre Neuroeducación y dimensiones de Realidad Aumentada

Dimensión de Realidad Aumentada	r	IC95%	Sig. bilateral	N
Colaboración	.805**	[.663, .895]	< .001	40
Personalización	.690**	[.481, .823]	< .001	40
Inmersión	.591**	[.342, .762]	< .001	40
Interacción Real-Virtual	.570**	[.316, .748]	< .001	40

Nota. Las dimensiones se presentan en orden descendente de magnitud del coeficiente de correlación. ** $p < .001$.

4.5 Hipótesis específicas — Bloque 2: Realidad Aumentada y dimensiones de Neuroeducación

Las cuatro hipótesis específicas del segundo bloque examinan la asociación entre la variable Realidad Aumentada y cada dimensión operacional de la Neuroeducación. El patrón general de asociaciones es positivo y estadísticamente significativo en las cuatro dimensiones. Los resultados se presentan en orden descendente de magnitud del efecto y se sintetizan en la Tabla 4.

La dimensión Voluntad registró la asociación más alta del bloque: $r(38) = .811$, $p < .001$, IC95% [.669, .897], efecto grande (Cohen, 1988). La dimensión Aprendizaje obtuvo $r(38) = .675$, $p < .001$, IC95% [.461, .813], efecto grande (Cohen, 1988). La dimensión Actitud registró $r(38) = .657$, $p < .001$, IC95% [.433, .802], efecto grande (Cohen, 1988). La dimensión Afectividad presentó la asociación más baja del bloque: $r(38) = .646$, $p < .001$, IC95% [.419, .795], efecto grande (Cohen, 1988). Al igual que en el bloque anterior, ningún intervalo de confianza cruza el cero, lo que confirma la consistencia estadística de todas las asociaciones en dirección positiva.

Tabla 4. Correlaciones de Pearson entre Realidad Aumentada y dimensiones de Neuroeducación

Dimensión de Neuroeducación	r	IC95%	Sig. bilateral	N
Voluntad	.811**	[.669, .897]	< .001	40
Aprendizaje	.675**	[.461, .813]	< .001	40
Actitud	.657**	[.433, .802]	< .001	40
Afectividad	.646**	[.419, .795]	< .001	40

Nota. Las dimensiones se presentan en orden descendente de magnitud del coeficiente de correlación. ** $p < .001$.

4.6 Síntesis del patrón global de resultados

La totalidad de los nueve coeficientes de correlación obtenidos —uno correspondiente a la hipótesis general y ocho a las hipótesis específicas— son positivos, estadísticamente

significativos ($p < .001$) y clasificables como efecto grande conforme a las convenciones de Cohen (1988). Los valores oscilan entre $r = .570$ y $r = .818$ en el primer bloque de hipótesis específicas, y entre $r = .646$ y $r = .811$ en el segundo. En ningún caso los intervalos de confianza al 95% cruzan el cero, lo que confirma que todas las asociaciones identificadas son estadísticamente distinguibles de cero en la población con un margen de confianza del 95%. Este patrón homogéneo y de alta magnitud constituye el sustrato empírico sobre el que se apoya la discusión que se presenta a continuación.

5. Discusión

Los resultados del presente estudio evidencian un patrón uniforme de asociaciones positivas, estadísticamente significativas y de magnitud grande entre la neuroeducación y las tecnologías educativas con realidad aumentada. La correlación general entre ambas variables ($r = .818$) establece el patrón dominante, que se mantiene consistente en los ocho contrastes dimensionales, con valores que oscilan entre $r = .570$ y $r = .811$. Dentro de este patrón, las dimensiones Voluntad y Colaboración presentan las asociaciones más altas, mientras que Inmersión e Interacción con el Entorno Real y Virtual registran las más moderadas, aunque igualmente clasificables como efecto grande. Este diferencial de magnitudes —y no solo la direccionalidad positiva general— constituye el elemento más informativo del conjunto de resultados y el que requiere interpretación teórica más precisa.

La Self-Determination Theory (SDT; Deci y Ryan, 1985) ofrece el marco más adecuado para explicar el patrón diferencial observado. La SDT postula que la motivación intrínseca se activa cuando el entorno satisface tres necesidades psicológicas básicas: autonomía, competencia y relación social. Las dimensiones con asociaciones más altas en este estudio —Voluntad y Colaboración— corresponden precisamente a los satisfactores de autonomía y relación: la creación de proyectos AR propios activa la experiencia de agencia personal, mientras que el trabajo entre pares activa el sentido de pertenencia y vínculo social. Desde esta lectura, las

correlaciones más robustas no son aleatorias sino estructuralmente coherentes con los principios de la SDT. Las dimensiones con asociaciones más moderadas —Inmersión e Interacción Real-Virtual— involucran un procesamiento cognitivo de mayor demanda técnica, cuyo beneficio neuroeducativo podría ser más gradual a medida que los estudiantes desarrollan competencia en el manejo de la plataforma. La Multimedia Learning Theory (Mayer, 2001), que sostiene que los entornos multisensoriales optimizan el procesamiento cognitivo al distribuir información entre canales visual y auditivo, ofrece un marco complementario para estas dimensiones, aunque su activación plena podría requerir mayor experiencia con la tecnología que la disponible en el período estudiado. Esta última interpretación es tentativa y requiere verificación en estudios longitudinales.

Los resultados convergen con la evidencia acumulada en la literatura sobre efectos positivos de la AR sobre la motivación, la actitud y el aprendizaje en contextos de educación secundaria (Amores-Valencia et al., 2022; Sakr y Abdullah, 2024). En particular, la asociación entre AR y las dimensiones afectivo-volitivas de la neuroeducación es consistente con los hallazgos de De La Cruz et al. (2025), quienes identifican que las intervenciones tecnológicas alineadas con principios neuroeducativos se asocian con mejoras en motivación intrínseca y autorregulación emocional. Asimismo, la asociación entre AR y la dimensión Aprendizaje es coherente con los efectos positivos sobre resultados de aprendizaje documentados en entornos AR interactivos (Lin y Yu, 2023). Sin embargo, estos estudios examinan los efectos de la AR sobre variables motivacionales o cognitivas de manera global, sin desagregarlos por dimensiones específicas de un constructo neuroeducativo. El presente estudio avanza en esta dirección al revelar que la magnitud de la asociación varía según la dimensión considerada, lo que la literatura previa no había documentado de forma sistemática. En este sentido, el estudio responde parcialmente al señalamiento de Koumpouros (2024) sobre la necesidad de investigaciones con diseños más precisos y medidas más específicas en el campo de la AR educativa. Una condición que los

resultados sugieren —aunque sin poder confirmarla causalmente— es que la estructuración pedagógica del uso de AR puede ser relevante para la magnitud de las asociaciones observadas, en línea con lo sostenido por Lampropoulos et al. (2022) respecto a la superioridad de la AR integrada en marcos pedagógicos sólidos frente a su uso aislado.

Los resultados tienen implicaciones directas para el diseño instruccional en entornos de informática técnica secundaria: dado que las dimensiones volitiva y colaborativa presentan las asociaciones más altas, las actividades AR deberían priorizar la creación autónoma de proyectos y el trabajo entre pares por sobre la exposición pasiva a contenidos inmersivos. Como limitación principal, el diseño correlacional no permite inferir causalidad ni generalizar los hallazgos más allá del contexto estudiado. Investigaciones futuras deberían avanzar hacia diseños experimentales o cuasi-experimentales que examinen el efecto directo de la AR sobre las dimensiones neuroeducativas con mayor sensibilidad a las asociaciones aquí identificadas, e incorporar muestras más amplias que fortalezcan la potencia estadística y la transferibilidad de los resultados.

Referencias bibliográficas

- Amores-Valencia, A., Burgos, D., y Branch-Bedoya, J. (2022). Influence of motivation and academic performance in the use of Augmented Reality in education. A systematic review. *Frontiers in Psychology*, 13, Article 1011409. <https://doi.org/10.3389/fpsyg.2022.1011409>
- Artivive GmbH. (2025). *Augmented reality in education*. <https://www.artivive.com/how-to-use-augmented-reality-in-education>
- Barroso Rodríguez, K. A. (2022). La realidad aumentada en el proceso de enseñanza-aprendizaje. *Technology Rain Journal*, 1(2), 16–27. <https://doi.org/10.55204/trj.v1i2.27>

- Cabero Almenara, J. (2016). La realidad aumentada en educación: Una revisión de la literatura. *Revista de Educación a Distancia (RED)*, 50, 1–27.
- Chang, W. (2021). The cognitive study of immersive experience in science and art exhibition. En T. Ahram & C. Falcão (Eds.), *Advances in human factors in training, education, and learning sciences* (Vol. 12776, pp. 231–240). Springer. https://doi.org/10.1007/978-3-030-78114-9_25
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2.^a ed.). Lawrence Erlbaum Associates.
- Creswell, J. W., y Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5.^a ed.). SAGE Publications.
- Crogman, H., Cano, V., Pacheco, E., Sonawane, R., y Boroon, R. (2025). Virtual reality, augmented reality, and mixed reality in experiential learning: Transforming educational paradigms. *Education Sciences*, 15(3), Article 303. <https://doi.org/10.3390/educsci15030303>
- De La Cruz, E., Gago-Valiente, F., Gavín-Chocano, Ó., y Pérez-Navío, E. (2025). Education, neuroscience, and technology: A review of applied models. *Information*, 16(8), 664. <https://doi.org/10.3390/info16080664>
- Deci, E. L., y Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press.
- Dorregaray Limachi, J. (2020). *Neuroeducación y estrategias de aprendizaje en estudiantes de la Escuela Profesional de Tecnología Médica de la Universidad Peruana Los Andes – 2019* [Tesis de maestría, Universidad Peruana de Ciencias e Informática]. https://repositorio.upci.edu.pe/bitstream/handle/upci/144/T-DORREGARAY_LIMACHI_JOEL.pdf?isAllowed=y&sequence=1
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5.^a ed.). SAGE Publications.

- Guamán Chisag, J. L. (2023). Aprendizaje colaborativo y su influencia efectiva en el rendimiento académico. *MQRInvestigar*, 7(1), 2291–2309. <https://doi.org/10.56048/MQR20225.7.1.2023.2291-2309>
- Hernández-Sampieri, R., y Mendoza, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education.
- Howard-Jones, P. A. (2011). *Investigación neuroeducativa*. La Muralla.
- Instituto Nacional de Estadística e Informática. (2025). Estado de la niñez y adolescencia: Trimestre enero-febrero-marzo 2025 [Informe técnico].
- Jiang, H., Zhu, D., Chugh, R., Turnbull, D., y Jin, W. (2025). Virtual reality and augmented reality-supported K-12 STEM learning: Trends, advantages and challenges. *Education and Information Technologies*, 30, 12827–12863. <https://doi.org/10.1007/s10639-024-13210-z>
- Koç, A., y Kanadlı, S. (2025). Effect of interactive learning environments on learning outcomes in science education: A network meta-analysis. *Journal of Science Education and Technology*, 34, 681–703. <https://doi.org/10.1007/s10956-025-10202-7>
- Koumpouros, Y. (2024). Revealing the true potential and prospects of augmented reality in education. *Smart Learning Environments*, 11, Article 5. <https://doi.org/10.1186/s40561-023-00288-0>
- Lampropoulos, G., Keramopoulos, E., Diamantaras, K., y Evangelidis, G. (2022). Augmented reality and gamification in education: A systematic literature review of research, applications, and empirical studies. *Applied Sciences*, 12(13), Article 6809. <https://doi.org/10.3390/app12136809>
- Lin, Y., y Yu, Z. (2023). A meta-analysis of the effects of augmented reality technologies in interactive learning environments (2012–2022). *Computer Applications in Engineering Education*, 31(5), 1111–1131. <https://doi.org/10.1002/cae.22628>

- Luque Ordóñez, J. (2020). *Realidad virtual y realidad aumentada*. Revista Digital de ACTA, (63). https://www.acta.es/medios/articulos/ciencias_y_tecnologia/063001.pdf
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139164603>
- Mora, F. (2021). *Neuroeducación: Solo se puede aprender aquello que se ama* (3.^a ed.). Alianza Editorial. ISBN 978-84-1362-522-5
- Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. <https://doi.org/10.1007/s10474-010-9438-y>
- Rodríguez, M., & Barragán, H. (2017). Entornos virtuales de aprendizaje como apoyo a la enseñanza presencial para potenciar el proceso educativo. *Killkana Social*, 1(2), 7–14. <https://doi.org/10.26871/killkanasocial.v1i2.29>
- Sakr, A., y Abdullah, T. (2024). Virtual, augmented reality and learning analytics impact on learners, and educators: A systematic review. *Education and Information Technologies*, 29, 19913–19962. <https://doi.org/10.1007/s10639-024-12602-5>
- Teran-Pazmiño, E. M., Cadena-Morales, L. S., González-González, L. P., Guamán-Sánchez, N. de J., & León-Flores, M. C. (2024). Tecnología y personalización del aprendizaje. *Revista Científica Retos de la Ciencia*, 1(4), 115–129. <https://doi.org/10.53877/rc.8.19e.202409.10>
- Triviño Pérez, J. P., Barberán Moreira, J. J., Ordóñez Sánchez, D. J., Lojano Villacis, J. P., & Quintero Ibarra, J. A. (2023). Realidad aumentada como herramienta educativa para aprender electroneumática. *Revista Científica Multidisciplinar G-ner@ndo*, 4(2). <https://revista.gnerando.org/revista/index.php/RCMG/article/view/136>
- UCV. (2022). *Código de ética en investigación*. Universidad César Vallejo.

UCV. (2025). *Reglamento de trabajos conducentes a grados y títulos*. Universidad César Vallejo.

Capítulo 7

Aula virtual y la neuroeducación en los estudiantes de nivel inicial

Gregoria Dueñas Huamani⁸

RESUMEN

La expansión acelerada de las aulas virtuales en la educación inicial ha precedido a la comprensión científica de cómo los entornos virtuales interactúan con los procesos neuroeducativos en preescolares, especialmente en contextos periféricos con restricciones estructurales de conectividad. Este estudio examinó la relación entre el uso del aula virtual y los procesos neuroeducativos en 30 estudiantes de cinco años de la Institución Educativa Inicial N° 269 "Las Palmeras de Tambopata," Puerto Maldonado, Madre de Dios, Perú. Se adoptó un diseño cuantitativo no experimental correlacional transversal. Los datos fueron recolectados mediante una guía de observación estructurada de elaboración propia, validada por juicio de expertos (V de Aiken = 0,98) con alta consistencia interna (α = 0,840 para aula virtual; α = 0,862 para neuroeducación), y analizados mediante el coeficiente Rho de Spearman tras confirmación de distribución no normal (Shapiro-Wilk, p = ,000). Los resultados revelaron una correlación positiva muy alta y estadísticamente significativa entre el uso del aula virtual y los procesos neuroeducativos (ρ = 0,862; p = ,000). El análisis dimensional identificó una gradación de correlaciones entre las cinco dimensiones del aula virtual: accesibilidad (ρ = 0,882), interactividad (ρ = 0,862), usabilidad (ρ = 0,844), evaluación (ρ = 0,834) y contenidos (ρ = 0,702), todas estadísticamente significativas. Se concluye que el aula virtual se asocia

⁸ Especialista en neuroeducación por la Universidad César Vallejo, Perú. E-mail: duenhuagr@ucvvirtual.edu.pe, ORCID: <https://orcid.org/0009-0008-7150-2936>

robustamente con los procesos neuroeducativos incluso bajo condiciones de conectividad restringida, y que las dimensiones de acceso y participación activa son más determinantes para la activación neuroeducativa que las dimensiones de calidad didáctica — hallazgo con implicaciones directas para la priorización de inversión educativa en contextos periféricos latinoamericanos.

Palabras clave: aula virtual; neuroeducación; educación inicial; primera infancia; conectividad digital; contextos periféricos.

Abstract

The rapid expansion of virtual classrooms in early childhood education has outpaced scientific understanding of how virtual environments interact with neuroeducational processes in preschoolers, particularly in peripheral contexts with structural connectivity constraints. This study examined the relationship between virtual classroom use and neuroeducational processes in 30 five-year-old children attending Institución Educativa Inicial N° 269 "Las Palmeras de Tambopata," Puerto Maldonado, Madre de Dios, Peru. A quantitative, non-experimental, correlational, cross-sectional design was adopted. Data were collected using a researcher-designed structured observation guide validated through expert judgment (Aiken's $V = 0.98$) with high internal consistency ($\alpha = .840$ for virtual classroom; $\alpha = .862$ for neuroeducation), and analyzed using Spearman's Rho following confirmation of non-normal distribution (Shapiro-Wilk, $p = .000$). Results revealed a very high and statistically significant positive correlation between virtual classroom use and neuroeducational processes ($\rho = .862$; $p = .000$). Dimensional analysis identified a gradient of correlations across the five virtual classroom dimensions: accessibility ($\rho = .882$), interactivity ($\rho = .862$), usability ($\rho = .844$), evaluation ($\rho = .834$), and content quality ($\rho = .702$), all statistically significant. It is concluded that virtual classrooms are robustly associated with neuroeducational processes even under conditions of restricted connectivity, and that dimensions of access and active participation are more

determinant for neuroeducational activation than dimensions of instructional design quality — a finding with direct implications for the prioritization of educational investment in peripheral Latin American contexts.

Keywords: virtual classroom; neuroeducation; early childhood education; young children; digital connectivity; peripheral contexts.

INTRODUCCIÓN

La digitalización acelerada de la educación inicial, intensificada por la pandemia de COVID-19 y sostenida en su período posterior, ha reposicionado el aula virtual como medio educativo primario para niños en edad preescolar a escala global. El cierre de escuelas que afectó a más de 1.600 millones de estudiantes en todo el mundo obligó a los sistemas educativos a migrar hacia plataformas digitales con escasa preparación y una comprensión aún más limitada de las consecuencias cognitivas para los estudiantes más jóvenes (UNICEF, 2021). Si bien esta transición generó un volumen sin precedentes de investigación sobre tecnología en educación, la expansión de los entornos virtuales ha superado sistemáticamente la comprensión científica de cómo estos entornos interactúan con los mecanismos neurológicos que gobiernan el aprendizaje en niños de tres a seis años — la ventana de desarrollo de mayor plasticidad cerebral y, simultáneamente, de mayor vulnerabilidad ante diseños instruccionales desalineados con el modo en que el cerebro infantil aprende (Williams et al., 2024). Esta asimetría entre expansión tecnológica y comprensión neuroeducativa constituye el problema empírico y teórico que el presente artículo aborda.

La investigación producida entre 2020 y 2026 ha documentado progresivamente que las aulas virtuales pueden sostener el desarrollo cognitivo, lingüístico y social en preescolares cuando están estructuradas, son lúdicas y cuentan con mediación adulta activa (Pottosina & Lambrechts, 2023; Nurhayati et al., 2023). Las tecnologías avanzadas — realidad virtual, realidad aumentada y aprendizaje basado en juegos digitales — muestran efectos especialmente

consistentes sobre la atención, la retención de memoria, la adquisición de conceptos y el pensamiento creativo en esta franja etaria (Williams et al., 2024; Behnamnia et al., 2022; Rapti et al., 2024). La efectividad de estos entornos no es, sin embargo, incondicional: los resultados dependen críticamente del ajuste entre el diseño instruccional y las características cognitivas de los niños, de la calidad del andamiaje adulto y de las condiciones de acceso a la infraestructura tecnológica necesaria para la participación (He & Li, 2023; Cheung & Ostrosky, 2023; Hu et al., 2021). Cuando estas condiciones están ausentes o son inconsistentes, los mismos entornos virtuales que apoyan el aprendizaje en condiciones óptimas pueden generar sobrecarga cognitiva, desvinculación o estancamiento del desarrollo (Putri et al., 2025; Stites et al., 2021). Lo que la evidencia disponible no resuelve todavía, sin embargo, es qué dimensiones específicas del entorno virtual — su interactividad, accesibilidad, usabilidad, calidad de contenidos y sistemas de evaluación — tienen el mayor peso relativo en la activación de los procesos neuroeducativos en preescolares. Los estudios existentes examinan la efectividad de intervenciones tecnológicas específicas sin desagregar los componentes estructurales del entorno virtual como predictores diferenciales de los resultados cognitivos y neuroeducativos. Esta granularidad analítica está ausente de la literatura.

La neuroeducación proporciona el marco conceptual mejor posicionado para abordar esta brecha. Como campo interdisciplinario que integra neurociencia, psicología del desarrollo y pedagogía, la neuroeducación identifica los mecanismos cerebrales que subyacen al aprendizaje efectivo y utiliza este conocimiento para optimizar el diseño instruccional (Williams et al., 2024; Pineda et al., 2025). En la primera infancia específicamente, los procesos neuroeducativos más activos y maleables durante la ventana de desarrollo de tres a seis años son la atención selectiva, la consolidación de la memoria, la motivación intrínseca, la regulación emocional y la plasticidad cerebral — cada uno de los cuales puede activarse o suprimirse en función de las características del entorno de aprendizaje al que el niño está

expuesto (Immordino-Yang & Damasio, 2007; Cabrera et al., 2024). Los entornos virtuales son neuroeducativamente relevantes precisamente porque sus características de diseño — qué tan interactivos son, qué tan accesibles, cuán demandante cognitivamente es su interfaz — modulan directamente estos procesos. Las revisiones neuroeducativas del aprendizaje digital en la primera infancia confirman que los diseños virtuales emocionalmente estimulantes, activos y apropiados para el desarrollo sostienen el desarrollo afectivo y cognitivo (Oviedo et al., 2020; Pineda et al., 2025). Sin embargo, la aplicación sistemática de marcos neuroeducativos al análisis de aulas virtuales estándar — no de tecnologías inmersivas avanzadas, sino de las plataformas que constituyen la realidad educativa cotidiana de la mayoría de los preescolares en el mundo en desarrollo — permanece insuficientemente desarrollada. De manera crítica, ningún estudio identificado en la literatura disponible ha examinado qué dimensiones específicas del entorno de aula virtual estándar ejercen la mayor influencia diferencial sobre los procesos neuroeducativos de los preescolares, privando a educadores y tomadores de decisiones de la evidencia dimensional necesaria para priorizar el diseño instruccional y la asignación de recursos.

Dos brechas convergentes motivan el presente estudio. La primera es conceptual y metodológica: la literatura sobre aulas virtuales y procesos neuroeducativos en la primera infancia carece de granularidad dimensional. La investigación disponible documenta si los entornos virtuales funcionan, y bajo qué condiciones generales, pero no identifica la contribución relativa de dimensiones estructurales específicas — accesibilidad, interactividad, usabilidad, calidad de contenidos, evaluación — a la activación de los procesos neuroeducativos. Este nivel de especificidad es necesario si la investigación ha de generar orientación accionable para el diseño instruccional y la política educativa; sin embargo, está sistemáticamente ausente de los 34 estudios empíricos identificados en la literatura más reciente sobre aulas virtuales y neuroeducación en preescolares. La segunda brecha es

contextual: la base de evidencia está geográficamente concentrada en contextos con infraestructura digital estable — Europa, Asia Oriental, Norteamérica y América Latina urbana — dejando a las regiones periféricas estructuralmente subrepresentadas. En el Perú, donde las estadísticas oficiales indican que solo el 51,8% de la población rural tenía acceso a internet al cierre de 2023 (INEI, 2023), y donde regiones amazónicas como Madre de Dios han sido formalmente identificadas por el gobierno nacional como zonas prioritarias de intervención en conectividad digital (MTC, 2023), las condiciones bajo las cuales operan las aulas virtuales son fundamentalmente distintas de las documentadas en la literatura existente. Esta brecha contextual no es meramente geográfica — es analíticamente consecuente, porque las restricciones de conectividad pueden moderar la relación entre la calidad del entorno virtual y la activación neuroeducativa de maneras que los estudios realizados bajo condiciones de infraestructura estable no capturan. Ambas brechas en conjunto definen un espacio de investigación que el presente estudio está diseñado para ocupar.

Este estudio examina la relación entre el uso del aula virtual y los procesos neuroeducativos en niños de cinco años que asisten a una institución educativa inicial pública en Madre de Dios, Perú — una región amazónica con limitaciones estructurales de conectividad documentadas, representativa de las condiciones que comparten millones de preescolares en América Latina periférica. Los objetivos específicos son determinar la correlación entre cada una de las cinco dimensiones del aula virtual — interactividad, accesibilidad, usabilidad, calidad de contenidos y evaluación — y los procesos neuroeducativos de los estudiantes participantes. Al aportar evidencia dimensional sobre qué aspectos del entorno virtual están más fuertemente asociados con la activación neuroeducativa bajo condiciones de conectividad restringida, este estudio contribuye tanto empíricamente, generando evidencia desde un contexto sistemáticamente subrepresentado, como analíticamente, al introducir un marco dimensional para el estudio de la calidad del aula virtual en la investigación neuroeducativa de la primera infancia.

METODOLOGÍA

2.1 Diseño de investigación

El estudio adoptó un diseño cuantitativo, no experimental, correlacional y transversal. El enfoque cuantitativo se justifica porque el objetivo de la investigación requiere establecer la magnitud y dirección de la relación entre el uso del aula virtual y los procesos neuroeducativos mediante coeficientes estadísticos, lo que demanda medición numérica y análisis inferencial (Creswell & Creswell, 2018). El diseño es no experimental porque la variable independiente —el uso del aula virtual— no fue manipulada; hacerlo habría resultado metodológicamente inapropiado y éticamente indefendible dado que los participantes eran niños de cinco años en su entorno natural de aprendizaje (Cohen et al., 2018). El nivel correlacional responde al objetivo de determinar el grado de asociación entre variables sin establecer atribución causal (Hernández-Sampieri & Mendoza, 2018). Finalmente, el carácter transversal refleja que los datos fueron recolectados en un único momento temporal, proporcionando una imagen sincrónica de la relación entre ambas variables durante el período académico 2025.

2.2 Participantes

2.2.1 Descripción de la muestra y contexto

Los participantes fueron 30 niños de cinco años matriculados en el aula "Palo Rosa" de la Institución Educativa Inicial N° 269 "Las Palmeras de Tambopata," ubicada en Puerto Maldonado, Madre de Dios, Perú. El estudio constituyó un censo de la totalidad de la población estudiantil activa en el único grupo con implementación sistemática de aula virtual disponible en la institución durante el período de estudio (abril–junio de 2025), y no una muestra probabilística extraída de una población mayor. Esta decisión elimina el error muestral dentro del grupo observado y es coherente con las recomendaciones metodológicas para investigación educativa en contextos geográficamente periféricos donde el acceso poblacional es inherentemente limitado (Etikan et al., 2016).

El contexto institucional resulta relevante para la interpretación de los hallazgos. Madre de Dios es una región amazónica caracterizada por brechas significativas en infraestructura digital; los propios datos recolectados durante el estudio revelaron que el 26,7% de los participantes experimentó acceso irregular a las plataformas virtuales, cifra consistente con reportes nacionales que documentan déficits persistentes de conectividad en provincias amazónicas peruanas (INEI, 2023). Este contexto sociotecnológico distingue al presente estudio de investigaciones realizadas en entornos urbanos con infraestructura digital estable, y debe considerarse al evaluar las condiciones bajo las cuales operan las aulas virtuales en la educación inicial.

2.2.2 Criterios de inclusión y exclusión

Los criterios de inclusión fueron: matrícula activa en el aula "Palo Rosa" para el año académico 2025, participación regular en sesiones de aula virtual durante el período de recolección de datos, y disponibilidad para la observación directa por parte de la investigadora. Los criterios de exclusión fueron: ausencia sistemática de las sesiones virtuales durante el período de observación e imposibilidad técnica de conexión a la plataforma a lo largo de toda la ventana de recolección. Ningún participante cumplió los criterios de exclusión en su totalidad; los tres estudiantes con conectividad irregular fueron observados durante las sesiones en que lograron conexión, sin imputación de valores faltantes ni exclusión del conjunto de datos.

En cuanto al tamaño muestral, si bien $n=30$ se sitúa en el límite inferior de lo convencionalmente aceptado para análisis correlacional, su adecuación debe evaluarse en relación con el procedimiento estadístico específico empleado. Para la Rho de Spearman con tamaños de efecto esperados en el rango medio-alto —que la literatura sobre entornos virtuales de aprendizaje y procesos cognitivos en la primera infancia reporta de manera sistemática (Falloon, 2020; Zosh et al., 2022)— $n=30$ proporciona sensibilidad suficiente para la detección de asociaciones significativas (Field, 2018). Los autores reconocen, no obstante, que este

tamaño muestral limita la potencia estadística para detectar correlaciones débiles y restringe la generalización de los hallazgos a contextos que compartan características institucionales y sociotecnológicas análogas.

2.3 Variables e instrumentos

2.3.1 Operacionalización de variables

La variable *aula virtual* fue operacionalizada en cinco dimensiones con sustento teórico: interactividad, accesibilidad, usabilidad, calidad de contenidos y evaluación. Esta estructura dimensional refleja los marcos establecidos para la evaluación de la calidad de entornos virtuales de aprendizaje, que convergen en estos componentes como determinantes primarios de experiencias de aprendizaje digital efectivas en la educación temprana (Bower, 2019; Means et al., 2014). La variable *neuroeducación* fue operacionalizada en cinco dimensiones derivadas de marcos neuroeducativos aplicados al desarrollo en la primera infancia: atención, memoria, motivación, emociones y plasticidad cerebral. Estas dimensiones representan los procesos cognitivos y afectivos centrales que la investigación neuroeducativa identifica como susceptibles de modulación mediante entornos de aprendizaje estructurados durante los primeros años de escolarización (Ansari et al., 2012; Immordino-Yang & Damasio, 2007). Cada dimensión fue medida a través de tres ítems, resultando en 15 ítems por variable. La correspondencia entre dimensiones, ítems y objetivos específicos del estudio se presenta en la matriz de operacionalización (véase Material Suplementario).

2.3.2 Tipo de instrumento y modalidad de aplicación

Se diseñó una guía de observación estructurada específicamente para este estudio. La elección de la metodología observacional sobre el autoinforme no fue incidental sino teóricamente determinada: los niños de cinco años aún no han desarrollado la capacidad metacognitiva necesaria para la autoevaluación confiable de sus propios estados cognitivos y emocionales (Flavell, 1979; Whitebread et al., 2009), lo que convierte a los cuestionarios autoadministrados

en instrumentos inválidos para esta población desde perspectivas tanto cognitivas como psicométricas. La investigadora, quien desempeñaba simultáneamente el rol de docente del aula, completó el protocolo de observación durante las sesiones activas de aula virtual realizadas entre abril y junio de 2025, registrando las manifestaciones conductuales e interaccionales de cada estudiante en cada ítem en tiempo real o dentro de los treinta minutos posteriores a la conclusión de cada sesión, con el fin de minimizar el sesgo de memoria retrospectiva. Las respuestas fueron registradas en una escala ordinal de cinco puntos que oscila entre 1 (Nunca) y 5 (Siempre).

El doble rol de la investigadora como docente y observadora constituye una fuente potencial de sesgo de expectativa, limitación que se reconoce explícitamente y se aborda en términos procedimentales en la sección 2.4.

2.3.3 Validez de contenido

El instrumento fue sometido a validación de contenido mediante juicio de expertos siguiendo un protocolo estructurado. Participaron en el proceso dos jueces, ambos con grado de doctor en psicología y ejerciendo como profesores universitarios activos, perfiles que se alinean con el doble fundamento teórico del instrumento en medición psicoeducativa y procesos neuroeducativos. Cada juez evaluó de manera independiente la totalidad de los ítems conforme a tres criterios estándar: claridad de la redacción, pertinencia respecto a la dimensión correspondiente y relevancia para el constructo más amplio. El coeficiente de validez de contenido fue calculado mediante la fórmula de la V de Aiken para cada ítem y para el instrumento en su conjunto. El resultado global fue $V=0,98$, superando sustancialmente el umbral de $V \geq 0,80$ establecido en la literatura psicométrica como indicador de alta validez de contenido (Aiken, 1985; Merino-Soto & Livia-Segovia, 2009). Los resultados a nivel de ítem confirmaron la ausencia de reactivos que requirieran eliminación o revisión sustantiva,

respaldando la integridad estructural de ambas escalas como operacionalizaciones de sus respectivos constructos.

2.3.4 Consistencia interna

La consistencia interna fue estimada mediante el coeficiente alfa de Cronbach calculado sobre el conjunto completo de datos ($n=30$). La escala de aula virtual arrojó $\alpha=0,840$ y la escala de neuroeducación arrojó $\alpha=0,862$. Ambos valores se ubican en el rango "bueno" según criterios psicométricos establecidos (George & Mallery, 2003), indicando que los ítems de cada escala miden un constructo subyacente suficientemente coherente para sustentar inferencias a nivel de variable y de dimensión. Estos niveles de fiabilidad son comparables a los reportados en instrumentos que evalúan calidad del aprendizaje digital y procesos neuroeducativos en contextos análogos de educación inicial en América Latina (Cedeño, 2021; Paredes, 2020), lo que sugiere que el desempeño psicométrico del presente instrumento es consistente con los estándares de medición de este dominio de investigación.

2.4 Procedimiento de recolección de datos

2.4.1 Fases del procedimiento

La recolección de datos se organizó en cuatro fases secuenciales. En la primera fase se obtuvo la autorización institucional formal de la dirección de la IEI N° 269 "Las Palmeras de Tambopata," seguida de la administración de formularios de consentimiento informado a los padres y tutores legales de los 30 participantes. La documentación de consentimiento especificó explícitamente los objetivos del estudio, los procedimientos de observación, el uso exclusivamente académico de los datos, el carácter voluntario de la participación y el derecho de retiro en cualquier momento sin consecuencia alguna.

En la segunda fase, el protocolo de observación fue estandarizado con anterioridad al inicio de la recolección de datos. Cada ítem fue operacionalizado mediante descriptores conductuales específicos y observables, con el fin de reducir la ambigüedad interpretativa durante la

observación en vivo. Este paso de operacionalización es particularmente relevante en la investigación mediante observación estructurada para garantizar que las decisiones de puntuación estén ancladas en referentes conductuales concretos y no en impresiones holísticas del observador (Pellegrini, 2004).

En la tercera fase se llevó a cabo la observación sistemática durante las sesiones de aula virtual a lo largo del período abril–junio de 2025. Cada participante fue observado a lo largo de múltiples sesiones, y las puntuaciones de los ítems reflejaron el juicio consolidado de la investigadora sobre el patrón conductual típico del estudiante dentro de la ventana de observación, y no un registro puntual de una única sesión. Las puntuaciones fueron registradas en la guía de observación estructurada de manera inmediata durante la sesión o dentro de los treinta minutos posteriores a su conclusión.

En la cuarta fase, todos los datos observacionales fueron codificados y transferidos a una matriz digital para su procesamiento estadístico mediante IBM SPSS Statistics versión 25.

2.4.2 Control de sesgos y limitaciones procedimentales

El doble rol de investigadora-docente introduce un riesgo de sesgo de expectativa que requiere discusión explícita. Las investigadoras que poseen conocimiento relacional previo de sus participantes pueden puntuar inconscientemente los ítems en direcciones consistentes con sus expectativas sobre estudiantes individuales (Rosenthal, 2002). Se implementaron tres decisiones procedimentales para atenuar este riesgo: el uso de descriptores conductuales anclados a cada punto de la escala —reduciendo la interpretación subjetiva—, la aplicación sistemática del protocolo sobre todos los participantes dentro de cada sesión —previniendo la atención selectiva— y la disciplina temporal de registrar las puntuaciones dentro de los treinta minutos posteriores a la observación —reduciendo la distorsión por memoria reconstructiva—. Estas medidas controlan parcialmente, pero no eliminan, el sesgo del observador, el cual permanece como limitación reconocida del estudio.

Adicionalmente, la irregularidad de conectividad documentada para el 26,7% de los participantes generó oportunidades de observación desiguales entre estudiantes. La decisión de observar a estos estudiantes únicamente durante las sesiones en que lograron conexión —en lugar de imputar puntuaciones o excluirllos— preserva la integridad de los datos al tiempo que introduce una posible subrepresentación de la experiencia de los estudiantes en situación de desventaja digital dentro de la muestra.

2.5 Análisis estadístico

Se calcularon estadísticos descriptivos —frecuencias, porcentajes, medias y desviaciones estándar— para cada variable y dimensión, con el fin de caracterizar la distribución de los patrones de uso del aula virtual y los niveles de los procesos neuroeducativos en la muestra con anterioridad al análisis inferencial. Estos resultados se presentan en formato tabular en la sección de Resultados.

Previamente al análisis correlacional, se examinó la normalidad de la distribución de ambas variables mediante la prueba de Shapiro-Wilk, seleccionada sobre la prueba de Kolmogorov-Smirnov por su mayor potencia estadística con muestras inferiores a 50 casos (Razali & Wah, 2011). Los resultados indicaron que ninguna de las dos variables siguió una distribución normal (ambas $p=,000$), descartando el uso del coeficiente r de Pearson, cuyo empleo requiere la satisfacción de este supuesto.

Dada la no normalidad confirmada de ambas distribuciones, el escalamiento ordinal del instrumento de medición y el tamaño muestral de $n=30$, se seleccionó el coeficiente de correlación por rangos de Spearman (ρ) como estadístico inferencial para todos los análisis bivariados (Siegel & Castellan, 1988). Este coeficiente fue calculado para la relación entre las dos variables principales y para cada combinación por pares de dimensiones de ambas variables, generando las correlaciones específicas correspondientes a cada hipótesis de

investigación. Se estableció $\alpha=,05$ como umbral de significancia estadística para todas las pruebas de hipótesis. Todos los análisis fueron realizados con IBM SPSS Statistics versión 25.

2.6 Consideraciones éticas

El estudio fue conducido en estricta adhesión a los principios éticos que rigen la investigación educativa con menores de edad. La autorización institucional fue otorgada formalmente por la dirección de la IEI N° 269 "Las Palmeras de Tambopata" con anterioridad al inicio de cualquier actividad de recolección de datos. El consentimiento informado por escrito fue obtenido de los padres o tutores legales de los 30 participantes; la documentación de consentimiento proporcionó información clara sobre la naturaleza y propósito del estudio, los procedimientos de observación a emplear, el uso exclusivamente académico de los datos resultantes y el derecho incondicional de retirar la participación en cualquier etapa sin consecuencia negativa de ningún tipo. Todos los datos fueron completamente anonimizados a lo largo de la recolección, el almacenamiento y el análisis, garantizando que ningún participante individual pueda ser identificado a partir de los resultados reportados. La metodología observacional empleada no introdujo intervención alguna más allá de las actividades habituales de aula virtual de los estudiantes, sin representar riesgo adicional de orden físico, psicológico o académico para los participantes. El estudio fue conducido de conformidad con las directrices éticas de la Universidad César Vallejo y con los estándares éticos establecidos por la American Educational Research Association (AERA, 2011) para investigación educativa con participantes humanos.

RESULTADOS

3.1 Caracterización descriptiva de las variables

El análisis descriptivo reveló perfiles de distribución diferenciados entre las dos variables del estudio. En relación con el uso del aula virtual, el 90,0% de los estudiantes fue observado en la categoría "casi siempre" y el 10,0% restante en "a veces," lo que indica una integración elevada

y sostenida de esta herramienta en el proceso de enseñanza durante el período de observación. El análisis por dimensiones, presentado en la Tabla 1, revela variabilidad interna relevante. La dimensión evaluación mostró el perfil más alto, con el 96,7% de los estudiantes en las categorías "casi siempre" o "siempre," seguida de interactividad (90,0% en "casi siempre"). Las dimensiones contenidos y accesibilidad mostraron perfiles intermedios, con el 76,7% y el 73,3% en "casi siempre," respectivamente. La dimensión usabilidad presentó el perfil más heterogéneo: únicamente el 6,7% de los estudiantes fue observado en "casi siempre," el 76,7% en "a veces" y el 16,7% en "casi nunca," constituyendo la dimensión con mayor proporción de respuestas en categorías bajas y anticipando su posición en el análisis correlacional posterior.

Tabla 1. Perfiles descriptivos de la variable Aula Virtual y sus dimensiones (n=30)

Dimensión	Casi nunca %	A veces %	Casi siempre %	Siempre %
Aula Virtual (general)	—	10,0	90,0	—
Evaluación	—	—	96,7	3,3
Interactividad	—	10,0	90,0	—
Contenidos	—	23,3	76,7	—
Accesibilidad	—	26,7	73,3	—
Usabilidad	16,7	76,7	6,7	—

Nota. Las dimensiones se presentan en orden descendente de frecuencia en la categoría modal. Los porcentajes corresponden a frecuencias observadas mediante guía de observación estructurada aplicada durante sesiones de aula virtual (abril–junio 2025).

En cuanto a la variable neuroeducación, el 96,7% de los estudiantes fue observado en la categoría "a veces" y únicamente el 3,3% en "casi siempre," con ausencia total de observaciones en las categorías "siempre," "casi nunca" y "nunca." Esta distribución fuertemente concentrada en una categoría intermedia indica que los procesos neuroeducativos

—atención, memoria, motivación, regulación emocional y plasticidad cerebral— se activaron de manera parcial y no sistemática durante las sesiones observadas, lo cual es coherente con la irregularidad de acceso digital documentada en la muestra y con la heterogeneidad en la usabilidad percibida de las plataformas. Es conceptualmente importante distinguir aquí dos planos analíticos: la concentración en "a veces" describe el nivel absoluto de activación neuroeducativa observada, no la estructura de su relación con el uso del aula virtual. Estos son planos estadísticamente independientes — una variable puede mostrar nivel moderado y covariación alta con otra simultáneamente — y es precisamente esa estructura relacional la que se examina en las subsecciones siguientes mediante análisis inferencial.

3.2 Verificación de la hipótesis general

El análisis de correlación de Spearman entre el uso del aula virtual y los procesos neuroeducativos arrojó $\rho=0,862$ ($p=,000$; $n=30$), resultado que permite rechazar la hipótesis nula y aceptar la hipótesis alternativa general: existe una relación positiva, muy alta y estadísticamente significativa entre ambas variables en los estudiantes de nivel inicial observados (Cohen, 1988).

La interpretación sustantiva de este coeficiente requiere considerar el contexto distribucional en que emerge. El hecho de que el 90,0% de los estudiantes use el aula virtual en la categoría "casi siempre" mientras que solo el 3,3% muestre procesos neuroeducativos en esa misma categoría implica que $\rho=0,862$ no deriva de una elevación paralela de ambas variables, sino de una covariación sistemática dentro de sus respectivos rangos de distribución. Dicho de otro modo, la asociación es analíticamente no obvia — no es el resultado de que ambas variables estén igualmente elevadas — lo que refuerza su relevancia empírica. Esta correlación de magnitud muy alta emerge, además, en un contexto caracterizado por restricciones tecnológicas documentadas, donde el 26,7% de los participantes experimentó acceso irregular a las plataformas virtuales, condición que, lejos de atenuar la asociación, sugiere que la relación

entre entorno virtual y procesos neuroeducativos opera de manera robusta incluso bajo condiciones subóptimas de conectividad.

Debe precisarse, no obstante, que el diseño no experimental del estudio no permite establecer dirección causal entre las variables: la correlación observada es consistente tanto con la hipótesis de que el uso del aula virtual activa procesos neuroeducativos como con la de que estudiantes con mayor activación cognitiva aprovechan más sistemáticamente los entornos virtuales disponibles. Esta cautela interpretativa es particularmente relevante dado que la distribución de la variable neuroeducación, concentrada en una categoría intermedia, introduce una restricción de rango que puede atenuar el coeficiente observado respecto de la asociación poblacional real. En consecuencia, $\rho=0,862$ debe interpretarse como estimación conservadora de la magnitud de la relación. La magnitud del hallazgo es, en todo caso, consistente con la literatura que documenta asociaciones positivas entre la calidad de entornos virtuales de aprendizaje y la activación de procesos cognitivos en la educación temprana (Falloon, 2020; Zosh et al., 2022), lo que otorga validez convergente al resultado. La estructura interna de esta relación — su distribución desigual entre las cinco dimensiones del aula virtual — se examina en la subsección siguiente, donde la gradación de coeficientes constituye en sí misma un hallazgo analítico.

3.3 Análisis integrado de hipótesis específicas

El análisis de correlación de Spearman entre cada dimensión del aula virtual y la variable neuroeducación arrojó coeficientes estadísticamente significativos en los cinco casos ($p<,05$), lo que permite rechazar las cinco hipótesis nulas específicas y aceptar H1 a H5 en su totalidad. Los resultados se presentan en la Tabla 2 en orden descendente de magnitud del coeficiente.

Tabla 2. Correlaciones de Spearman entre dimensiones del aula virtual y neuroeducación (n=30)

Dimensión	ρ	p	Decisión
Accesibilidad	0,882	,001	Se rechaza H_0

Interactividad	0,862	,001	Se rechaza H_0
Usabilidad	0,844	,000	Se rechaza H_0
Evaluación	0,834	,002	Se rechaza H_0
Contenidos	0,702	,000	Se rechaza H_0

Nota. ρ = coeficiente de correlación por rangos de Spearman. Todas las correlaciones son positivas. Nivel de significancia $\alpha=,05$. Las dimensiones se presentan en orden descendente de magnitud del coeficiente.

La inspección de la Tabla 2 revela que los cinco coeficientes son sustancialmente altos en términos absolutos — el mínimo ($\rho=0,702$) supera el umbral convencional de efecto grande (Cohen, 1988) — pero exhiben una gradación que no es aleatoria sino interpretable como jerarquía funcional de las dimensiones del aula virtual en relación con los procesos neuroeducativos. Esta gradación, que oscila entre $\rho=0,882$ y $\rho=0,702$, constituye el hallazgo analítico central de esta subsección y se examina a continuación en tres niveles.

En el polo superior de la gradación se sitúan accesibilidad ($\rho=0,882$) e interactividad ($\rho=0,862$), las dos dimensiones con correlaciones más altas. Este patrón es teóricamente coherente y puede interpretarse en términos de jerarquía funcional: accesibilidad opera como condición de posibilidad de cualquier proceso neuroeducativo mediado digitalmente — sin acceso estable a la plataforma, ninguna otra dimensión puede activarse — lo que explica que su correlación con neuroeducación sea la más alta de las cinco. El dato descriptivo de que el 26,7% de los participantes experimentó acceso irregular, reportado en la sección anterior, refuerza esta interpretación en el contexto específico de Madre de Dios: la variabilidad en el acceso se traduce directamente en variabilidad en la activación neuroeducativa, generando una covariación especialmente pronunciada. La interactividad, por su parte, es la dimensión que activa de manera más directa los mecanismos neuroeducativos centrales — atención sostenida, memoria de trabajo y motivación intrínseca — al generar demanda cognitiva activa sobre el

estudiante en lugar de mantenerlo en posición de receptor pasivo. Desde la teoría de la carga cognitiva, la participación interactiva reduce la carga extrínseca al orientar la atención hacia la tarea, liberando recursos cognitivos para el procesamiento intrínseco del contenido (Sweller et al., 2019); desde marcos neuroeducativos, la participación activa es condición necesaria para los procesos de consolidación neural que subyacen al aprendizaje significativo en la primera infancia (Immordino-Yang & Damasio, 2007).

En el polo medio de la gradación se ubican usabilidad ($\rho=0,844$) y evaluación ($\rho=0,834$), con coeficientes muy próximos entre sí. El hallazgo de evaluación es directamente interpretable: la retroalimentación sistemática — documentada en el 96,7% de los estudiantes como práctica presente en las categorías más altas — activa mecanismos de memoria de trabajo, autorregulación y motivación que la investigación neuroeducativa y pedagógica identifica consistentemente como potenciadores del aprendizaje (Hattie & Timperley, 2007; Shute, 2008). Una evaluación frecuente y pertinente no solo mide el aprendizaje sino que lo genera, al forzar la recuperación activa de información y activar procesos metacognitivos que consolidan las representaciones cognitivas. El hallazgo de usabilidad merece atención analítica particular porque introduce una tensión interpretativa productiva: esta dimensión presenta el perfil descriptivo más heterogéneo y problemático de las cinco — con el 76,7% de los estudiantes en "a veces" y el 16,7% en "casi nunca" — y sin embargo su correlación con neuroeducación ($\rho=0,844$) es la tercera más alta. Esta aparente paradoja sugiere que incluso una usabilidad percibida como parcialmente adecuada genera procesos neuroeducativos mensurables, lo cual podría indicar que los niños de cinco años desarrollan estrategias adaptativas ante interfaces de complejidad moderada, ajustando sus recursos cognitivos a las demandas de navegación sin que esto comprometa necesariamente la activación neuroeducativa. Alternativamente, la covariación alta con neuroeducación pese al bajo nivel de usabilidad percibida podría reflejar que, en este grupo específico, la mediación docente

durante las sesiones virtuales compensó parcialmente las limitaciones de usabilidad de las plataformas.

En el polo inferior de la gradación se sitúa la dimensión contenidos con la correlación más baja del conjunto ($\rho=0,702$), aunque estadísticamente significativa y sustancialmente alta en términos absolutos. La posición de esta dimensión en la gradación es interpretable en función de dos factores convergentes. En primer lugar, la calidad y relevancia de los contenidos digitales es la dimensión más dependiente de decisiones curriculares y pedagógicas externas al contexto inmediato de interacción estudiante-plataforma, y más susceptible a la variabilidad de recursos en contextos de infraestructura limitada; en Madre de Dios, donde el 23,3% de los estudiantes percibió los contenidos como disponibles solo "a veces," la heterogeneidad en la calidad y disponibilidad del material pudo haber introducido varianza que atenuó el coeficiente. En segundo lugar, la relevancia neuroeducativa de los contenidos depende de manera crítica de cómo estos se presentan y no solo de que estén disponibles — un contenido de alta calidad curricular pero con diseño visual inadecuado para la franja etaria de cinco años puede no activar los procesos atencionales y motivacionales que la neuroeducación identifica como precondiciones del aprendizaje (Ansari et al., 2012). Pese a esta posición inferior en la gradación, $\rho=0,702$ confirma que la calidad de contenidos mantiene una asociación sustantiva con los procesos neuroeducativos — su correlación es la más baja del conjunto pero continúa siendo un predictor relevante de la activación cognitiva observada.

La gradación completa —de $\rho=0,882$ en accesibilidad a $\rho=0,702$ en contenidos— revela una jerarquía funcional interpretable: las dimensiones que operan como condiciones de acceso y participación activa muestran correlaciones más altas con los procesos neuroeducativos que las dimensiones que operan como variables de diseño y contenido. Este patrón sugiere que, en la educación inicial mediada digitalmente y en contextos con restricciones de infraestructura, garantizar el acceso equitativo y la participación activa es más determinante para los procesos

neuroeducativos que optimizar el diseño de contenidos — hallazgo con implicaciones directas para la priorización de recursos en política educativa en contextos similares. Esta interpretación es consistente con marcos que posicionan la presencia activa y el acceso como precondiciones de la construcción cognitiva en entornos de aprendizaje mediados (Garrison & Anderson, 2003), y trasciende el contexto específico de Madre de Dios para ofrecer implicaciones para contextos amazónicos y periféricos análogos en la región latinoamericana. Los hallazgos en su conjunto son discutidos en relación con la literatura existente y sus implicaciones teóricas y prácticas en la sección siguiente.

DISCUSIÓN

Los hallazgos del presente estudio revelan una correlación positiva muy alta y estadísticamente significativa entre el uso del aula virtual y los procesos neuroeducativos en estudiantes de nivel inicial en Madre de Dios ($\rho=0,862$; $p=,000$), con cinco correlaciones dimensionales significativas que oscilan entre $\rho=0,882$ para accesibilidad y $\rho=0,702$ para contenidos. Esta gradación dimensional — y no solo la magnitud global de la asociación — constituye el hallazgo analítico central que la discusión examina.

La magnitud de la correlación general es interpretable desde el marco de Comunidades de Indagación (Garrison & Anderson, 2003), el cual, si bien fue desarrollado originalmente para entornos virtuales de educación superior, establece principios de presencia aplicables a todo contexto de aprendizaje virtual con mediación docente estructurada — condición que el presente estudio satisface a través de la participación activa de la investigadora-docente durante las sesiones observadas. Desde este marco, el coeficiente obtenido refleja que el aula virtual activó presencias cognitiva, social y didáctica suficientes para generar covariación sistemática con los procesos neuroeducativos, incluso bajo condiciones de infraestructura restringida. Esta robustez es coherente con los postulados de la neuroplasticidad cerebral (Immordino-Yang & Damasio, 2007; Williams et al., 2024), que sostienen que el cerebro infantil puede reorganizar

sus conexiones neuronales en respuesta a entornos de aprendizaje estructurados incluso cuando las condiciones de acceso son subóptimas. La robustez del coeficiente global no implica, sin embargo, uniformidad dimensional — la resiliencia opera en el nivel de la covariación agregada, mientras que la distribución desigual de esa resiliencia entre dimensiones es precisamente lo que la gradación revela y lo que el marco teórico permite explicar.

La gradación dimensional constituye el aporte analítico más original de este estudio. El marco de Comunidades de Indagación establece que las presencias de acceso y participación son precondiciones de la presencia cognitiva y didáctica — sin ellas, ninguna otra variable del entorno puede activar procesos de aprendizaje. Esta distinción jerárquica entre precondiciones y variables de calidad mapea directamente sobre la gradación observada: accesibilidad e interactividad, las dos dimensiones con correlaciones más altas, operan exactamente como precondiciones del sistema — su ausencia o irregularidad compromete sistémicamente la activación neuroeducativa con independencia de la calidad de los demás componentes. Usabilidad, evaluación y contenidos, por su parte, operan como variables de calidad didáctica que optimizan, pero no habilitan la activación neuroeducativa, lo que explica su posición intermedia y baja en la gradación. La posición más baja de contenidos ($p=0,702$) es adicionalmente interpretable desde la Teoría de la Carga Cognitiva (Sweller et al., 2019): en el contexto de esta muestra, donde el 26,7% de los estudiantes experimentó acceso irregular, la variabilidad en la disponibilidad de contenidos pudo haber introducido carga extrínseca que atenuó la asociación entre esta dimensión y los procesos neuroeducativos — la carga cognitiva que impone la irregularidad de acceso al contenido consume recursos que de otro modo estarían disponibles para el procesamiento neuroeducativo intrínseco.

Los hallazgos son consistentes con investigaciones previas que documentan relaciones positivas entre entornos virtuales y desarrollo cognitivo en preescolares (Espino-Díaz et al., 2020; Hu et al., 2021), y con los hallazgos de Falloon (2020), quien identificó la interacción

activa estudiante-contenido como determinante central del aprendizaje digital en la primera infancia. La primacía de la accesibilidad coincide con Stites et al. (2021), quienes señalaron el acceso irregular como el principal obstáculo para el aprendizaje a distancia en preescolares, y con Molina (2022), quien documentó el impacto negativo de las restricciones tecnológicas en el aprendizaje temprano en zonas rurales peruanas. La relación entre interactividad y procesos cognitivos es consistente con Cedeño (2021), quien concluyó que las actividades interactivas optimizan el uso de las funciones cerebrales durante el aprendizaje, y con Paredes (2020), quien reportó que el uso de aulas virtuales tiene efecto real en el aprendizaje de los niños. Sin embargo, ninguno de estos estudios examina la contribución diferencial de dimensiones específicas del entorno virtual — este vacío es precisamente el que el presente estudio llena al demostrar empíricamente que las dimensiones no son equivalentes en su relación con la activación neuroeducativa. La única tensión con la literatura proviene de Putri et al. (2025), quienes advierten sobre riesgo de sobrecarga cognitiva en entornos digitales para preescolares; los coeficientes altos observados en este estudio bajo condiciones de usabilidad parcial sugieren que la mediación docente activa pudo haber funcionado como andamiaje que atenuó ese riesgo, hipótesis que estudios futuros deberían examinar directamente.

Este estudio presenta tres limitaciones que acotan la interpretación de los hallazgos. El diseño correlacional no permite inferir causalidad entre las variables. El tamaño muestral ($n=30$) y el contexto específico de Madre de Dios limitan la generalización a contextos con características institucionales y sociotecnológicas análogas. El doble rol investigadora-docente introduce riesgo de sesgo de expectativa que las medidas procedimentales adoptadas controlan parcialmente pero no eliminan; simultáneamente, esta misma condición proporcionó la mediación docente natural que los datos sugieren como factor explicativo de los coeficientes observados bajo condiciones subóptimas — tensión que estudios futuros con observadores independientes deberían desambiguar. Se recomienda el desarrollo de estudios cuasi-

experimentales con muestras mayores que examinen el efecto causal de intervenciones específicas sobre cada dimensión del aula virtual en contextos periféricos latinoamericanos, así como estudios que midan directamente el rol mediador de la intervención docente en la relación entre accesibilidad y activación neuroeducativa.

CONCLUSIONES

Los hallazgos del presente estudio demuestran que el uso del aula virtual mantiene una relación positiva muy alta y estadísticamente significativa con los procesos neuroeducativos en estudiantes de nivel inicial en Madre de Dios, Perú ($p=0,862$; $p=,000$). La magnitud de esta asociación bajo condiciones documentadas de restricción de conectividad — donde el 26,7% de los participantes experimentó acceso irregular a las plataformas — indica que la relación entre entorno virtual y activación neuroeducativa no es un fenómeno exclusivo de contextos con infraestructura digital estable, sino que opera con robustez en condiciones de infraestructura subóptima representativas de millones de preescolares en regiones periféricas de América Latina. Esta conclusión trasciende el contexto específico del estudio para ofrecer evidencia relevante a todo contexto amazónico y periférico con características sociotecnológicas análogas.

El hallazgo analíticamente más original del estudio reside, sin embargo, no en la magnitud de la correlación global sino en la gradación que articula su estructura interna. Las cinco dimensiones del aula virtual muestran correlaciones positivas significativas con los procesos neuroeducativos, pero no equivalentes: accesibilidad ($p=0,882$) e interactividad ($p=0,862$) — dimensiones que operan como condiciones de acceso y participación activa — muestran las correlaciones más altas, mientras que usabilidad ($p=0,844$), evaluación ($p=0,834$) y contenidos ($p=0,702$) — dimensiones de calidad didáctica — muestran correlaciones progresivamente menores, aunque sustancialmente altas en términos absolutos. Esta gradación establece una conclusión de alcance práctico y teórico directo: en contextos de infraestructura digital

restringida, garantizar el acceso estable y la participación activa en el entorno virtual es más determinante para la activación neuroeducativa que optimizar el diseño de contenidos o la complejidad de la interfaz. La jerarquía funcional de dimensiones no es un artefacto estadístico sino un patrón interpretable que el marco de Comunidades de Indagación predice y que los datos de este estudio confirman empíricamente por primera vez en el contexto de la educación inicial latinoamericana periférica.

Este estudio realiza tres contribuciones al campo. Empíricamente, genera evidencia sobre la relación aula virtual–neuroeducación desde un contexto sistemáticamente subrepresentado en la literatura internacional — una zona amazónica peruana con restricciones estructurales de conectividad — demostrando que esta relación es robusta más allá de los contextos urbanos y de infraestructura estable en los que se concentra la evidencia existente. Analíticamente, introduce un marco dimensional para el estudio de la relación entre entornos virtuales y procesos neuroeducativos en la primera infancia, demostrando que las dimensiones del aula virtual no son intercambiables en su capacidad de activar procesos cognitivos y que su jerarquía funcional es empíricamente detectable y teóricamente coherente. Teóricamente, extiende la aplicabilidad empírica del marco de Comunidades de Indagación al contexto de la educación inicial mediada digitalmente en zonas periféricas, contexto para el cual este marco no había sido previamente validado, ampliando así su alcance explicativo más allá de la educación superior en línea para la que fue originalmente desarrollado.

Los hallazgos tienen implicaciones directas para dos niveles de decisión. En el nivel de política educativa, la primacía de la accesibilidad en la gradación dimensional indica que las inversiones en tecnología educativa para contextos periféricos deben priorizar la garantía de conectividad estable y equitativa sobre el desarrollo de materiales digitales sofisticados — un reordenamiento de prioridades que invierte la lógica habitual de los programas de tecnología educativa, los cuales tienden a concentrarse en la producción de contenidos sin resolver

primero las condiciones de acceso que determinan si esos contenidos pueden activar procesos neuroeducativos. En el nivel de diseño instruccional, la alta correlación de la interactividad con los procesos neuroeducativos establece que los entornos virtuales para preescolares deben organizarse alrededor de la participación activa del estudiante como principio rector, por encima de la complejidad visual de los contenidos o la sofisticación de la interfaz, especialmente cuando los recursos tecnológicos disponibles son limitados.

REFERENCIAS

- American Educational Research Association. (2011). *Code of ethics*. American Educational Research Association.
[https://www.aera.net/Portals/38/docs/About_AERA/CodeOfEthics\(1\).pdf](https://www.aera.net/Portals/38/docs/About_AERA/CodeOfEthics(1).pdf)
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142.
<https://doi.org/10.1177/0013164485451012>
- Ansari, D., De Smedt, B., & Grabner, R. H. (2012). Neuroeducation: A critical overview of an emerging field. *Neuroethics*, 5(2), 105–117. <https://doi.org/10.1007/s12152-011-9119-3>
- Behnamnia, N., Kamsin, A., Ismail, M., & Hayati, S. (2022). A review of using digital game-based learning for preschoolers. *Journal of Computers in Education*.
<https://doi.org/10.1007/s40692-022-00240-0>
- Bower, M. (2019). Technology-mediated learning theory. *British Journal of Educational Technology*, 50(3), 1035–1048. <https://doi.org/10.1111/bjet.12771>
- Cabrera, M., Maldonado, C., Calle, T., & Sánchez, A. (2024). The importance of neuroeducation in the teaching-learning of mathematics in preschoolers. *Salud, Ciencia y Tecnología - Serie de Conferencias*. <https://doi.org/10.56294/sctconf2024.1029>

- Cedeño, M. (2021). *Habilidades neuroeducativas en proceso de aprendizaje en nivel inicial* [Tesis de segunda especialidad]. Universidad César Vallejo.
- Cheung, W., & Ostrosky, M. (2023). Supporting preschoolers' motor development in virtual environments: Listening to teachers' voices. *Early Childhood Education Journal*, 1–10.
<https://doi.org/10.1007/s10643-023-01492-w>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Espino-Díaz, L., Fernández-Caminero, G., Hernández-Lloret, C., González-González, H., & Álvarez-Castillo, J. (2020). Analyzing the impact of COVID-19 on education professionals. Toward a paradigm shift: ICT and neuroeducation as a binomial of action. *Sustainability*, 12, 5646. <https://doi.org/10.3390/su12145646>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4.
<https://doi.org/10.11648/j.ajtas.20160501.11>
- Falloon, G. (2020). From digital literacy to digital competence: The teacher digital competency (TDC) framework. *Educational Technology Research and Development*, 68(5), 2449–2472. <https://doi.org/10.1007/s11423-020-09767-4>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911.
<https://doi.org/10.1037/0003-066X.34.10.906>

- Garrison, D. R., & Anderson, T. (2003). *E-learning in the 21st century: A framework for research and practice*. Routledge.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Allyn & Bacon.
- He, X., & Li, Y. (2023). Effect of learning methods and cognitive characteristics on preschoolers' online English attainment. *Education and Information Technologies*, 1–19. <https://doi.org/10.1007/s10639-022-11551-1>
- Hernández-Sampieri, R., & Mendoza, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education.
- Hu, X., Chiu, M., Leung, W., & Yelland, N. (2021). Technology integration for young children during COVID-19: Towards future online teaching. *British Journal of Educational Technology*, 52, 1513–1537. <https://doi.org/10.1111/bjet.13106>
- Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education*, 1(1), 3–10. <https://doi.org/10.1111/j.1751-228X.2007.00004.x>
- Instituto Nacional de Estadística e Informática. (2023). *Estadísticas de las tecnologías de información y comunicación en los hogares: IV trimestre 2023*. INEI. <https://www.inei.gob.pe/media/MenuRecursivo/boletines/>
- Lau, E., & Lee, K. (2020). Parents' views on young children's distance learning and screen time during COVID-19 class suspension in Hong Kong. *Early Education and Development*, 32, 863–880. <https://doi.org/10.1080/10409289.2020.1843925>
- Means, B., Toyama, Y., Murphy, R., & Bakia, M. (2014). *Learning online: What research tells us about whether, when and how*. Routledge.

- Merino-Soto, C., & Livia-Segovia, J. (2009). Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken. *Anales de Psicología*, 25(1), 169–171.
- Ministerio de Transportes y Comunicaciones. (2023). *Proyecto Conecta Selva: Instalación de internet satelital en comunidades rurales de Loreto, Amazonas, Madre de Dios y Ucayali*. MTC. <https://www.gob.pe/mtc>
- Molina, C. (2022). *Efectos de la educación en niños de 5 años en zonas rurales en Kalabamba y Kotaras, Apurímac* [Tesis de maestría]. Universidad César Vallejo.
- Naida, R., Berezovska, L., Bulgakova, O., Kravets, N., & Savchenkova, M. (2024). Integrating innovative pedagogical technologies into early childhood education training programs. *Conhecimento & Diversidade*. <https://doi.org/10.18316/rcd.v16i41.11551>
- Nurhayati, S., Hidayat, A., Zahra, D., Fitri, A., & Fasa, M. (2023). The effectiveness of virtual classroom learning in Islamic early childhood education. *KnE Social Sciences*. <https://doi.org/10.18502/kss.v8i4.12927>
- Oviedo, V., Cevallos, M., & Zavala, A. (2020). La neuroeducación en el aprendizaje de los preescolares. *UNESUM-Ciencias*, 4, 83–92. <https://doi.org/10.47230/unsum-ciencias.v4.n1.2021.231>
- Paredes, L. (2020). *Instrumento de las aulas virtuales y su impacto en el aprendizaje de niños de 4 años en las escuelas en Lima* [Tesis de maestría]. Universidad César Vallejo.
- Pellegrini, A. D. (2004). *Observing children in their natural worlds: A methodological primer* (2nd ed.). Lawrence Erlbaum Associates.
- Pineda, M., Lara, P., Cornejo, M., Condori, E., & Choque, D. (2025). Neuroeducation in initial basic education: Strategies and benefits for early learning. *Seminars in Medical Writing and Education*. <https://doi.org/10.56294/mw2024465>

- Pottosina, Y., & Lambrechts, L. (2023). Transforming early education: The evolution of online learning for pre-schoolers. *Child in a Digital World*.
<https://doi.org/10.61365/forum.2023.024>
- Putri, R., Arum, R., & Arum, C. (2025). Use of digital educational media for early childhood cognitive stimulation. *Early Childhood Research Journal (ECRJ)*.
<https://doi.org/10.23917/ecrj.v7i2.8782>
- Rapti, S., Sapounidis, T., & Tselegkaridis, S. (2024). Investigating educators' and students' perspectives on virtual reality enhanced teaching in preschool. *Early Childhood Education Journal*, 53, 1107–1118. <https://doi.org/10.1007/s10643-024-01659-z>
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Rosenthal, R. (2002). Covert communication in classrooms, clinics, courtrooms, and cubicles. *American Psychologist*, 57(11), 839–849. <https://doi.org/10.1037/0003-066X.57.11.839>
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). McGraw-Hill.
- Stites, M., Sonneschein, S., & Galczyk, S. (2021). Preschool parents' views of distance learning during COVID-19. *Early Education and Development*, 32, 923–939.
<https://doi.org/10.1080/10409289.2021.1930936>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292.
<https://doi.org/10.1007/s10648-019-09465-5>

- Fondo de las Naciones Unidas para la Infancia. (2021). *COVID-19 and school closures: One year of education disruption*. UNICEF. <https://data.unicef.org/resources/one-year-of-covid-19-and-school-closures/>
- Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., Almeqdad, Q., & Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning*, 4(1), 63–85. <https://doi.org/10.1007/s11409-008-9033-1>
- Williams, R., Dowker, A., Evers, C., Pradeep, K., Sulur, R., Priya, A., Thangavelu, A. P., Aswathy, S., Jisha, V. G., & Vaisakhi, V. S. (2024). Neuroeducation: Understanding neural dynamics in learning and teaching. *Frontiers in Education*, 9, 1437418. <https://doi.org/10.3389/feduc.2024.1437418>
- Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., Solis, S. L., & Whitebread, D. (2022). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.824680>

CONCLUSIONES GENERALES

Los resultados integrados a lo largo de los siete capítulos permiten afirmar que el aprendizaje, en los contextos educativos contemporáneos, debe ser comprendido como un fenómeno inherentemente multifactorial, en el que convergen dimensiones biológicas, cognitivas y tecnológicas que interactúan de manera dinámica. La evidencia presentada muestra que variables como la calidad del sueño, la estimulación multisensorial, la exposición a experiencias musicales y el uso de tecnologías educativas no actúan de forma aislada, sino que configuran un entramado de influencias que condiciona el desempeño académico y los procesos de adquisición de conocimientos. Esta perspectiva integrada permite superar enfoques reduccionistas que han intentado explicar el aprendizaje desde una única dimensión, evidenciando la necesidad de modelos más complejos que contemplen la interdependencia entre factores internos y externos al estudiante.

Asimismo, los hallazgos ponen de manifiesto una variabilidad significativa en los efectos observados, tanto en función del nivel educativo como del tipo de intervención analizada. Mientras que en contextos de educación inicial se evidencia una asociación más clara entre estrategias neuroeducativas y aprendizaje multisensorial, en niveles educativos superiores los resultados tienden a ser más heterogéneos, particularmente en el caso de variables como la calidad del sueño y el rendimiento académico. De igual forma, las intervenciones basadas en tecnología educativa muestran efectos diferenciados según su diseño pedagógico y su grado de alineación con principios neurocognitivos. Esta heterogeneidad no debe interpretarse como una debilidad de la evidencia, sino como una manifestación de la complejidad del fenómeno

educativo, que exige ser abordado desde perspectivas analíticas más sofisticadas y contextualizadas.

Desde el punto de vista teórico, los resultados obtenidos refuerzan la validez de la neuroeducación como marco integrador para el estudio del aprendizaje. La convergencia entre los hallazgos empíricos y los principios fundamentales de este campo —como la neuroplasticidad, la centralidad de la atención y el papel modulador de la emoción— sugiere que los procesos educativos pueden beneficiarse significativamente de una comprensión más profunda del funcionamiento cerebral. En particular, la evidencia relativa a la consolidación de la memoria, el impacto de la estimulación multisensorial y la influencia de factores fisiológicos como el sueño aporta sustento a la idea de que el aprendizaje implica transformaciones estructurales en el cerebro que son sensibles a las condiciones del entorno y a las experiencias educativas.

No obstante, los resultados también permiten problematizar la tendencia a establecer relaciones directas y lineales entre variables educativas. En varios de los estudios analizados, especialmente aquellos de naturaleza correlacional, no se observaron asociaciones estadísticamente significativas o se identificaron relaciones de magnitud débil entre variables que teóricamente se esperaba que estuvieran fuertemente vinculadas. Este hallazgo pone en evidencia la necesidad de adoptar modelos explicativos que incorporen variables mediadoras y moderadoras, tales como la motivación, el contexto socioeducativo, la regulación emocional y los hábitos de estudio. En este sentido, el aprendizaje se configura como un proceso no lineal, en el que los efectos de una variable pueden depender de la presencia o ausencia de otras, lo que exige un enfoque analítico más complejo y multidimensional.

En términos de implicaciones prácticas, los resultados de este libro sugieren la necesidad de replantear el diseño de las estrategias pedagógicas desde una perspectiva informada por la evidencia neuroeducativa. En primer lugar, se destaca la importancia de considerar los procesos

cognitivos básicos —como la atención y la memoria— en la planificación de las actividades de enseñanza, promoviendo metodologías activas que favorezcan la participación del estudiante y la consolidación del aprendizaje. En segundo lugar, se subraya el valor del aprendizaje multisensorial, especialmente en etapas tempranas, como un medio para potenciar la codificación de la información y facilitar la adquisición de habilidades fundamentales. En tercer lugar, se reconoce el potencial de la tecnología educativa como herramienta para enriquecer los entornos de aprendizaje, siempre que su uso esté guiado por principios pedagógicos claros y no se limite a una incorporación superficial o instrumental.

A nivel institucional, estos hallazgos ponen de relieve la necesidad de fortalecer la formación docente en el campo de la neuroeducación y la tecnología educativa. La capacidad de los docentes para interpretar la evidencia científica y traducirla en prácticas pedagógicas efectivas constituye un factor clave para mejorar la calidad de la educación. Asimismo, se hace evidente la importancia de promover políticas educativas que favorezcan la integración de enfoques interdisciplinarios y que reconozcan la complejidad de los procesos de aprendizaje. En este sentido, la articulación entre investigación, formación docente y práctica educativa se presenta como un elemento fundamental para avanzar hacia sistemas educativos más coherentes y basados en evidencia.

No obstante, es necesario reconocer las limitaciones inherentes al conjunto de estudios que conforman esta obra. En primer lugar, una parte importante de la evidencia empírica se basa en diseños no experimentales, lo que limita la posibilidad de establecer relaciones causales entre las variables analizadas. En segundo lugar, los tamaños muestrales en algunos estudios son reducidos y se circunscriben a contextos educativos específicos, lo que restringe la generalización de los resultados. En tercer lugar, la heterogeneidad en los instrumentos de medición y en los indicadores de aprendizaje dificulta la comparación directa entre estudios. Finalmente, la ausencia de diseños longitudinales impide analizar la evolución de los efectos a

lo largo del tiempo, lo que constituye una limitación relevante en el estudio de procesos complejos como el aprendizaje.

A partir de estas limitaciones, se abren diversas líneas de investigación futura que resultan fundamentales para el desarrollo del campo. En primer lugar, se recomienda la implementación de estudios longitudinales que permitan analizar la evolución de los procesos de aprendizaje y la estabilidad de los efectos observados. En segundo lugar, se sugiere el desarrollo de modelos multivariados que integren múltiples variables neuroeducativas, cognitivas y contextuales, con el fin de comprender mejor las interacciones entre ellas. En tercer lugar, se destaca la necesidad de ampliar la investigación en contextos latinoamericanos, donde las particularidades socioculturales y educativas pueden influir de manera significativa en los procesos de aprendizaje. Finalmente, el avance de tecnologías como la inteligencia artificial generativa plantea nuevas oportunidades y desafíos que requieren ser explorados desde una perspectiva crítica y fundamentada en la evidencia.



ISBN: 978-9907-834-01-7



9 789907 834017