

Evaluación formativa mediada por inteligencia artificial generativa en Educación General Básica.

Formative Assessment Mediated by Generative Artificial Intelligence in Basic General Education

Sánchez-Vásquez, María Cristina¹; Sánchez-Vásquez, María Eugenia²; Sánchez-Sánchez, Diana Katherine³; Vera-Vera, Ney Agustín⁴.

¹ U. E. Juan Antonio Vergara Alcívar; Ecuador, Portoviejo; <https://orcid.org/0009-0001-7062-3639>; mariac.sanchez@docentes.educacion.edu.ec

² U. E. Soberanía Esmeraldeña; Ecuador, La Concordia; <https://orcid.org/0009-0004-8686-1860>; eugesanchez523@gmail.com

³ Escuela de Educación Básica San Eloy; Ecuador, Portoviejo; <https://orcid.org/0009-0001-9178-0756>; diana.sanchez@docentes.educacion.edu.ec

⁴ Universidad Técnica de Manabí; Ecuador, Junín; <https://orcid.org/0009-0008-3594-3128>; nvera0129@utm.edu.ec

Cita: Sánchez-Vásquez, M. C., Sánchez-Vásquez, M. E., Sánchez-Sánchez, D. K., & Vera-Vera, N. A. (2026). Evaluación formativa mediada por inteligencia artificial generativa en Educación General Básica. *Innova Science Journal*, 4(2), 845-856. <https://doi.org/10.63618/omd/isj/v4/n2/318>

Recibido: 20/12/2025

Aceptado: 20/04/2026

Publicado: 30/04/2026



Copyright: © 2026 por los autores. Este artículo es un artículo de acceso abierto distribuido bajo los términos y condiciones de la [Licencia Creative Commons, Atribución-NoComercial 4.0 Internacional. \(CC BY-NC\)](https://creativecommons.org/licenses/by-nc/4.0/).

(<https://creativecommons.org/licenses/by-nc/4.0/>)

¹ Autor Correspondencia

 <https://doi.org/10.63618/omd/isj/v4/n2/318>

Resumen: La evaluación formativa mediada por inteligencia artificial generativa constituye un campo emergente en Educación General Básica, debido a su potencial para fortalecer la retroalimentación, la personalización del aprendizaje y la toma de decisiones docentes. Este estudio tuvo como objetivo analizar la producción científica reciente sobre sus aportes pedagógicos, riesgos éticos y condiciones de implementación. Se desarrolló una revisión bibliográfica documental, descriptiva-analítica y cualitativa, basada en 17 estudios publicados entre 2021 y 2026, organizados en una matriz de 12 campos. Los resultados evidenciaron que la IA generativa puede apoyar la retroalimentación inmediata, el diseño de rúbricas, la identificación de dificultades y la autorregulación estudiantil, siempre que exista mediación docente. Se concluye que su valor formativo no depende de la automatización, sino de su integración crítica, ética y pedagógicamente orientada.

Palabras clave: evaluación formativa; inteligencia artificial generativa; Educación General Básica; retroalimentación; mediación docente.

Abstract: Formative assessment mediated by generative artificial intelligence constitutes an emerging field in Basic General Education because of its potential to strengthen feedback, learning personalization, and teachers' decision-making. This study aimed to analyze recent scientific production on its pedagogical contributions, ethical risks, and implementation conditions. A documentary, descriptive-analytical, and qualitative literature review was conducted, based on 17 studies published between 2021 and 2026 and organized through a 12-field documentary matrix. The findings showed that generative AI can support immediate feedback, rubric design, identification of learning difficulties, and student self-regulation, provided that teacher mediation is maintained. It is concluded that its formative value does not depend on automation itself, but on its critical, ethical, and pedagogically guided integration into teaching and assessment practices.

Keywords: formative assessment; generative artificial intelligence; Basic General Education; feedback; teacher mediation

1. Introducción

La educación contemporánea se encuentra en una etapa de transición pedagógica marcada por la expansión de tecnologías inteligentes, plataformas adaptativas y sistemas capaces de generar texto, retroalimentación y recursos didácticos en tiempo real. En este escenario, la inteligencia artificial generativa deja de ser un asunto exclusivamente técnico y empieza a ubicarse en el centro de las decisiones educativas, especialmente cuando se vincula con la enseñanza, la evaluación y el acompañamiento del aprendizaje. Kasneci et al. (2023) sostienen que los grandes modelos de lenguaje abren posibilidades para personalizar contenidos, apoyar la interacción educativa y ampliar los recursos disponibles para docentes y estudiantes; sin embargo, también advierten que su uso exige alfabetización crítica, criterios éticos y comprensión de sus límites. Esta tensión convierte a la inteligencia artificial generativa en un campo pertinente para la investigación educativa actual, no por su novedad tecnológica, sino por su capacidad de reconfigurar prácticas pedagógicas sensibles.

Desde una mirada amplia, la evaluación constituye uno de los núcleos más decisivos del proceso educativo, porque permite reconocer avances, dificultades, ritmos y necesidades de aprendizaje. En particular, la evaluación formativa adquiere relevancia porque no se orienta únicamente a calificar resultados, sino a producir evidencias útiles para mejorar la enseñanza y apoyar la autorregulación del estudiante. Xuan et al. (2022) identifican que la evaluación formativa tiene efectos positivos, aunque moderados, en el rendimiento lector de estudiantes K-12, y subrayan que su efectividad depende de factores como la diferenciación pedagógica, la participación del estudiante y la adaptación cultural de las estrategias. En la misma línea, Sortwell et al. (2024) muestran que la evaluación formativa en educación básica y secundaria presenta efectos favorables sobre el aprendizaje, aunque con variaciones según el tipo de intervención, la calidad metodológica y el contexto de aplicación.

El vínculo entre evaluación formativa e inteligencia artificial se fortalece porque ambas dimensiones convergen en un mismo punto: la necesidad de generar retroalimentación oportuna, comprensible y accionable. Swiecki et al. (2022) plantean que las prácticas tradicionales de evaluación suelen ofrecer fotografías parciales del desempeño estudiantil, por lo que la inteligencia artificial puede contribuir a construir miradas más continuas, contextualizadas y adaptativas del aprendizaje. No obstante, Hopfenbeck et al. (2023) precisan que la inteligencia artificial no debe reemplazar las prácticas docentes de evaluación formativa, sino facilitar procesos como la retroalimentación, la autoevaluación, la coevaluación y el uso pedagógico de evidencias. Esta afirmación es clave para el presente estudio, porque sitúa la mediación tecnológica dentro de una lógica pedagógica, no dentro de una lógica de sustitución profesional.

En un nivel más específico, la inteligencia artificial generativa introduce nuevas posibilidades para la Educación General Básica, al permitir la elaboración de comentarios personalizados, preguntas orientadoras, rúbricas, ejemplos graduados, reformulación de instrucciones y apoyos diferenciados según niveles de desempeño. Alfarwan (2025) señala que la investigación sobre inteligencia artificial generativa en educación K-12 aún se encuentra en consolidación y requiere más evidencia sobre usos reales en el aula, más allá de percepciones generales o debates abstractos. Lin y Tan (2025), por su parte, describen la inteligencia artificial generativa en K-12 como un

sistema donde objetivos, actividades, roles y resultados educativos evolucionan de manera interdependiente. Esta lectura resulta especialmente útil para Educación General Básica, porque el aprendizaje escolar no depende solo del acceso a una herramienta, sino de cómo el docente organiza la tarea, interpreta la respuesta generada y acompaña la mejora del estudiante.

A pesar de sus posibilidades, el uso de inteligencia artificial generativa en evaluación formativa plantea un problema científico relevante: todavía no existe suficiente claridad sobre las condiciones pedagógicas, éticas y metodológicas que permiten convertir la retroalimentación automatizada en una experiencia realmente formativa. Farrokhnia et al. (2024) advierten que ChatGPT y herramientas similares ofrecen oportunidades educativas, pero también presentan amenazas relacionadas con la comprensión limitada del contexto, la integridad académica, la discriminación algorítmica y la reducción de habilidades cognitivas superiores. Tlili et al. (2023) también muestran que el entusiasmo por los chatbots educativos convive con preocupaciones sobre precisión, dependencia, sesgos y uso acrítico de la información. Por ello, el problema no consiste únicamente en saber si la inteligencia artificial generativa “sirve” para evaluar, sino en determinar cuándo, cómo, con qué criterios y bajo qué supervisión docente puede aportar a una evaluación formativa justa, contextualizada y pedagógicamente válida.

La pertinencia de este estudio se justifica en la necesidad de revisar críticamente la literatura reciente para evitar dos extremos igualmente débiles: la adopción acrítica de la inteligencia artificial como solución automática y el rechazo absoluto que desconoce sus usos pedagógicos emergentes. Zhang et al. (2026) reconocen que la inteligencia artificial generativa puede favorecer el rendimiento académico, la motivación y la personalización del aprendizaje en K-12, pero también advierten riesgos vinculados con contenidos erróneos, dependencia tecnológica, dificultades de comprensión y privacidad. Maksimchuk et al. (2025) refuerzan esta perspectiva al sostener que la inteligencia artificial generativa puede apoyar la evaluación formativa en K-12 mediante retroalimentación y aplicaciones prácticas, siempre que se mantenga el juicio docente y un sistema equilibrado de evaluación. En consecuencia, una revisión bibliográfica sobre este tema permite ordenar evidencias, identificar vacíos y proponer criterios para una integración responsable en Educación General Básica.

El presente trabajo tiene como objetivo principal analizar la producción científica reciente sobre la evaluación formativa mediada por inteligencia artificial generativa en Educación General Básica, con énfasis en sus aportes pedagógicos, riesgos éticos y condiciones de implementación docente. Se parte del supuesto de que la inteligencia artificial generativa puede enriquecer la evaluación formativa cuando se emplea como mediación orientada a mejorar la retroalimentación, la autorregulación y la toma de decisiones pedagógicas; no obstante, su valor educativo depende de la calidad de los criterios de evaluación, de la supervisión profesional del docente y de la protección de principios como equidad, transparencia, privacidad y pertinencia contextual. Así, esta revisión no busca celebrar la tecnología, sino comprender críticamente su lugar dentro de una evaluación que sigue siendo, ante todo, una práctica humana de acompañamiento, mejora y justicia educativa.

2. Materiales y Métodos

La investigación se desarrolló bajo un diseño documental, de tipo revisión bibliográfica, nivel descriptivo-analítico y modalidad cualitativa, orientado a examinar la producción científica reciente sobre evaluación formativa mediada por inteligencia artificial generativa en Educación General Básica. El estudio se apoyó en un procedimiento de búsqueda, selección y síntesis inspirado en las recomendaciones de transparencia de la declaración PRISMA 2020, propuesta por Page et al. (2021), debido a que este protocolo permitió organizar con mayor claridad las fases de identificación, cribado, elegibilidad e inclusión de documentos. Para delimitar el campo de análisis, se utilizaron como descriptores principales las expresiones: “evaluación formativa”, “formative assessment”, “inteligencia artificial generativa”, “generative artificial intelligence”, “ChatGPT”, “large language models”, “Educación General Básica”, “basic education” y “K-12 education”. La búsqueda se realizó en bases académicas y repositorios de amplia circulación científica, como Scopus, Web of Science, ERIC, SciELO, Redalyc, Google Scholar y bases editoriales de acceso abierto, considerando que la literatura reciente reconoce el crecimiento acelerado de investigaciones sobre IA generativa en educación K-12, aunque con vacíos todavía visibles en prácticas reales de aula y evaluación formativa.

La población documental estuvo conformada por artículos científicos, revisiones sistemáticas, revisiones bibliográficas, capítulos académicos y documentos especializados publicados entre 2021 y 2026, en idioma español, inglés o portugués, relacionados con inteligencia artificial generativa, evaluación formativa, retroalimentación, autorregulación del aprendizaje y educación básica. Se incluyeron estudios que abordaron de manera explícita el uso pedagógico de herramientas generativas en contextos escolares, procesos de retroalimentación, diseño de rúbricas, acompañamiento docente, personalización del aprendizaje o análisis ético de la IA en educación. Se excluyeron documentos centrados exclusivamente en educación superior, textos de opinión sin respaldo metodológico, publicaciones duplicadas, trabajos sin acceso al resumen o al texto completo, estudios sobre inteligencia artificial no generativa y documentos que trataron la evaluación únicamente desde una perspectiva sumativa o estandarizada. También se eliminaron registros que, luego de la lectura del título, resumen y palabras clave, no presentaron relación directa con el problema científico del estudio. Esta delimitación resultó pertinente porque investigaciones recientes, como las de Maksimchuk et al. (2025), han señalado que la IA generativa puede apoyar la evaluación formativa en K-12 mediante retroalimentación, diseño de tareas y acompañamiento del aprendizaje, pero siempre con mediación docente y criterios éticos claros.

El procedimiento metodológico se organizó en cinco fases: primero, se definieron las palabras clave y ecuaciones de búsqueda mediante operadores booleanos; segundo, se identificaron los documentos disponibles en las bases seleccionadas; tercero, se realizó una depuración inicial por año, idioma, tipo de publicación y pertinencia temática; cuarto, se efectuó una lectura analítica de los textos seleccionados; y quinto, se sistematizó la información en una matriz documental con campos referidos a autor, año, país o contexto, objetivo, metodología, nivel educativo, relación con la evaluación formativa, uso de IA generativa, principales hallazgos, riesgos señalados y aporte al presente estudio. El análisis se realizó mediante codificación temática, agrupando los

aportes en tres categorías centrales: contribuciones pedagógicas, riesgos éticos y condiciones de implementación docente. En cuanto a los aspectos éticos, no se trabajó con seres humanos, datos sensibles ni intervención directa en instituciones educativas; por tanto, no se requirió consentimiento informado ni autorización institucional. Sin embargo, se respetaron los principios de integridad académica, citación responsable, trazabilidad de fuentes y uso transparente de la información, especialmente porque la literatura reciente advierte que la IA generativa en educación requiere atención a la privacidad, dependencia tecnológica, precisión de la información y protección del juicio profesional docente.

3. Resultados

Esta sección puede dividirse por subtítulos. Debe proporcionar una descripción concisaEl análisis documental permitió identificar 17 estudios publicados entre 2021 y 2026 que aportaron evidencia directa o complementaria sobre la evaluación formativa mediada por inteligencia artificial generativa en Educación General Básica. El corpus seleccionado integró revisiones sistemáticas, metaanálisis, artículos empíricos, estudios conceptuales y trabajos de reflexión académica. La evidencia mostró que el tema se encontró en una fase de consolidación, debido a que la evaluación formativa ya contaba con una base empírica sólida en K-12, mientras que la inteligencia artificial generativa apareció como una mediación emergente que todavía requería criterios pedagógicos, éticos y metodológicos más precisos. Xuan et al. (2022), Yao et al. (2024) y Sortwell et al. (2024) confirmaron la relevancia de la evaluación formativa en el aprendizaje escolar, mientras que Alfarwan (2025), Zhang et al. (2026) y Martin et al. (2024) ubicaron la inteligencia artificial como un campo de expansión reciente dentro de la educación K-12.

3.1. Caracterización del corpus documental revisado

La revisión permitió agrupar los estudios en cuatro núcleos temáticos: a) investigaciones sobre evaluación formativa en K-12; b) estudios sobre inteligencia artificial en educación escolar; c) publicaciones centradas en IA generativa, ChatGPT o modelos de lenguaje; y d) trabajos que vincularon de forma directa la IA generativa con retroalimentación, acompañamiento o evaluación formativa. Esta organización permitió reconocer que la literatura más fuerte no se concentró únicamente en herramientas tecnológicas, sino en la relación entre retroalimentación, toma de decisiones pedagógicas, autorregulación del aprendizaje y mediación docente. En este sentido, la matriz compacta permitió visualizar los aportes centrales de cada documento, como se muestra en la tabla 1.

Tabla 1.

Matriz documental compacta de estudios incluidos sobre evaluación formativa e inteligencia artificial generativa

Nº	Referencia abreviada	Contexto/nivel	Diseño o tipo de estudio	Aporte principal al presente estudio
1	Maksimchuk et al. (2025)	K-12	Paper de congreso	Vinculó directamente IA generativa y evaluación formativa, destacando

				feedback, diseño de tareas y juicio docente.
2	Meyer et al. (2024)	Secundaria	Estudio empírico	Evidenció que el feedback generado por LLM mejoró revisión textual, motivación y emociones positivas.
3	Steiss et al. (2024)	Escritura escolar	Estudio comparativo	Comparó feedback humano y ChatGPT según criterios de claridad, precisión, tono y mejora.
4	Cheah et al. (2025)	K-12	Encuesta docente	Mostró preparación, prácticas y barreras de docentes para integrar IA generativa.
5	Alfarwan (2025)	K-12	Revisión sistemática	Identificó usos, impactos y vacíos de la IA generativa en educación escolar.
6	Zhang et al. (2026)	K-12	Revisión sistemática	Sistematizó tendencias, beneficios y desafíos de la IA generativa en educación K-12.
7	Martin et al. (2024)	K-12	Revisión sistemática	Delimitó aplicaciones de IA en aprendizaje, evaluación, enseñanza y analítica escolar.
8	Crompton y Burke (2024)	Educación general	Revisión sistemática	Analizó usos educativos de ChatGPT y sus desafíos para docentes y estudiantes.
9	Kasneji et al. (2023)	Educación general	Artículo conceptual	Explicó oportunidades y riesgos de los grandes modelos de lenguaje en educación.
10	Luckin et al. (2022)	Formación docente	Artículo conceptual	Fundamentó la preparación docente para comprender y usar IA con sentido educativo.
11	Xuan et al. (2022)	K-12	Metaanálisis	Confirmó efecto positivo, aunque moderado, de la evaluación formativa en lectura.
12	Yao et al. (2024)	K-12	Metaanálisis	Confirmó la utilidad de la evaluación formativa en el rendimiento académico escolar.
13	Sortwell et al. (2024)	K-12	Revisión de metaanálisis	Mostró efectos positivos variables de la evaluación formativa en el aprendizaje.

14	Proaño Sánchez et al. (2025)	Educación básica	Revisión sistemática	Aportó evidencia en español sobre IA generativa, competencias y escenarios interdisciplinarios.
15	Chiluisa Aimara et al. (2025)	Ecuador/Educación básica	Artículo aplicado	Relacionó IA, retroalimentación en tiempo real y progreso académico en educación básica.
16	Trujillo Casadiego (2024)	Aula escolar	Artículo de reflexión	Analizó ChatGPT como apoyo para retroalimentación personalizada y autonomía estudiantil.
17	Zapata-Ros (2024)	Iberoamérica	Artículo teórico	Problematizó la IA generativa como reto para la evaluación y la pedagogía.

Nota: Elaborado por los autores

En la tabla 1 se observó que la producción científica revisada no presentó una distribución homogénea. Los estudios sobre evaluación formativa en K-12 ofrecieron una base empírica más consolidada, especialmente mediante metaanálisis y revisiones de metaanálisis, mientras que los estudios sobre inteligencia artificial generativa se concentraron en revisiones recientes, experiencias iniciales y análisis de oportunidades y riesgos. Este resultado indicó que el campo todavía se encontraba en una etapa de transición: existía claridad sobre la importancia de evaluar formativamente, pero la mediación con IA generativa aún necesitaba más estudios empíricos en Educación General Básica, especialmente en contextos latinoamericanos.

3.2. Aportes pedagógicos identificados

Los resultados mostraron que el aporte pedagógico más recurrente de la inteligencia artificial generativa se relacionó con la retroalimentación inmediata, personalizada y orientada a la mejora. Meyer et al. (2024) reportaron que el feedback generado por modelos de lenguaje incrementó el desempeño en revisión textual, la motivación hacia la tarea y las emociones positivas en estudiantes de secundaria; este hallazgo resultó relevante porque la evaluación formativa dependió, en gran medida, de que el estudiante recibiera información comprensible para mejorar su producción. Steiss et al. (2024) también evidenciaron que ChatGPT pudo producir retroalimentación formativa valorable cuando se evaluó con criterios como claridad, precisión, tono de apoyo y orientación a la mejora.

Otro resultado relevante fue que la IA generativa amplió las posibilidades de apoyo al docente en tareas asociadas con la evaluación: generación de preguntas, construcción de rúbricas, identificación de errores frecuentes, adaptación de instrucciones y elaboración de comentarios diferenciados. Maksimchuk et al. (2025) destacaron que la IA generativa pudo fortalecer prácticas de evaluación formativa en K-12 siempre que se conservara el juicio docente y un sistema equilibrado de evaluación. De forma complementaria, Trujillo (2024) señaló que ChatGPT favoreció la retroalimentación personalizada, la identificación de patrones de error y la autonomía del estudiante; sin

embargo, este aporte se interpretó como mediación pedagógica y no como sustitución del acompañamiento humano.

3.3. Riesgos, límites y tensiones éticas

El análisis también permitió identificar riesgos que condicionaron el valor educativo de la inteligencia artificial generativa. Entre los más frecuentes se encontraron la producción de respuestas incorrectas, los sesgos algorítmicos, la dependencia tecnológica, la privacidad de datos, el plagio, la reducción del pensamiento crítico y la posibilidad de que la retroalimentación automatizada se aceptara sin revisión docente. Kasneci et al. (2023) advirtieron que los grandes modelos de lenguaje ofrecieron beneficios para personalización e interacción educativa, pero exigieron nuevas alfabetizaciones para comprender sus límites. Farrokhnia et al. (2024) añadieron que ChatGPT presentó fortalezas como respuestas personalizadas y en tiempo real, aunque también amenazas asociadas con exactitud, integridad académica y uso acrítico.

En los estudios revisados en español se repitió una preocupación similar: la IA generativa podía enriquecer el aprendizaje, pero su incorporación en educación básica exigía infraestructura, formación docente, criterios éticos y acompañamiento institucional. Proaño et al. (2025) identificaron oportunidades para el desarrollo de competencias cognitivas e interdisciplinarias, aunque también señalaron cuestiones éticas emergentes. Chiluisa et al. (2025) relacionaron la IA con retroalimentación en tiempo real y progreso académico, pero reconocieron desafíos vinculados con infraestructura y capacitación. Estos hallazgos resultaron importantes porque conectaron el debate global con realidades educativas más cercanas a América Latina y Ecuador.

3.4. Condiciones de implementación docente

Los resultados evidenciaron que la mediación docente fue la condición más importante para que la IA generativa aportara a la evaluación formativa. Cheah et al. (2025) mostraron que la integración cotidiana de GenAI en K-12 todavía presentó barreras relacionadas con preparación, prácticas docentes y contexto escolar. Luckin et al. (2022) reforzaron esta idea al plantear que la preparación para la IA no consistió solo en aprender herramientas, sino en desarrollar una comprensión contextualizada de la tecnología, sus usos y sus efectos en profesiones concretas. Por ello, la evidencia revisada sugirió que el docente necesitó actuar como diseñador de criterios, curador de respuestas, intérprete de evidencias y garante ético del proceso evaluativo.

La síntesis de resultados permitió establecer que la IA generativa tuvo mayor valor cuando se integró en una secuencia evaluativa clara: definición de criterios de logro, recolección de evidencias, generación de retroalimentación, revisión docente, participación del estudiante y ajuste de la enseñanza. Hopfenbeck et al. (2023) sostuvieron que la IA podía apoyar la evaluación formativa a escala, especialmente en el aula, pero este apoyo debía comprenderse dentro de prácticas pedagógicas responsables. En consecuencia, el resultado central de la revisión fue que la IA generativa no funcionó como una solución autónoma para evaluar mejor, sino como una mediación que dependió de la calidad de los criterios, de la supervisión docente y de la participación activa del estudiante.

En conjunto, los resultados respondieron al objetivo del estudio al mostrar que la evaluación formativa mediada por inteligencia artificial generativa en Educación General Básica se configuró como un campo promisorio, pero todavía inestable en términos de evidencia empírica específica. La literatura revisada permitió sostener que sus principales aportes se ubicaron en la retroalimentación personalizada, la eficiencia docente, la detección temprana de dificultades y el fortalecimiento de la autorregulación; mientras que sus límites se asociaron con sesgos, privacidad, dependencia, desigualdad de acceso y pérdida de profundidad pedagógica. Así, el hallazgo integrador fue que la IA generativa solo adquirió valor formativo cuando se subordinó a una intención pedagógica clara, humanizada y éticamente supervisada.

4. Discusión

Los resultados de esta revisión confirman que la evaluación formativa mediada por inteligencia artificial generativa constituye un campo pertinente, pero todavía en consolidación dentro de la Educación General Básica. La evidencia revisada permite interpretar que la fortaleza del tema no reside únicamente en la novedad tecnológica, sino en su relación con un problema pedagógico persistente: cómo ofrecer retroalimentación oportuna, comprensible y útil para mejorar el aprendizaje. En este sentido, Xuan et al. (2022) y Yao et al. (2024) sostienen que la evaluación formativa tiene efectos positivos en el aprendizaje K-12, aunque estos varían según la calidad de la intervención, el contexto y la forma en que el docente utiliza la evidencia. De manera complementaria, Sortwell et al. (2024) muestran que los efectos de la evaluación formativa pueden ser favorables, pero no deben interpretarse de forma automática ni homogénea. Estos antecedentes permiten valorar los hallazgos del presente estudio con prudencia: la IA generativa no crea por sí misma una evaluación más formativa; su aporte aparece cuando fortalece procesos ya pedagógicamente intencionados, como la retroalimentación, la revisión de errores, la autoevaluación y el ajuste de la enseñanza.

En relación con la hipótesis de trabajo, los resultados respaldan la idea de que la inteligencia artificial generativa puede enriquecer la evaluación formativa cuando funciona como mediación docente y no como reemplazo del criterio profesional. Meyer et al. (2024) evidencian que la retroalimentación generada por modelos de lenguaje puede aumentar la revisión textual, la motivación y las emociones positivas de estudiantes de secundaria; Steiss et al. (2024) también muestran que ChatGPT puede producir retroalimentación formativa valorable cuando se analiza según criterios de claridad, precisión, orientación a la mejora y tono de apoyo. Sin embargo, Hopfenbeck et al. (2023) advierten que la IA debe facilitar, no sustituir, las prácticas docentes de evaluación formativa, mientras que Maksimchuk et al. (2025) insisten en la importancia del juicio docente y de sistemas evaluativos equilibrados. Desde esta perspectiva, el alcance del estudio se ubica en una zona crítica: la IA generativa puede ampliar la frecuencia, variedad y personalización del feedback, pero su calidad formativa depende de que el docente revise, contextualice y transforme la respuesta automática en una orientación pedagógica significativa.

Las limitaciones identificadas muestran que la integración de IA generativa en Educación General Básica todavía enfrenta desafíos importantes de evidencia, ética y formación docente. Alfarwan (2025) señala que la investigación sobre IA generativa en K-12 crece con rapidez, pero aún requiere más estudios empíricos situados en aulas

reales; Cheah et al. (2025) confirma que la preparación, las prácticas y las barreras docentes siguen siendo factores decisivos para su implementación. A ello se suman los riesgos señalados por Kasneci et al. (2023) y Farrokhnia et al. (2024), quienes advierten sobre sesgos, errores, dependencia tecnológica, integridad académica y necesidad de alfabetización crítica. Por tanto, las futuras investigaciones deberían avanzar hacia diseños empíricos, longitudinales y contextualizados en educación básica, especialmente en América Latina, para examinar cómo la IA generativa afecta la retroalimentación, la autorregulación, la equidad y la toma de decisiones docentes. En consecuencia, este estudio aporta una lectura crítica del campo: la IA generativa puede ser una aliada de la evaluación formativa, pero solo cuando se subordina a criterios pedagógicos, éticos y humanos claramente definidos.

5. Conclusiones

La revisión bibliográfica permitió cumplir el objetivo de analizar la producción científica reciente sobre la evaluación formativa mediada por inteligencia artificial generativa en Educación General Básica, identificando que su valor no depende de la automatización de la evaluación, sino de su capacidad para ampliar las oportunidades de retroalimentación, personalización, seguimiento del aprendizaje y apoyo a la toma de decisiones docentes. Los resultados permiten concluir que la IA generativa puede convertirse en una mediación pedagógica relevante cuando se integra a procesos formativos con criterios claros, revisión humana y sentido ético; sin embargo, pierde valor educativo cuando se utiliza como sustituto del juicio profesional, como corrector automático o como recurso aislado de la planificación didáctica. El principal aporte de esta investigación consiste en ordenar críticamente un campo emergente y mostrar que la evaluación formativa con IA generativa exige una articulación equilibrada entre tecnología, pedagogía y responsabilidad docente. En consecuencia, la Educación General Básica requiere avanzar hacia modelos de implementación donde la IA apoye la retroalimentación inmediata, la autorregulación estudiantil y la identificación temprana de dificultades, sin debilitar la interacción humana que sostiene el aprendizaje escolar. Esta conclusión reafirma que la innovación evaluativa no se legitima por el uso de herramientas avanzadas, sino por su contribución real a una enseñanza más justa, contextualizada y orientada a la mejora continua.

Referencias Bibliográficas

- Alfarwan, A. (2025). Generative AI use in K-12 education: A systematic review. *Frontiers in Education*, 10. <https://doi.org/10.3389/educ.2025.1647573>
- Cheah, Y. H., Lu, J., y Kim, J. (2025). Integrating generative artificial intelligence in K-12 education: Examining teachers' preparedness, practices, and barriers. *Computers and Education: Artificial Intelligence*, 8, 100363. <https://doi.org/10.1016/j.caeai.2025.100363>
- Chiluisa-Aimara, M. S., Castro-Campos, N. G., Garzón-Paredes, E. A., y Núñez-Gordon, B. I. (2025). Inteligencia artificial en la evaluación formativa de la educación básica: Estrategias para mejorar el progreso académico de los estudiantes. *Polo del Conocimiento*, 10(4), 680–697. <https://doi.org/10.23857/pc.v10i4.9340>

- Crompton, H., y Burke, D. (2024). The Educational Affordances and Challenges of ChatGPT: State of the Field. *TechTrends*, 68(2), 380–392. <https://doi.org/10.1007/s11528-024-00939-0>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., y Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474. <https://doi.org/10.1080/14703297.2023.2195846>
- Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., y McGrane, J. A. (2023). Challenges and opportunities for classroom-based formative assessment and AI: A perspective article. *Frontiers in Education*, 8. <https://doi.org/10.3389/educ.2023.1270700>
- Kasneji, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneji, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Lin, X., y Tan, H. (2025). A Systematic Review of Generative AI in K–12: Mapping Goals, Activities, Roles, and Outcomes via the 3P Model. *Systems*, 13(10). <https://doi.org/10.3390/systems13100840>
- Luckin, R., Cukurova, M., Kent, C., y du Boulay, B. (2022). Empowering educators to be AI-ready. *Computers and Education: Artificial Intelligence*, 3, 100076. <https://doi.org/10.1016/j.caeai.2022.100076>
- Maksimchuk, M., Roeber, E., y Store, D. (2025). Generative AI in the K–12 Formative Assessment Process: Enhancing Feedback in the Classroom. En J. Wilson, C. Ormerod, y M. Beiting Parrish (Eds.), *Proceedings of the Artificial Intelligence in Measurement and Education Conference (AIME-Con): Full Papers* (pp. 107–110). National Council on Measurement in Education (NCME). <https://aclanthology.org/2025.aimecon-main.12/>
- Martin, F., Zhuang, M., y Schaefer, D. (2024). Systematic review of research on artificial intelligence in K-12 education (2017–2022). *Computers and Education: Artificial Intelligence*, 6, 100195. <https://doi.org/10.1016/j.caeai.2023.100195>
- Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., y Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, 100199. <https://doi.org/10.1016/j.caeai.2023.100199>
- 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., ... Alonso-Fernández, S. (2021). Declaración PRISMA 2020: Una guía actualizada para la publicación de revisiones sistemáticas. *Revista Española de Cardiología*, 74(9), 790–799. <https://doi.org/10.1016/j.recesp.2021.06.016>

- Proaño-Sánchez, A. Y., Ambuludi-Abrigo, M. G., Román-Laínez, F. R., Córdova-Vivanco, M. J., y Guaycha-Maza, A. R. (2025). Inteligencia Artificial Generativa en la Educación Básica: Oportunidades y Consideraciones para el Diseño de Escenarios Interdisciplinarios. *Estudios y Perspectivas Revista Científica y Académica*, 5(2), 1563–1580. <https://doi.org/10.61384/r.c.a...v5i2.1226>
- Sortwell, A., Trimble, K., Ferraz, R., Geelan, D. R., Hine, G., Ramirez-Campillo, R., Carter-Thuiller, B., Gkintoni, E., y Xuan, Q. (2024). A Systematic Review of Meta-Analyses on the Impact of Formative Assessment on K-12 Students' Learning: Toward Sustainable Quality Education. *Sustainability*, 16(17). <https://doi.org/10.3390/su16177826>
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., y Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91(101894). <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., y Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075. <https://doi.org/10.1016/j.caeai.2022.100075>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., y Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Trujillo-Casadiego, H. E. (2024). Inteligencia Artificial en el aula: Transformando la evaluación formativa con ChatGPT. *Revista Universitaria de Informática RUNIN*, (18), 16–21.
- Xuan, Q., Cheung, A., y Sun, D. (2022). The effectiveness of formative assessment for enhancing reading achievement in K-12 classrooms: A meta-analysis. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.990196>
- Yao, Y., Amos, M., Snider, K., y Brown, T. (2024). The Impact of Formative Assessment on K-12 Learning: A Meta-Analysis. *Educational Research and Evaluation*, 29, 452–475. <https://doi.org/10.1080/13803611.2024.2363831>
- Zapata-Ros, M. (2024). IA generativa y ChatGPT en Educación: Un reto para la evaluación y ¿una nueva pedagogía? *Revista Paraguaya de Educación a Distancia (REPED)*, 5(1), 12–44.
- Zhang, T., Lai, Y. C., y Yu, P. L. H. (2026). Generative artificial intelligence in K-12 education: A systematic review. *Research and Practice in Technology Enhanced Learning*, 21, 034–034. <https://doi.org/10.58459/rptel.2026.21034>

CONFLICTO DE INTERESES

“Los autores declaran no tener ningún conflicto de intereses”.