Sage Sphere in Artificial Intelligence

ISSN: 3103-1307

Vol. 4 Núm. 1 (2026)

SSAI-Vol.4. N1. 010

Evaluación del pensamiento crítico antes y después de aplicar una estrategia pedagógica con inteligencia artificial en primaria

Assessment of critical thinking before and after implementing a pedagogical strategy using artificial intelligence in primary education

Autores:
Elizabeth Virginia Filipps Díaz
Universidad Mayor, Real y Pontificia de San Francisco Xavier de Chuquisaca, Sucre, Bolivia.
elyfilipps@gmail.com
https://orcid.org/0009-0005-4270-2179

Autor de correspondencia: *Elizabeth Virginia Filipps Díaz*, elyfilipps@gmail.com

Enviado: 25-05-2026 **Revisado:** 30-09-2026
Aceptado: 05-11-2026 **Publicado:** 06-11-2026

Cómo citar este artículo:
Filipps Díaz, E. V. (2026). Assessment of critical thinking before and after implementing a pedagogical strategy using artificial intelligence in primary education. *Sage Sphere in Artificial Intelligence*, *4*(1), 1-26. https://doi.org/10.63688/hpnnrf37

© 2026; Los autores. Este es un artículo en acceso abierto, distribuido bajo los términos de una licencia Creative Commons (https://creativecommons.org/licenses/by/4.0) que permite el uso, distribución y reproducción en cualquier medio, siempre que la obra original sea correctamente citada.

ISSN: 3103-1307

Vol. 4 Núm. 1 (2026)

RESUMEN

Este artículo analiza los cambios en el pensamiento crítico de estudiantes de quinto y sexto de primaria de la Unidad Educativa Guido Villagómez Loma (Potosí, Bolivia), tras la implementación de una estrategia pedagógica mediada por inteligencia artificial durante 2025. La investigación adoptó un enfoque mixto y un diseño cuasiexperimental con dos grupos intactos: experimental (22 estudiantes) y control (21 estudiantes). La intervención se desarrolló mediante cinco fases: alfabetización digital crítica, exploración guiada, trabajo individual y colaborativo, debate y transferencia. Se aplicaron una preprueba y una posprueba para evaluar las dimensiones de análisis, evaluación, explicación y autorregulación, con un máximo de 80 puntos. Estos datos se complementaron con observaciones de aula y producciones escolares. La puntuación media del grupo experimental pasó de 34,14 a 52,23 puntos, mientras que la del control aumentó de 34,29 a 37,48; la diferencia entre las ganancias medias fue de 14,90 puntos a favor del grupo experimental. Los registros de aula mostraron mayor contraste de información, justificación de respuestas y revisión de errores. Se concluye que la estrategia se asoció con avances en el pensamiento crítico, aunque la ausencia de aleatorización y el uso de una sola aula por condición impiden atribuir estos resultados exclusivamente a la inteligencia artificial.

Palabras clave: pensamiento crítico; inteligencia artificial; educación primaria; estrategia pedagógica; evaluación cuasiexperimental.

ABSTRACT

This article analyzes changes in the critical thinking of fifth- and sixth-grade primary students at the Guido Villagómez Loma Educational Unit (Potosí, Bolivia) following the implementation of an artificial intelligence-mediated pedagogical strategy in 2025. The research employed a mixed-methods approach and a quasi-experimental design involving two intact groups: an experimental group (22 students) and a control group (21 students). The intervention was carried out in five phases: critical digital literacy, guided exploration, individual and collaborative work, debate, and transfer. A pre-test and a post-test were administered to assess the dimensions of analysis, evaluation, explanation, and self-regulation, with a maximum score of 80 points. These data were complemented by classroom observations and student work. The mean score of the experimental group rose from 34.14 to 52.23 points, while that of the control group increased from 34.29 to 37.48; the difference in mean gains was 14.90 points in favor of the experimental group. Classroom records showed greater contrast of information, justification of responses, and error review. It is concluded that the strategy was associated with improvements in critical thinking, although the lack of randomization and the use of only one classroom per condition prevent attributing these results exclusively to artificial intelligence.

Keywords: critical thinking; artificial intelligence; primary education; pedagogical strategy; quasi-experimental evaluation.

Doi: https://doi.org/10.63688/hpnnrf37

ISSN: 3103-1307
Vol. 4 Núm. 1 (2026)

1. INTRODUCCIÓN

La incorporación de inteligencia artificial a las actividades escolares planteó un problema pedagógico concreto: cómo aprovechar sus respuestas para ampliar el aprendizaje sin reducir las oportunidades de razonamiento del alumnado. En primaria, la posibilidad de recibir explicaciones inmediatas tuvo que examinarse junto con la capacidad de comprenderlas, contrastarlas y justificar su aceptación. Walter (2024) vinculó la alfabetización en inteligencia artificial, la formulación de instrucciones y el pensamiento crítico con una integración educativa orientada por objetivos explícitos. Su planteamiento aportó criterios conceptuales para analizar el uso de estas herramientas, aunque procedió principalmente de educación superior y no constituyó evidencia experimental sobre niños de primaria.

La perspectiva de Benvenuti et al. (2023) situó la adquisición de nuevas competencias en la relación entre inteligencia artificial y desarrollo humano, con atención al pensamiento crítico, la creatividad y la resolución de problemas. Este enfoque permitió considerar la tecnología dentro de una experiencia formativa y no como garantía de aprendizaje. En el ámbito escolar, Kim et al. (2022) recogieron las perspectivas de diez docentes sobre el diseño de la colaboración entre estudiantes e inteligencia artificial y señalaron la importancia de las tareas auténticas, la evaluación del proceso y el análisis de errores. Ambos trabajos sustentaron la necesidad de organizar las interacciones alrededor de exigencias cognitivas reconocibles.

La revisión de Yim y Su (2025) identificó herramientas y estrategias de alfabetización en inteligencia artificial desde educación infantil hasta secundaria, entre ellas proyectos, actividades lúdicas y colaboración entre personas y sistemas. Sus resultados mostraron diversidad de recursos, objetivos y formas de evaluación. Esta heterogeneidad impidió equiparar el aprendizaje de conceptos de inteligencia artificial con el desarrollo del pensamiento crítico: comprender el funcionamiento básico de una herramienta, programar una secuencia y evaluar la consistencia de una afirmación correspondieron a desempeños relacionados, pero distintos. Por ello, el estudio delimitó su resultado principal en habilidades de análisis, evaluación, explicación y autorregulación.

La evidencia directamente vinculada con primaria aportó antecedentes pertinentes, aunque con objetos de medición diferentes. Perdana et al. (2025) desarrollaron un módulo de indagación con inteligencia artificial para mejorar la alfabetización científica en cuarto grado;

Doi: https://doi.org/10.63688/hpnnrf37

su resultado principal no fue una medida idéntica de pensamiento crítico. Yulianti y Herpratiwi (2024) evaluaron un módulo basado en ciencia, ambiente, tecnología y sociedad que favoreció esta capacidad sin aislar un componente de inteligencia artificial. La comparación entre ambos trabajos indicó que las mejoras escolares pudieron depender tanto del recurso tecnológico como de la indagación y la pertinencia de las tareas. Esta distinción orientó la interpretación de la intervención realizada en Potosí.

En un campo próximo, Benavides y Zambrano-Ramírez (2023) estudiaron la comprensión lectora autorregulada apoyada en tecnología en educación básica. Su investigación ofreció un antecedente para examinar la relación entre selección de tareas, autoevaluación y desempeño lector, sin demostrar por sí misma un efecto de la inteligencia artificial generativa. La revisión de Xu y Ouyang (2022) también mostró que la aplicación de inteligencia artificial en educación STEM involucró técnicas, entornos y sujetos diversos. Estos antecedentes hicieron necesario definir qué se intervino y qué se evaluó, evitando reunir bajo una misma categoría aplicaciones educativas con funciones diferentes.

La literatura sobre inteligencia artificial generativa aportó resultados favorables y advertencias que debieron interpretarse según el diseño de cada estudio. Deng et al. (2025) sintetizaron investigaciones experimentales sobre ChatGPT y encontraron mejoras en desempeño académico y disposiciones de pensamiento de orden superior, entre otros resultados. Li et al. (2026) revisaron estudios universitarios y relacionaron los beneficios cognitivos con diseños de indagación y apoyos instruccionales explícitos. Estas síntesis respaldaron la plausibilidad de una intervención guiada, pero no permitieron trasladar directamente sus magnitudes a una muestra escolar boliviana ni identificar como equivalentes rendimiento académico, pensamiento creativo y pensamiento crítico.

Los riesgos de dependencia tampoco se redujeron a una valoración general de la tecnología. Zhai et al. (2024) examinaron las consecuencias de la confianza excesiva en sistemas de diálogo sobre capacidades cognitivas; José et al. (2025), en un artículo de opinión, discutieron la tensión entre apoyo y sustitución del esfuerzo intelectual. Flores González et al. (2025) analizaron teóricamente oportunidades y riesgos, incluida la delegación del juicio. La diferencia entre revisión sistemática y reflexión conceptual se mantuvo explícita, porque cada tipo de fuente aportó un respaldo distinto. En conjunto, estos trabajos justificaron

actividades en las que el estudiante tuvo que verificar información y elaborar una respuesta propia.

La discusión sobre mediación docente reforzó esa orientación, Chiappe et al. (2025) revisaron la relación entre inteligencia artificial generativa y prácticas del profesorado, mientras Bolaños Gómez et al. (2026) concluyeron, para educación superior, que la relación con el pensamiento crítico dependió del encuadre pedagógico. En un estudio cualitativo preuniversitario, Sánchez Muñoz et al. (2025) describieron cómo la formulación y revisión de instrucciones coexistió con dificultades de vigilancia epistémica y reflexión crítica. Estos resultados permitieron formular una pregunta compartida: qué actividades conservaron la responsabilidad del estudiante sobre la interpretación y la conclusión, aun cuando recibió asistencia automatizada.

La dimensión ética se vinculó con decisiones de enseñanza y organización, Chan (2023) propuso un marco universitario que articuló pedagogía, gobernanza y condiciones operativas; Chan y Hu (2023) documentaron percepciones estudiantiles sobre beneficios, exactitud y privacidad. Moya y Eaton (2023) examinaron recomendaciones de integridad desde distintos niveles institucionales y advirtieron vacíos en su concreción pedagógica. Estos antecedentes no se emplearon como prueba de eficacia en primaria, sino como apoyo para analizar reglas de uso, responsabilidades y protección de información. En la experiencia estudiada, los acuerdos de aula y la supervisión de las respuestas formaron parte de la estrategia.

La investigación se desarrolló en la Unidad Educativa Guido Villagómez Loma de Potosí, durante la gestión 2025, en condiciones de conectividad intermitente y equipamiento heterogéneo. La intervención previó respuestas impresas y actividades presenciales para mantener el trabajo cuando la conexión falló. La justificación se apoyó en una necesidad de aprendizaje: los registros iniciales mostraron puntuaciones especialmente bajas en autorregulación y explicación. Se buscó que el alumnado dispusiera de oportunidades sostenidas para reconocer ideas, evaluar afirmaciones, expresar razones y revisar su razonamiento, mediante tareas articuladas con los contenidos escolares.

El pensamiento crítico se definió operativamente a través de cuatro dimensiones. El proceso de análisis comprendió la identificación de ideas y relaciones; la evaluación se expresó en el examen de afirmaciones y la distinción entre hechos y opiniones; la explicación exigió razones comprensibles y pertinentes; la autorregulación implicó reconocer

dificultades y plantear revisiones. Esta delimitación correspondió al instrumento empleado y no pretendió abarcar todas las concepciones de pensamiento crítico. La evaluación por dimensiones permitió examinar si el avance fue uniforme y evitó interpretar cualquier mejora en el puntaje total como dominio completo de la competencia.

La variable independiente fue la participación en una estrategia pedagógica con inteligencia artificial, organizada en alfabetización digital crítica, exploración guiada, trabajo individual y colaborativo, debate y transferencia. La variable dependiente fue el desempeño en pensamiento crítico, medido antes y después de la intervención. El estudio adoptó un enfoque mixto y un diseño cuasiexperimental longitudinal con dos grupos intactos, complementado por observación de aula y revisión de producciones. El objetivo general fue evaluar los cambios del pensamiento crítico tras aplicar la estrategia en estudiantes de quinto y sexto de primaria de la institución, comparando las variaciones del grupo experimental con las del grupo de control durante la gestión escolar 2025.

2. METODOLOGÍA

La investigación integró un componente cuantitativo de comparación de desempeños y un componente cualitativo de descripción de la implementación. Las pruebas aportaron puntuaciones de cuatro dimensiones, mientras la guía de observación, las producciones escritas y el diario de campo permitieron contextualizar la interacción con las herramientas. El enfoque mixto se concretó en la interpretación conjunta de ambas fuentes; no se atribuyó a la información cualitativa una representatividad estadística ni se trató la coincidencia entre registros como demostración de un mecanismo causal.

El estudio fue evaluativo y explicativo en su propósito y cuasiexperimental en su diseño, el trabajo se realizó con dos cursos previamente constituidos, uno en la condición experimental y otro en la condición de control, sin asignación aleatoria de estudiantes. Ambos fueron evaluados antes y después de la intervención. El control continuó con la programación curricular habitual, mientras el experimental participó en la estrategia. La comparación permitió describir evoluciones diferentes bajo el mismo calendario escolar, aunque no eliminó las posibles diferencias entre cursos, docentes o experiencias externas.

El contexto fue la Unidad Educativa Guido Villagómez Loma, en la ciudad de Potosí, Bolivia. La población de interés correspondió a estudiantes de quinto y sexto de primaria de

la institución. La muestra incluyó 43 participantes: 22 en el experimental y 21 en el control. La selección de cursos disponibles se consideró no probabilística; no se documentó un cálculo prospectivo del tamaño muestral ni el número total de estudiantes elegibles. Tampoco se especificó de manera inequívoca qué grado correspondió a cada condición, por lo que no se supuso equivalencia curricular entre grupos únicamente a partir de sus puntajes iniciales.

La participación se basó en autorización institucional, consentimiento informado escrito de madres, padres o tutores y asentimiento de los estudiantes. La tesis estableció como criterios la escolaridad regular en la institución, la autorización de participación y la ausencia de circunstancias que interfirieran significativamente con las actividades. Los anexos utilizaron códigos E01–E22 y C01–C21, que permitieron emparejar las mediciones sin identificar a los menores. Se encontraron registros completos para los mismos 43 códigos en ambos momentos; esta coincidencia no permitió reconstruir asistencia por sesión ni eventuales incidencias no consignadas.

La variable de intervención comprendió una secuencia pedagógica de cinco fases y la variable de resultado incluyó análisis, evaluación, explicación y autorregulación. Cada dimensión tuvo un máximo de 20 puntos y el total, de 80. El instrumento presentó textos breves y situaciones relacionadas con comprensión de fenómenos, evaluación de afirmaciones, justificación de opiniones y revisión del propio aprendizaje. Las preguntas exigieron respuestas con palabras propias y razones, de modo que la evaluación se centró en el desempeño argumentativo y reflexivo, no en una encuesta de percepción sobre el uso de tecnología.

La preprueba se aplicó en marzo de 2025 y la posprueba en el mes de noviembre, con un tiempo estimado de 60 minutos y condiciones comparables para ambos grupos. El anexo del instrumento describió versiones con contenidos equivalentes y reordenados. Se utilizó una rúbrica para la calificación; sin embargo, su presentación por niveles de 0 a 5 puntos no explicitó completamente la transformación a los 20 puntos por dimensión, especialmente en explicación. Por ello, el análisis utilizó las puntuaciones ya consignadas y no reconstruyó una regla de calificación ausente. Esta limitación afectó la replicabilidad exacta de la medición.

La validez de contenido se sustentó en revisión por expertos de pedagogía y pensamiento crítico y en una aplicación piloto dirigida a ajustar claridad y comprensión de las consignas. No se informaron el número de jueces, índices de acuerdo, coeficientes de

consistencia interna, estabilidad temporal o concordancia entre calificadores. La equivalencia declarada de las versiones tampoco estuvo acompañada de un estudio psicométrico. En consecuencia, se distinguió el respaldo a la pertinencia del contenido de la evidencia de confiabilidad, y no se atribuyó al instrumento un coeficiente que la documentación no reportó.

La primera fase de la estrategia se desarrolló durante tres semanas y abordó nociones accesibles de inteligencia artificial y acuerdos de uso seguro, ético y reflexivo. La segunda comprendió cuatro semanas y ocho sesiones de modelado docente sobre preguntas, contraste de respuestas y verificación. La tercera se extendió diez semanas, con veinte sesiones de trabajo individual y en pares sobre tareas curriculares. La cuarta ocupó cuatro semanas e incluyó cuatro debates colectivos y rúbricas sencillas de autoevaluación. La quinta duró dos semanas y culminó en carteles sobre el cuestionamiento de información de internet y su aplicación a situaciones cotidianas.

Las cinco fases sumaron 23 semanas planificadas dentro del periodo escolar comprendido entre las mediciones. La documentación no detalló todas las fechas, la duración de cada sesión ni un registro individual de exposición; por ello, las 23 semanas no se equipararon a una intervención continua de marzo a noviembre. Las herramientas se describieron como asistentes de diálogo, recursos adaptativos e interactivos, seleccionados por accesibilidad y bajos requerimientos. No se identificaron de forma completa productos, versiones, configuraciones o instrucciones de uso, de modo que el procedimiento pudo reconstruirse en su secuencia pedagógica, pero no reproducirse exactamente en su componente tecnológico.

La docente supervisó y complementó las respuestas de los sistemas, vinculó las actividades con contenidos escolares y promovió preguntas de verificación. Ante fallas de conexión se utilizaron materiales impresos y debates presenciales preparados con antelación. En el proceso hubo algunas interrupciones o demoras en aproximadamente el 15 % de las sesiones, sin presentar una matriz que permitiera comprobar ese porcentaje. Se mantuvo como descripción contextual y no como una medida exacta de exposición a la inteligencia artificial. Las actividades sin conexión formaron parte de la adaptación de la estrategia y no constituyeron una condición experimental separada.

La recolección cualitativa se apoyó en una guía con indicadores de preguntas precisas, contraste con fuentes, solicitud de justificación, detección de contradicciones,

expresión de razones y reconocimiento de errores. Se revisaron los episodios de implementación descritos en el diario y las producciones escolares. La sistematización se organizó por dimensiones y por momentos de la intervención. No se documentaron un corpus completo de transcripciones, codificación independiente o acuerdo entre observadores; los resultados cualitativos se presentaron como evidencia descriptiva del proceso y no como frecuencias de comportamiento de todos los participantes.

Para el análisis descriptivo, se aplicó pruebas t para muestras relacionadas e independientes y tamaños del efecto. Antes de incorporar estos valores al análisis, se cotejaron sus resúmenes con los registros individuales de los anexos 1 y 2. Se verificaron 43 pares de observaciones y la igualdad entre la suma de dimensiones y el total de cada registro. Se encontraron diferencias en medias, dispersión y resultados inferenciales frente al capítulo de resultados y al anexo 3. Para evitar combinar series incompatibles, todas las tablas cuantitativas del artículo se recalcularon desde los anexos individuales, sin modificar sus puntuaciones.

El análisis principal presentó medias, desviaciones estándar muestrales, cambios absolutos y diferencias entre ganancias. Como análisis exploratorio, se compararon las puntuaciones totales iniciales, finales y sus cambios mediante t de Welch, con significación bilateral de 0,05, y se calculó un intervalo de confianza del 95 % para la diferencia entre ganancias. Se prefirió este procedimiento porque no exigió igualdad de varianzas. Sus resultados se interpretaron bajo el supuesto de independencia entre estudiantes, que pudo incumplirse al compartir aula; al existir una sola clase por condición, no fue posible separar el efecto de la estrategia del efecto del curso ni estimar de forma fiable la dependencia intraclase.

Los niveles de logro se recalcularon con los umbrales establecidos en la tesis: inicial, hasta 40 %; en desarrollo, de 41 % a 70 %; consolidado, de 71 % a 100 %. Para las puntuaciones enteras observadas de 0 a 20, estos intervalos correspondieron a 0–8, 9–14 y 15–20 puntos. La clasificación se realizó por dimensión, evitando confundir un porcentaje relativo de incremento con el porcentaje del máximo de la escala. La integración final contrastó la dirección de los cambios con los registros de verificación, debate y reflexión, manteniendo explícita la incertidumbre derivada de la consistencia documental.

3. RESULTADOS

Los resultados se organizaron en el diagnóstico inicial, la ejecución de la estrategia, los cambios de cada grupo y la comparación final. La intervención y las dos mediciones estuvieron documentadas; la principal dificultad surgió en la correspondencia entre datos individuales y estadísticas resumidas. La tabla 1 presentó esta diferencia para el puntaje total. Los cálculos posteriores procedieron exclusivamente de los anexos 1 y 2, cuya transcripción se comprobó contra las páginas del PDF; no se mezclaron sus valores con los del anexo estadístico.

Tabla 1

Contraste del puntaje total resumido y recalculado

Indicador	Resumen de tesis	Anexos individuales
Experimental preprueba	34,77	34,14
Experimental posprueba	53,77	52,23
Control preprueba	35,05	34,29
Control posprueba	37,24	37,48
Ganancia experimental	19,00	18,09
Ganancia del control	2,19	3,19
Diferencia final	16,53	14,75

Nota. Elaboración propia con base a la aplicación de la Estrategia pedagógica de aplicación de inteligencia artificial para el desarrollo del pensamiento crítico en educación primaria.

La discrepancia superó una diferencia de redondeo. El resumen de la tesis indicó un aumento experimental de 19,00 puntos y un incremento de 2,19 en el control, mientras los registros individuales produjeron cambios de 18,09 y 3,19, respectivamente. La diferencia final entre grupos también varió: 16,53 puntos en el resumen y 14,75 en los anexos. La dirección de la ventaja experimental se mantuvo, pero su magnitud y la interpretación de la evolución del control requirieron corrección. La documentación disponible no permitió determinar cuál serie reprodujo las hojas originales de calificación.

Tabla 2

Desempeño inicial por dimensión y grupo

Dimensión	Experimental M (DE)	Control M (DE)
Análisis	9,23 (1,27)	9,19 (1,29)
Evaluación	8,64 (1,14)	8,71 (1,10)
Explicación	8,32 (0,99)	8,33 (1,06)

Dimensión	Experimental M (DE)	Control M (DE)
Autorregulación	7,95 (1,05)	8,05 (0,86)
Total	34,14 (4,28)	34,29 (4,19)

Nota. Experimental n = 22; control n = 21. M = media; DE = desviación estándar muestral. Máximo por dimensión: 20; total: 80. Elaboración propia con base a la aplicación de la Estrategia pedagógica de aplicación de inteligencia artificial para el desarrollo del pensamiento crítico en educación primaria.

Las medias iniciales recalculadas fueron próximas: 34,14 puntos en el experimental y 34,29 en el control, equivalentes a 42,67 % y 42,86 % del máximo de la prueba. La diferencia total fue de −0,15 puntos. En ambos grupos, análisis obtuvo la media dimensional más alta y autorregulación la más baja. La comparación exploratoria del total mediante Welch no detectó una diferencia inicial, t (40,97) = −0,12, p = 0,908. Este resultado describió compatibilidad de las puntuaciones observadas, sin demostrar equivalencia de todas las características escolares o personales.

Tabla 3

Niveles iniciales del grupo experimental

Dimensión	Inicial n (%)	En desarrollo n (%)	Consolidado n (%)
Análisis	6 (27,3)	16 (72,7)	0 (0,0)
Evaluación	10 (45,5)	12 (54,5)	0 (0,0)
Explicación	11 (50,0)	11 (50,0)	0 (0,0)
Autorregulación	15 (68,2)	7 (31,8)	0 (0,0)

Nota. n = 22. Se aplicaron los umbrales de la tesis a los datos individuales: inicial 0–8, en desarrollo 9–14 y consolidado 15–20 puntos. Elaboración propia con base a la aplicación de la Estrategia pedagógica de aplicación de inteligencia artificial para el desarrollo del pensamiento crítico en educación primaria.

La reclasificación del experimental mostró mayor proporción inicial en autorregulación, con 15 de 22 estudiantes, seguida de explicación, con 11. Ningún registro alcanzó 15 puntos o más en alguna dimensión, por lo que no se reprodujo el 4,5 % de nivel consolidado informado en las tablas originales. El diagnóstico apoyó la necesidad de fortalecer revisión y justificación, pero la distribución precisa dependió de aplicar los umbrales a los registros individuales. La diferencia fue relevante para evitar presentar una progresión de niveles construida sobre porcentajes incompatibles con las puntuaciones.

Tabla 4

Fases ejecutadas durante la gestión escolar 2025

Fase	Duración	Evidencia documentada
Inducción y alfabetización digital crítica	3 semanas	Siete acuerdos de uso seguro, ético y reflexivo.
Exploración guiada y modelado	4 semanas	Ocho sesiones con preguntas, contraste y verificación.
Trabajo individual y colaborativo	10 semanas	Veinte sesiones individuales y en pares; alternativas impresas.
Socialización, debate y metacognición	4 semanas	Cuatro debates y autoevaluación acompañada.
Cierre y transferencia	2 semanas	Carteles sobre cuestionamiento de información de internet.

Nota. Elaboración propia con base a la aplicación de la Estrategia pedagógica de aplicación de inteligencia artificial para el desarrollo del pensamiento crítico en educación primaria.

Las evidencias de implementación mostraron que se ejecutaron actividades diferenciadas durante las cinco fases, con acuerdos expuestos en el aula, sesiones modeladas, trabajo individual y en pares, debates y carteles finales. El registro sustentó la existencia de una intervención aplicada, no únicamente el diseño de una propuesta. No permitió, sin embargo, calcular dosis individuales ni atribuir a una fase determinada el cambio de una dimensión. La tabla sintetizó la secuencia y los productos descritos, sin convertir su cumplimiento declarado en una medida independiente de calidad o fidelidad.

Tabla 5

Cambios del grupo experimental

Dimensión	Pre M (DE)	Pos M (DE)	Cambio
Análisis	9,23 (1,27)	13,82 (1,74)	+4,59
Evaluación	8,64 (1,14)	13,18 (1,76)	+4,55
Explicación	8,32 (0,99)	12,82 (1,68)	+4,50
Autorregulación	7,95 (1,05)	12,41 (1,65)	+4,45
Total	34,14 (4,28)	52,23 (6,74)	+18,09

Nota. n = 22. Cambio = posprueba − preprueba, calculado antes de redondear. Elaboración propia con base a la aplicación de la Estrategia pedagógica de aplicación de inteligencia artificial para el desarrollo del pensamiento crítico en educación primaria.

La media experimental total aumentó 18,09 puntos, de 34,14 a 52,23, mientras la mediana pasó de 34 a 54. El rango inicial fue de 27 a 43 y el final, de 40 a 63 puntos. Las ganancias dimensionales fueron cercanas: 4,59 en análisis, 4,55 en evaluación, 4,50 en explicación y 4,45 en autorregulación. La proximidad de estos cambios mostró un avance distribuido, sin diferencias suficientes para sostener que una dimensión respondió de manera claramente superior. Autorregulación conservó la media final más baja, pese a su mejora absoluta.

Tabla 6

Cambios del grupo de control

Dimensión	Pre M (DE)	Pos M (DE)	Cambio
Análisis	9,19 (1,29)	10,14 (1,35)	+0,95
Evaluación	8,71 (1,10)	9,71 (1,10)	+1,00
Explicación	8,33 (1,06)	8,90 (1,09)	+0,57
Autorregulación	8,05 (0,86)	8,71 (1,10)	+0,67
Total	34,29 (4,19)	37,48 (4,56)	+3,19

Nota. n = 21. Cambio = posprueba − preprueba, calculado antes de redondear. Elaboración propia con base a la aplicación de la Estrategia pedagógica de aplicación de inteligencia artificial para el desarrollo del pensamiento crítico en educación primaria.

El control también mejoró. Su media total aumentó 3,19 puntos y la mediana pasó de 34 a 38, con rangos de 28–43 en la preprueba y 30–47 en la posprueba. El mayor incremento dimensional se produjo en evaluación, con 1,00 punto, seguido de análisis, con 0,95; explicación y autorregulación aumentaron 0,57 y 0,67. Estos registros no sustentaron la ausencia de avances en el control descrita en los resúmenes estadísticos de la tesis. La comparación pertinente se centró en la mayor magnitud de la ganancia experimental y no en suponer un control inmóvil.

Tabla 7

Comparaciones exploratorias del puntaje total

Comparación	Diferencia E − C	t de Welch	gl	p
Preprueba	−0,15	−0,12	40,97	0,908
Posprueba	14,75	8,44	37,02	< 0,001
Ganancias	14,90	24,25	24,79	< 0,001

Nota. E = experimental; C = control; gl = grados de libertad. IC del 95 % para la diferencia entre ganancias: [13,63; 16,17]. Cálculos bilaterales desde los anexos 1 y 2. La inferencia

supuso independencia individual y no ajustó agrupación por aula; se presentó con alcance exploratorio.

En la posprueba, la diferencia descriptiva favoreció al experimental en todas las dimensiones. Fue de 3,68 puntos en análisis, 3,47 en evaluación, 3,91 en explicación y 3,69 en autorregulación. La separación final total fue de 14,75 puntos, mientras la diferencia entre ganancias alcanzó 14,90. El intervalo exploratorio del 95 % para esta última fue de 13,63 a 16,17 puntos, condicionado al supuesto de independencia individual. Su amplitud no incorporó la incertidumbre por asignación de aulas, medición o discrepancias entre series documentales, por lo que no se interpretó como precisión causal de la intervención.

La clasificación de la posprueba indicó que cinco de los 22 estudiantes del experimental alcanzaron nivel consolidado en al menos dos dimensiones, equivalente al 22,7 %, frente a ninguno de los 21 del control. Este resultado no reprodujo los porcentajes de 72,7 % y 4,8 % del resumen de la tesis. La media global experimental correspondió a 65,28 % del máximo y permaneció dentro del intervalo en desarrollo. Se observó una mejora del desempeño, pero no una consolidación generalizada de la competencia en toda la muestra.

Los registros cualitativos describieron una transición en el experimental. Durante las primeras semanas predominó la confianza en las respuestas automatizadas; entre las semanas 6 y 8 se consignaron preguntas más precisas, solicitudes de razones y comparación con textos escolares. Entre las semanas 12 y 15 aparecieron episodios de detección de imprecisiones, reformulación de preguntas y discusión entre pares. Estos episodios fueron compatibles con los objetivos de evaluación y explicación, aunque no permitieron cuantificar cuántos estudiantes mantuvieron esas prácticas ni demostrar su persistencia fuera de las sesiones.

La autoevaluación acompañada y los debates aportaron oportunidades para reconocer dificultades y justificar decisiones. Las alternativas impresas sostuvieron el contraste de información cuando la conexión se interrumpió. La convergencia entre los avances medidos y esos registros resultó plausible, pero no identificó un mediador estadístico ni separó el aporte de la inteligencia artificial del tiempo adicional de práctica, el apoyo docente y la interacción entre compañeros. La lectura de los hallazgos quedó condicionada a comprobar las hojas originales que dieron origen a las dos series numéricas de la tesis.

4. DISCUSIÓN

ISSN: 3103-1307
Vol. 4 Núm. 1 (2026)

El estudio mostró una evolución favorable del pensamiento crítico en el grupo experimental y una ganancia menor en el control. La ventaja descriptiva se mantuvo al recalcular los registros individuales, aunque cambiaron las medias, los niveles de logro y la interpretación del control respecto de los resúmenes de la tesis. La primera implicación fue metodológica: una conclusión sobre la intervención exigió coherencia entre puntuaciones, tablas y análisis. La coincidencia de la dirección del resultado no justificó conservar cifras incompatibles. Por ello, la discusión se apoyó en la serie reproducible de los anexos y mantuvo abierta la corroboración de su correspondencia con las pruebas originales.

La mejora experimental resultó compatible con la literatura que relacionó beneficios cognitivos y diseño pedagógico. Li et al. (2026) identificaron resultados favorables cuando ChatGPT se integró en actividades de indagación, regulación metacognitiva y argumentación. Bolaños Gómez et al. (2026) también destacaron el encuadre formativo en su revisión universitaria. La experiencia de Potosí compartió esas exigencias mediante preguntas, comparación de respuestas y elaboración propia de conclusiones. La convergencia se refirió a la orientación pedagógica; no estableció que los estudios universitarios validaran la prueba escolar ni que las magnitudes fueran directamente comparables.

Deng et al. (2025) ofrecieron un respaldo experimental agregado para determinados resultados de aprendizaje con ChatGPT, pero su síntesis reunió medidas y entornos diferentes. En el presente estudio, el resultado evaluado fue una prueba contextualizada de cuatro dimensiones y no una calificación curricular general. Las mejoras debieron interpretarse dentro de ese instrumento, cuya evidencia psicométrica fue incompleta. La comparación con el metaanálisis no autorizó calificar el efecto local como superior al de otros estudios, especialmente cuando la unidad de intervención fue un aula y las estadísticas individuales pudieron sobreestimar la precisión.

El análisis dimensional indicó ganancias experimentales muy próximas y una persistencia de diferencias en el nivel final. La autorregulación mejoró, pero continuó por debajo de análisis, mientras explicación mantuvo exigencias de organización y justificación. La intervención ofreció ejercicios para esas capacidades, aunque no midió por separado el esfuerzo cognitivo que requirió cada tarea. La interpretación que atribuyó la menor puntuación a una carga metacognitiva mayor quedó como hipótesis plausible y no como hallazgo demostrado. Tian

Doi: https://doi.org/10.63688/hpnnrf37

y Zhang (2026) aportaron evidencia transversal universitaria sobre inmersión y dependencia, pero no probaron esa diferencia entre dimensiones en niños.

La mediación docente apareció como un componente observable de la estrategia. Kim et al. (2022) situaron la colaboración con inteligencia artificial en tareas auténticas, análisis de errores y evaluación del proceso. Lee et al. (2024) diseñaron, en educación superior, un mecanismo que exigió formular ideas propias antes de consultar ChatGPT y revisar respuestas con su retroalimentación. La experiencia escolar incluyó modelado y acompañamiento, aunque no documentó que todas las sesiones siguieran exactamente esa secuencia. Ambos antecedentes ayudaron a interpretar por qué la organización de la interacción pudo importar, sin atribuir al caso procedimientos no consignados.

Los resultados de Castro Macías (2025) constituyeron un antecedente regional más próximo por su diseño cuasiexperimental mixto en clases de Física de bachillerato ecuatoriano. Ese estudio evaluó pensamiento crítico con una intervención tecnológica, pero trabajó con estudiantes de otra edad y disciplina. La similitud del diseño permitió comparar la necesidad de una condición de control y de orientación pedagógica; las diferencias impidieron extrapolar resultados a primaria. La experiencia de Potosí amplió el ámbito contextual de la discusión, aunque una sola institución no representó la diversidad educativa boliviana.

Perdana et al. (2025) aportaron un antecedente directamente escolar sobre indagación socialmente situada con inteligencia artificial, pero evaluaron alfabetización científica. Su resultado apoyó la viabilidad de combinar tareas de investigación y asistencia tecnológica en primaria. Yulianti y Herpratiwi (2024) mostraron mejoras del pensamiento crítico mediante un módulo contextualizado sin aislar inteligencia artificial. La consideración conjunta de ambos trabajos reforzó una objeción pertinente: el avance local pudo corresponder a la secuencia de indagación, argumentación y reflexión en su conjunto. El diseño de Potosí no comparó esa misma secuencia con y sin inteligencia artificial, de modo que no estimó su aporte incremental.

La lectura cualitativa favoreció una explicación basada en el contraste de información. Sánchez Muñoz et al. (2025) describieron, en estudiantes preuniversitarios, revisión de respuestas y problemas de vigilancia epistémica durante actividades con inteligencia artificial generativa. En Potosí, los episodios de comparación con textos escolares y discusión entre pares sugirieron que las respuestas automatizadas pudieron convertirse en

materiales examinables. Sin embargo, los registros no ofrecieron un seguimiento completo de cada estudiante. La inferencia se limitó a la presencia de esas prácticas durante la implementación, sin afirmar que todos adquirieran un hábito estable de verificación.

Gil-Duran y Mogas (2026) aportaron una diferencia que enriqueció la interpretación. En su estudio de formación inicial docente, una respuesta escrita más extensa o formalmente desarrollada no implicó necesariamente un cambio de interpretación ni un pensamiento más autónomo. Esta advertencia resultó relevante para valorar las producciones del alumnado: la calidad superficial de una respuesta no bastó para identificar aprendizaje. La prueba escolar solicitó razones y revisión, pero no documentó una evaluación ciega ni una separación completa entre ayuda recibida y desempeño autónomo. Por ello, los cambios en puntuación se leyeron junto con las condiciones de medición y no como evidencia suficiente de autonomía duradera.

Benavides y Zambrano-Ramírez (2023) permitieron relacionar el trabajo de comprensión con la autoevaluación apoyada en tecnología. Su objeto no fue la combinación de tareas conectadas y desconectadas, por lo que no se les atribuyó demostrar ventajas de las alternativas impresas. En el caso estudiado, estas alternativas mantuvieron disponibles las preguntas y las fuentes cuando falló internet. Su utilidad quedó respaldada por la descripción de continuidad del trabajo; no se midió si produjeron mejores resultados que las sesiones conectadas. La adaptación tecnológica tuvo un alcance práctico comprobable sin requerir una afirmación comparativa no evaluada.

La revisión de Yim y Su (2025) mostró que aprender sobre inteligencia artificial y aprender mediante ella involucró herramientas y objetivos diversos. En Potosí, la alfabetización inicial preparó el uso, mientras el resultado principal se definió en habilidades críticas. Esta diferenciación evitó interpretar la familiaridad tecnológica como equivalente de pensamiento crítico. Mitevska-Poceva (2025) propuso una integración teórica de juego e inteligencia artificial para primaria; su contribución ayudó a comprender posibilidades de motivación y participación, pero no demostró un efecto experimental de gamificación en la intervención. La tesis no documentó una condición lúdica diferenciada que permitiera atribuirle parte del cambio.

La organización del aprendizaje también se vinculó con condiciones institucionales. Oyarvide Estupiñán et al. (2024) estudiaron acceso, habilidades y apoyo para

herramientas digitales en universitarios; sus hallazgos orientaron la atención a factores de implementación y no explicaron directamente el desempeño de niños. Chiappe et al. (2025) discutieron la relación entre inteligencia artificial y prácticas docentes. En la experiencia escolar, la selección de recursos accesibles y la preparación de materiales alternativos mostraron decisiones concretas de acompañamiento. No se midieron costos, carga laboral o sostenibilidad institucional, de modo que la ejecución durante una gestión no acreditó viabilidad permanente.

La formación del profesorado apareció como una condición de posible transferencia. López Proaño y Abad Arroyo (2025) recogieron percepciones y experiencias sobre inteligencia artificial en formación docente inicial; Rodríguez Vela et al. (2026) analizaron cambios en conocimiento y usos de la herramienta entre estudiantes de educación preescolar. Estos trabajos no evaluaron aulas de niños preescolares ni comprobaron pensamiento crítico de alumnos de primaria. Su aporte consistió en situar la preparación profesional como un problema distinto del desempeño escolar. La intervención local requirió modelado, selección de información y acompañamiento, competencias cuya formación no se estudió directamente.

Los riesgos de sustitución cognitiva ofrecieron una lectura complementaria. Zhai et al. (2024) revisaron problemas de sobredependencia, mientras José et al. (2025) discutieron conceptualmente el doble potencial de apoyo y erosión. Flores González et al. (2025) vincularon el uso reflexivo con principios pedagógicos y éticos. Las actividades de cuestionamiento documentadas en Potosí resultaron compatibles con una prevención de la aceptación automática, pero no se aplicó una escala de dependencia. La ausencia de esa medición impidió concluir que la estrategia redujera un riesgo psicológico específico, aunque sí permitió describir prácticas de verificación durante las sesiones.

Ismail et al. (2026) evaluaron un programa universitario basado en la duda, con análisis de textos generados, debate y verificación, y observaron mejoras en habilidades críticas y autoeficacia. La similitud con algunas actividades locales respaldó el interés de enseñar a cuestionar afirmaciones de manera sistemática. La diferencia de edad, instrumentos y asignación de participantes impidió utilizar ese estudio como confirmación directa de la eficacia en primaria. Además, el caso de Potosí no midió autoeficacia: reconocer errores o

pedir razones en una actividad no equivalió a demostrar confianza general para verificar conocimiento generado.

Aguirre Mejía et al. (2026) identificaron, en una intervención universitaria de escritura científica, una distancia entre manejo instrumental y profundidad conceptual, junto con problemas de verificación de fuentes. Este resultado contrastó con una interpretación automáticamente favorable de toda actividad con inteligencia artificial. El estudio escolar describió cambios positivos, pero no incluyó una prueba posterior de transferencia a materiales completamente nuevos. La comparación planteó la necesidad de examinar si el estudiante sostuvo sus razones ante información contradictoria y sin apoyo de la herramienta, en lugar de valorar únicamente el producto realizado dentro de la secuencia guiada.

Elshall et al. (2026) compararon cohortes de un curso universitario de ciencia de datos ambientales con distinta disponibilidad tecnológica y observaron mejoras en proyectos auténticos sin deterioro de exámenes conceptuales. Su estudio no midió directamente pensamiento crítico. Aportó una hipótesis de alineación entre capacidades de la herramienta, tarea y pedagogía que resultó útil para interpretar la intervención local. Esta afinidad teórica no eliminó una limitación del caso: no se identificaron completamente los sistemas y sus versiones. Sin esa información, una réplica pudo conservar la secuencia didáctica, pero variar de manera sustancial la asistencia automatizada.

Yan et al. (2024) analizaron oportunidades de apoyo personalizado y retroalimentación junto con errores de los modelos y problemas de evaluación. Yan et al. (2025), en un editorial, orientaron la atención hacia cognición, metacognición y agencia epistémica más allá de la eficiencia. La experiencia de Potosí no evaluó ahorro de tiempo, por lo que su contribución se ubicó en el desempeño y las prácticas de aula. La comparación sugirió que una intervención pudo valorarse por la actividad intelectual que promovió, aunque la información disponible no permitió establecer cuánto razonamiento permaneció autónomo después del cierre.

La evaluación exigió diferenciar el resultado observable de su atribución. Xia et al. (2024) revisaron transformaciones de la evaluación con inteligencia artificial en educación superior y destacaron autorregulación, responsabilidad e integridad. En el estudio escolar, la prueba de desempeño aportó una ventaja sobre una medida exclusivamente perceptiva, pero la ausencia de confiabilidad y concordancia entre calificadores limitó su precisión. La

discrepancia entre resúmenes y anexos agregó una dificultad independiente de la validez del instrumento. Ambas limitaciones requirieron revisión: mejorar la medición no resolvió por sí mismo la trazabilidad de las cifras ya registradas.

Chan (2023) y Moya y Eaton (2023) aportaron marcos de responsabilidad e integridad cuya adaptación escolar exigió considerar menores y mediación adulta. Los siete acuerdos de aula y el consentimiento estuvieron documentados, pero no se especificaron por completo condiciones de cuentas, tratamiento de datos o restricciones de las plataformas. La referencia a principios éticos no sustituyó esa información técnica. Chan y Hu (2023) documentaron preocupaciones universitarias por privacidad y exactitud; el estudio local no midió esas percepciones en niños, de modo que se evitó presentarlas como hallazgos propios.

La ganancia del control modificó la interpretación sustantiva. Al recalcular los anexos, el aprendizaje ordinario también presentó avances, aunque de menor magnitud. Esto fue compatible con maduración, práctica de la prueba o enseñanza curricular, pero el diseño no permitió identificar cuánto aportó cada factor. La ausencia de diferencias iniciales detectables no garantizó paralelismo de trayectorias sin intervención. Con una sola aula por condición, grado, docente y dinámica de curso pudieron confundirse con el tratamiento. Los valores inferenciales individuales describieron los registros bajo supuestos restrictivos y no resolvieron esas amenazas al diseño.

La regularidad de las ganancias también requirió cautela. En el control, todos los registros aumentaron exactamente un punto en evaluación, y los cambios totales de ambos grupos mostraron una dispersión reducida. Este patrón pudo corresponder a la calificación, a la construcción de las tablas o a otro proceso no documentado; no permitió inferir por sí mismo una irregularidad intencional. Sí justificó revisar las hojas originales y las reglas de puntuación antes de extraer conclusiones inferenciales definitivas. La comprobación documental se consideró necesaria porque los datos agregados y los individuales produjeron magnitudes diferentes.

El alcance teórico del estudio se situó en una interpretación de la inteligencia artificial como recurso dentro de una mediación pedagógica más amplia. Las tareas de contraste, la explicación a compañeros y la revisión acompañada ofrecieron mecanismos plausibles para el avance. No se midieron mediadores ni se compararon componentes aislados; por tanto, la explicación permaneció como interpretación apoyada por registros y literatura. El

alcance práctico consistió en una secuencia aplicada y adaptable, con alternativas ante conectividad irregular, cuya reproducción exigió completar detalles tecnológicos, temporales y de evaluación.

La experiencia permitió plantear una contribución contextual delimitada: se documentó una intervención escolar aplicada y una ventaja descriptiva del grupo que participó en ella. No se demostró eficacia general para primaria, superioridad de una plataforma ni consolidación del pensamiento crítico en la mayoría. Las diferencias documentales, el tamaño de la muestra y la ausencia de seguimiento limitaron la conclusión. El resultado más defendible fue una asociación favorable que justificó investigación adicional y revisión de las fuentes primarias de calificación, antes de orientar decisiones de expansión institucional.

5. CONCLUSIÓN

La evaluación de los registros individuales mostró que el grupo experimental aumentó su media total de 34,14 a 52,23 puntos, con una ganancia de 18,09. Las cuatro dimensiones avanzaron de manera próxima, aunque autorregulación conservó la media final más baja. La puntuación global permaneció en el intervalo en desarrollo, por lo que se concluyó una mejora del desempeño y no una consolidación generalizada del pensamiento crítico.

El control también mejoró, de 34,29 a 37,48 puntos, pero su ganancia de 3,19 fue inferior a la experimental. La diferencia entre ganancias alcanzó 14,90 puntos y estuvo acompañada por registros de mayor contraste y justificación durante la intervención. Se concluyó una asociación favorable con la estrategia completa; la falta de aleatorización, una sola aula por condición y la discrepancia entre anexos y resúmenes impidieron atribuir causalidad exclusiva a la inteligencia artificial.

Se recomendó contrastar las hojas originales de calificación y replicar el estudio con más aulas, documentación de exposición y análisis psicométricos del instrumento. Como línea futura, se propuso comparar una misma secuencia pedagógica con y sin inteligencia artificial, incorporar calificación independiente y realizar seguimiento posterior. Estas decisiones permitirían examinar el aporte incremental de la herramienta y la persistencia de las habilidades observadas.

Se planteó ampliar la investigación a instituciones urbanas y rurales con diferentes condiciones de conectividad, lengua de enseñanza y preparación docente. La experiencia de

Potosí aportó una secuencia contextualizada, pero no acreditó su transferencia automática. Los estudios comparativos deberían conservar los objetivos cognitivos y documentar las adaptaciones, para distinguir diferencias de implementación de diferencias en la respuesta del alumnado.

Se identificó como escenario prospectivo una integración escolar de inteligencia artificial orientada por verificación, explicación y revisión del razonamiento, con acompañamiento docente y alternativas sin conexión. Su posible institucionalización quedó condicionada a evidencia reproducible, protección de la información y evaluación de sostenibilidad. Los resultados no permitieron prometer efectos futuros, pero sustentaron el interés de estudiar bajo qué condiciones esas prácticas podrían mantenerse.

REFERENCIAS BIBLIOGRÁFICAS

Aguirre Mejía, E. T., Arias Hernández, L. E., Urbina Amador, A. L., Parada Avila, S. V., & Meraz Salazar, E. A. (2026). Pensamiento crítico y escritura científica mediada por IA: análisis de una intervención con estudiantes universitarios. *Ciencia Latina Revista Científica Multidisciplinar, 10*(1), 1381–1405. https://doi.org/10.37811/cl_rcm.v10i1.22290

Benavides, N., & Zambrano-Ramírez, J. (2023). Comprensión lectora autorregulada apoyada en tecnología en estudiantes de Educación Básica. *Revista Electrónica Educare, 27*(3), 1–19. https://doi.org/10.15359/ree.27-3.17221

Benvenuti, M., Cangelosi, A., Weinberger, A., Mazzoni, E., Benassi, M., Barbaresi, M., & Orsoni, M. (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. *Computers in Human Behavior, 148*, 107903. https://doi.org/10.1016/j.chb.2023.107903

Bolaños Gómez, A. C., Márquez Zambrano, L. G., Tagua Pomaina, J. E., & Vinueza Morales, M. G. (2026). Inteligencia artificial en educación superior: ¿Sustituto o complemento del pensamiento crítico? Revisión sistemática. *Revista Simón Rodríguez, 6*(11), 635–652. https://doi.org/10.62319/simonrodriguez.v.6i11.157

Castro Macías, N. M. (2025). Aplicación de la inteligencia artificial como recurso para desarrollar pensamiento crítico en clases de Física. *Revista Internacional Horizonte Científico, 3*(2), 1–19. https://doi.org/10.64747/7wvch719

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*(1), 38. https://doi.org/10.1186/s41239-023-00408-3

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*(1), 43. https://doi.org/10.1186/s41239-023-00411-8

Chiappe, A., Sanmiguel, C., & Sáez Delgado, F. M. (2025). IA generativa versus profesores: reflexiones desde una revisión de la literatura. *Pixel-Bit. Revista de Medios y Educación, 72*, 119–137. https://doi.org/10.12795/pixelbit.107046

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

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

Filipps Díaz, E. V. (2026). *Estrategia pedagógica de aplicación de la inteligencia artificial para el desarrollo del pensamiento crítico en educación primaria* [Tesis doctoral inédita, Universidad Mayor, Real y Pontificia de San Francisco Xavier de Chuquisaca].

Flores González, E., Jojoa Santacruz, R., & Trujillo Rodríguez, A. (2025). Cuando la tecnología piensa: desafíos y oportunidades para el pensamiento crítico. *Fedumar Pedagogía y Educación, 12*(1), 174–182. https://doi.org/10.31948/fpe.v12i1.4972

Gil-Duran, N., & Mogas, J. (2026). Interacción de los estudiantes con ChatGPT durante una tarea de interpretación de fuentes en la formación inicial del profesorado: implicaciones para el pensamiento crítico. *Social Sciences, 15*(8), 517. https://doi.org/10.3390/socsci15080517

Ismail, A. F., Neyef Al-Saliti, R. A. M., Al-Muoaeweed, O., & Mohamed Hamid, A. A. (2026). The effectiveness of a doubt-based learning program in enhancing critical thinking and self-efficacy in verifying AI-generated knowledge among university

students. *Journal of Educational and Social Research, 16*(3), 327. https://doi.org/10.36941/jesr-2026-0335

Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., S, M., & Joseph, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. *Frontiers in Psychology, 16*, 1550621. https://doi.org/10.3389/fpsyg.2025.1550621

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies, 27*(5), 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

Lee, H.-Y., Chen, P.-H., Wang, W.-S., Huang, Y.-M., & Wu, T.-T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education, 21*(1), 16. https://doi.org/10.1186/s41239-024-00447-4

Li, C., Cui, H., & Hagedorn, L. S. (2026). The cognitive impact of ChatGPT in higher education: A systematic review of critical and creative thinking outcomes. *Computers & Education: Artificial Intelligence, 10*, 100571. https://doi.org/10.1016/j.caeai.2026.100571

López Proaño, A. A., & Abad Arroyo, A. A. (2025). Impacto de la IA en el desarrollo de habilidades y competencias del siglo XXI. *Ciencia y Reflexión, 4*(2), 801–836. https://doi.org/10.70747/cr.v4i2.321

Mitevska-Poceva, M. (2025). De los juegos a los algoritmos: integración de juegos e IA para fomentar el pensamiento crítico en los estudiantes. *KAIROS: Revista de Medios y Comunicación, 4*(2), 83–96. https://doi.org/10.64370/KEJO9642

Moya Figueroa, B., & Eaton, S. E. (2023). Examining recommendations for artificial intelligence use with integrity from a scholarship of teaching and learning lens. *RELIEVE – Revista Electrónica de Investigación y Evaluación Educativa, 29*(2). https://doi.org/10.30827/relieve.v29i2.29295

Oyarvide Estupiñán, N., Tenorio Canchingre, E., Oyarvide Ibarra, R., Oyarvide Ramírez, H., & Racines Alban, T. (2024). Factores influyentes para el uso de herramientas digitales en estudiantes universitarios. *Revista Científica de Salud y Desarrollo Humano, 5*(2), 346–366. https://doi.org/10.61368/rsdhv5i2.141

Perdana, R., Yulianti, D., & Bertiliya, W. A. (2025). Enhancing scientific literacy through AI-integrated Inquiry Social Complexity module in primary education. *Journal of Educational Management and Instruction (JEMIN), 5*(2), 388–404. https://doi.org/10.22515/jemin.v5i2.10891

Rodríguez Vela, B. R., Tapia Martínez, J. R., & Castro Luna, E. J. (2026). Percepciones y usos de la inteligencia artificial en la formación inicial docente: estudio cuasiexperimental en estudiantes de educación preescolar. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo, 16*(32), e1109. https://doi.org/10.23913/ride.v16i32.2973

Sánchez Muñoz, J., Flores-Eraña, G., Silva-Campos, J., Chavira-Quintero, R., & Olais-Govea, J. (2025). GenAI as a cognitive mediator: A critical-constructivist inquiry into computational thinking in pre-university education. *Frontiers in Education, 10*, 1597249. https://doi.org/10.3389/feduc.2025.1597249

Tian, J., & Zhang, R. (2026). Outsourcing thinking to AI? Focused immersion, AI dependency, and the double-edged impact on critical thinking. *Humanities and Social Sciences Communications, 13*(1), 1015. https://doi.org/10.1057/s41599-026-07153-8

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*(1), 15. https://doi.org/10.1186/s41239-024-00448-3

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education, 21*(1), 40. https://doi.org/10.1186/s41239-024-00468-z

Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. *International Journal of STEM Education, 9*(1), 59. https://doi.org/10.1186/s40594-022-00377-5

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour, 8*(10), 1839–1850. https://doi.org/10.1038/s41562-024-02004-5

Yan, L., Pammer-Schindler, V., Mills, C., Nguyen, A., & Gašević, D. (2025). Beyond efficiency: Empirical insights on generative AI's impact on cognition,

metacognition and epistemic agency in learning. *British Journal of Educational Technology, 56*(5), 1675–1685. https://doi.org/10.1111/bjet.70000

Yim, I. H. Y., & Su, J. (2025). Artificial intelligence (AI) learning tools in K-12 education: A scoping review. *Journal of Computers in Education, 12*, 93–131. https://doi.org/10.1007/s40692-023-00304-9

Yulianti, D., & Herpratiwi, H. (2024). Development of a science, environment, technology, and society-based learning module to foster critical thinking in elementary students. *Journal of Education and Learning (EduLearn), 18*(4), 1372–1384. https://doi.org/10.11591/edulearn.v18i4.21713

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*(1), 28. https://doi.org/10.1186/s40561-024-00316-7

CONFLICTO DE INTERESES

No existe.

FINANCIAMIENTO

No aplica.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Elizabeth Virginia Filipps Díaz.
Metodología: Elizabeth Virginia Filipps Díaz.
Investigación: Elizabeth Virginia Filipps Díaz.
Redacción – borrador original: Elizabeth Virginia Filipps Díaz.
Redacción – revisión y edición: Elizabeth Virginia Filipps Díaz.