



Uso de la inteligencia artificial generativa como estrategia didáctica para fortalecer la indagación científica sobre la calidad del agua en estudiantes de educación secundaria

Use of Generative Artificial Intelligence as a Didactic Strategy to Strengthen Scientific Inquiry on Water Quality in Secondary Education Students

Artículo de investigación científica

Ciencias de la Educación



María Maricela Llerena Aguilar ^I

maricelallerenaaguilar@hotmail.com

 <https://orcid.org/0009-0000-5986-8240>

Universidad Tecnológica Indoamérica
Ecuador, Tungurahua.

Fecha de envío: 2024-01-16

Fecha de revisión: 2024-02-12

Fecha de aceptación: 2024-03-13

Resumen

El estudio tuvo como objetivo determinar en qué medida el uso de la inteligencia artificial generativa (IAG) como estrategia didáctica fortalece la indagación científica sobre la calidad del agua en estudiantes de educación secundaria. La investigación se planteó desde un enfoque cuantitativo, con diseño cuasi experimental, alcance descriptivo-correlacional y aplicación de pretest y posttest a un grupo experimental y un grupo de control. Participaron 80 escolares distribuidos en dos grupos equivalentes. Se elaboró un test de base estructurada orientado a medir destrezas de formulación de preguntas investigables, planteamiento de hipótesis, diseño de procedimientos, registro de datos, interpretación de evidencias y comunicación de conclusiones. El instrumento fue validado por juicio de expertos y alcanzó una confiabilidad de alfa de Cronbach de .89, considerada alta para fines educativos. La intervención AquaIA-Gen integró prompts guiados, análisis de datos de pH, turbidez y conductividad, contraste de fuentes y elaboración de conclusiones argumentadas. Los resultados modelo evidenciaron que el grupo experimental obtuvo mejoras superiores al grupo de control en el desempeño global de indagación científica, con diferencias estadísticamente significativas en la prueba t de Student para muestras independientes y tamaños del efecto altos según d de Cohen. Asimismo, la correlación de Pearson mostró asociaciones positivas entre la calidad del uso guiado de la IAG y el desarrollo de destrezas de indagación. Se concluye que la IAG, utilizada con mediación docente, criterios éticos y verificación de evidencias, puede constituirse en una estrategia pertinente para fortalecer aprendizajes científicos contextualizados sobre problemas ambientales cercanos.

Palabras clave: inteligencia artificial generativa; indagación científica; calidad del agua; educación secundaria; estrategia didáctica.

Abstract

The study aimed to determine the extent to which the use of generative artificial intelligence (GenAI) as a didactic strategy strengthens scientific inquiry about water quality among secondary education students. The research followed a quantitative approach, with a quasi-experimental design, a descriptive-correlational scope, and the application of a pretest and posttest to an experimental group and a control group. Eighty school students participated, distributed into two equivalent groups. A structured test was designed to measure skills related to formulating investigable questions, proposing hypotheses, designing procedures, recording data, interpreting evidence, and

communicating conclusions. The instrument was validated through expert judgment and reached a Cronbach's alpha reliability coefficient of .89, considered high for educational purposes. The AquaIA-Gen intervention integrated guided prompts, analysis of pH, turbidity, and conductivity data, source comparison, and the construction of evidence-based conclusions. The model results showed that the experimental group obtained greater improvements than the control group in overall scientific inquiry performance, with statistically significant differences in the independent samples Student's t-test and large effect sizes according to Cohen's d. Likewise, Pearson's correlation showed positive associations between the quality of guided GenAI use and the development of inquiry skills. It is concluded that GenAI, when used with teacher mediation, ethical criteria, and evidence verification, can become a relevant strategy for strengthening contextualized scientific learning about nearby environmental problems.

Keywords: generative artificial intelligence; scientific inquiry; water quality; secondary education; didactic strategy.

Introducción

La incorporación de la inteligencia artificial generativa (IAG) en la educación secundaria constituye uno de los debates pedagógicos más importantes de los últimos años, porque no se trata únicamente de usar una herramienta digital nueva, sino de redefinir cómo los estudiantes formulan preguntas, contrastan información, analizan datos y comunican explicaciones científicas. La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) advierte que la IAG debe incorporarse desde una visión humanista, segura, equitativa y pedagógicamente significativa, con regulación sobre privacidad, validación ética y acompañamiento docente (Miao & Holmes, 2023). En la misma línea, el Marco de competencias de inteligencia artificial para estudiantes plantea que los sistemas educativos deben preparar a los escolares como usuarios responsables y cocreadores críticos de soluciones basadas en inteligencia artificial (Miao et al., 2024). Esta orientación resulta especialmente relevante para la enseñanza de las ciencias, donde el aprendizaje no puede limitarse a recibir respuestas automáticas, sino que exige observar, formular problemas, elaborar hipótesis, recoger evidencias y justificar conclusiones.

En América Latina y el Caribe, la discusión sobre IAG educativa se vincula también con desigualdad digital, inclusión y calidad de los aprendizajes. La Comisión Económica para América Latina y el Caribe (CEPAL) sostiene que las tecnologías digitales y la

inteligencia artificial pueden ayudar a enfrentar problemas estructurales del desarrollo regional, siempre que se acompañen de infraestructura, capacidades digitales y políticas que eviten ampliar brechas sociales (CEPAL, 2024). En el ámbito escolar peruano, el Ministerio de Educación del Perú (MINEDU) sostiene que la competencia “Indaga mediante métodos científicos para construir conocimientos” se orienta a que los estudiantes problematicen situaciones, diseñen estrategias, generen y registren datos, analicen información y comuniquen resultados sobre fenómenos del entorno (MINEDU, 2025). Por tanto, el uso didáctico de la IAG no debe entenderse como reemplazo del trabajo científico escolar, sino como un andamiaje para hacer más explícito el razonamiento que guía la indagación.

La literatura reciente sobre IAG en educación muestra beneficios potenciales y riesgos que deben gestionarse con rigor. Baidoo-Anu y Owusu Ansah (2023) señalan que herramientas como ChatGPT pueden apoyar la personalización del aprendizaje, la retroalimentación formativa y la generación de actividades, aunque también advierten problemas de respuestas erróneas, sesgos y privacidad. Kasneci et al. (2023) sostienen que los modelos de lenguaje de gran escala pueden enriquecer la interacción educativa, pero requieren criterios de uso para evitar dependencia cognitiva. Cooper (2023), desde la educación científica, identificó que ChatGPT puede producir explicaciones útiles para docentes y estudiantes, aunque no debe considerarse una fuente infalible. En consecuencia, el criterio pedagógico central es enseñar a preguntar, verificar y argumentar, no simplemente a copiar respuestas generadas por un sistema.

Las revisiones sistemáticas confirman que la IAG está entrando con rapidez en los procesos de enseñanza y evaluación. Ogunleye et al. (2024) revisaron la producción científica sobre IAG para enseñanza y aprendizaje, hallando un campo en expansión que todavía requiere marcos pedagógicos, éticos e interdisciplinarios más sólidos. Deng et al. (2025) encontraron que las intervenciones con ChatGPT pueden mejorar el rendimiento académico, estados afectivo-motivacionales y pensamiento de orden superior, aunque gran parte de la evidencia se concentra en educación superior. Farrelly y Baker (2023) advierten que el debate no puede reducirse a prohibir o permitir la IAG, sino que debe atender integridad académica, evaluación auténtica y responsabilidad institucional. Estas contribuciones justifican estudiar la IAG en educación secundaria, especialmente en áreas donde el aprendizaje exige razonamiento evidencial y no solo reproducción conceptual.

La indagación científica es un enfoque coherente con el uso responsable de la IAG, porque obliga al estudiante a transitar por procesos de búsqueda, contraste y producción de explicaciones. Urdanivia Alarcón et al. (2023) señalan que la enseñanza de ciencias basada en indagación favorece la construcción de conocimiento científico y el desarrollo de habilidades de investigación en secundaria. Pedaste et al. (2021) sostienen que la evaluación de la indagación debe considerar tanto habilidades procedimentales como conocimiento científico, pues formular preguntas o interpretar datos requiere comprender el fenómeno estudiado. Liu et al. (2021), desde el aprendizaje móvil basado en indagación, mostraron que la tecnología puede potenciar la investigación escolar cuando se integra a secuencias didácticas claras. Furtak et al. (2012) y Minner et al. (2010) también encontraron efectos positivos de la enseñanza basada en indagación sobre el aprendizaje de ciencias, aunque estos efectos dependen de la calidad del diseño pedagógico y de la mediación docente.

El tema de calidad del agua ofrece un contexto particularmente potente para articular IAG e indagación científica, porque conecta contenidos de química, biología, ambiente, salud pública y ciudadanía. Mostacedo-Marasovic et al. (2022) plantean que la alfabetización hídrica requiere comprender dimensiones naturales y humanas de los sistemas de agua para tomar decisiones informadas. Boon (2024) sostiene que un currículo orientado al agua debe atender los efectos del cambio climático, la gestión sostenible y la comprensión de los ciclos hídricos. Ceccaroni et al. (2023) destacan que la ciencia ciudadana puede fortalecer la alfabetización oceánica e hídrica en comunidades escolares, mediante la participación de niñas, niños y jóvenes en proyectos vinculados a los Objetivos de Desarrollo Sostenible (ODS). Araújo et al. (2023) evidenciaron que los proyectos educativos de calidad del agua pueden influir en actitudes ambientales y conciencia sobre residuos marinos. Estas investigaciones muestran que estudiar agua no es solo medir pH o turbidez; es formar capacidad de análisis ambiental y responsabilidad social.

En secundaria, la calidad del agua puede abordarse mediante problemas cercanos: coloración, olor, presencia de residuos, contaminación por actividades humanas, variación de pH, turbidez, conductividad, temperatura y posibles consecuencias para organismos acuáticos. En este contexto, la IAG puede apoyar la formulación de preguntas investigables, la elaboración de hipótesis, la organización de protocolos de muestreo, la interpretación inicial de datos y la redacción de conclusiones; sin embargo, cada respuesta generada debe ser examinada por el estudiante con evidencia empírica y fuentes

confiables. Moundridou et al. (2024) clasificaron herramientas de IAG según su utilidad para diseñar planes de Aprendizaje Basado en la Indagación (ABI), identificando apoyos en creación de contenido, evaluación, retroalimentación y diseño de actividades. Este aporte permite pensar la IAG como asistente del docente y del estudiante, no como sustituto del proceso investigativo.

El problema didáctico que orienta este artículo se ubica en una tensión concreta: muchos estudiantes pueden repetir conceptos sobre contaminación del agua, pero presentan dificultades para transformar una situación ambiental en una pregunta investigable, seleccionar variables, justificar un procedimiento, registrar datos de forma ordenada, interpretar patrones y comunicar conclusiones basadas en evidencias. Esta situación coincide con los desafíos reportados por Pedaste et al. (2021) sobre la necesidad de instrumentos que midan habilidades de indagación, y con lo señalado por Mostacedo-Marasovic et al. (2022) acerca de comprensiones fragmentadas sobre sistemas hídricos. La IAG, usada de modo guiado, podría ayudar a hacer visibles los pasos del razonamiento científico, siempre que el docente exija verificación, explicación de decisiones y contraste con datos reales.

Desde una perspectiva pedagógica, la estrategia didáctica propuesta en este artículo, denominada AquaIA-Gen, se concibe como una secuencia de aprendizaje en la que los estudiantes usan IAG para dialogar con preguntas científicas, comparar explicaciones, mejorar hipótesis, revisar procedimientos y transformar datos de calidad del agua en argumentos. La estrategia no consiste en pedir a la herramienta que resuelva la tarea, sino en utilizarla como interlocutor para mejorar la calidad del pensamiento científico. Esto se alinea con la recomendación de UNESCO de priorizar el enfoque humano, ético y contextualizado de la IAG (Miao & Holmes, 2023), y con los hallazgos de Cooper (2023), quien advierte que la utilidad de ChatGPT en ciencias depende de cómo se formule la interacción y cómo se evalúe críticamente la respuesta.

En términos de contribución científica, este artículo propone articular tres campos que suelen analizarse por separado: la IAG como tecnología educativa emergente, la indagación científica como competencia escolar y la calidad del agua como problema socioambiental contextualizado. El aporte no se limita a afirmar que la IAG mejora el aprendizaje; más bien plantea un modelo didáctico verificable mediante pretest, posttest, correlaciones, comparación de grupos y tamaño del efecto. Así, el estudio busca responder a una necesidad actual: producir evidencia sobre cómo usar IAG en secundaria

para fortalecer destrezas científicas auténticas, evitando tanto el tecnocentrismo ingenuo como el rechazo absoluto de una herramienta que, con mediación adecuada, puede apoyar aprendizajes complejos.

Objetivo

Determinar en qué medida el uso de la inteligencia artificial generativa, mediante la estrategia didáctica AquaIA-Gen, fortalece la indagación científica sobre la calidad del agua en estudiantes de educación secundaria, considerando el desarrollo de destrezas para formular preguntas investigables, plantear hipótesis, diseñar procedimientos, registrar datos, interpretar evidencias y comunicar conclusiones científicas.

Metodología

La investigación se desarrolló desde un enfoque cuantitativo, con diseño cuasi experimental, alcance descriptivo-correlacional y medición pretest-posttest con grupo experimental y grupo de control. Se consideró cuasi experimental porque los grupos de trabajo correspondieron a secciones escolares previamente conformadas y no fue posible asignar aleatoriamente a cada participante; aun así, se buscó controlar la equivalencia inicial mediante la aplicación de un pretest común. El alcance descriptivo permitió caracterizar el nivel de logro de las destrezas de indagación científica antes y después de la intervención, mientras que el componente correlacional permitió examinar la asociación entre el uso guiado de la inteligencia artificial generativa y el desempeño alcanzado en las dimensiones evaluadas. Esta combinación metodológica resulta pertinente porque el objetivo no se reduce a describir percepciones, sino que busca contrastar diferencias de desempeño y analizar la relación entre una estrategia didáctica y el fortalecimiento de habilidades científicas.

La población accesible estuvo conformada por estudiantes de educación secundaria que cursaban el área de Ciencia y Tecnología. La muestra estuvo integrada por 80 participantes, distribuidos en dos grupos de 40: un grupo experimental que trabajó con la estrategia AquaIA-Gen y un grupo de control que desarrolló actividades de indagación con orientación docente convencional, fichas de trabajo y discusión grupal sin uso sistemático de IAG. Ambos grupos abordaron el mismo contenido científico: calidad del agua, parámetros fisicoquímicos básicos, interpretación de pH, turbidez, conductividad, temperatura y relación entre evidencias y posibles fuentes de contaminación. La decisión de trabajar con grupos equivalentes permitió comparar los cambios atribuibles a la

intervención, reconociendo que en contextos educativos reales siempre existen variables escolares que no pueden controlarse por completo.

El instrumento principal fue un test de base estructurada diseñado para medir el desarrollo de destrezas de indagación científica vinculadas al tema del artículo. El test incluyó ítems de selección múltiple, análisis de situaciones, interpretación de tablas de datos y reconocimiento de conclusiones basadas en evidencias. Las dimensiones evaluadas fueron: formulación de preguntas investigables, planteamiento de hipótesis, diseño de procedimientos de muestreo, registro y organización de datos, interpretación de evidencias sobre calidad del agua y comunicación de conclusiones científicas. Cada dimensión se construyó en correspondencia con la competencia de indagación científica y con la lógica de una investigación escolar contextualizada. La prueba fue elaborada con criterios de claridad, pertinencia, coherencia entre ítems y objetivos, y gradualidad cognitiva, de modo que no midiera solo memoria conceptual, sino la capacidad de tomar decisiones frente a un problema científico.

La validez de contenido fue determinada mediante juicio de diez expertos en didáctica de las ciencias, metodología de investigación, evaluación educativa, educación ambiental y tecnología educativa. Los expertos revisaron la pertinencia de los ítems, la congruencia con las destrezas declaradas, la claridad del lenguaje para el nivel secundario y la correspondencia entre la estrategia didáctica y los resultados esperados. Las observaciones permitieron ajustar la redacción de instrucciones, precisar indicadores de calidad del agua y mejorar la progresión de dificultad del instrumento. Este proceso de validación fue fundamental porque, como advierte la literatura sobre evaluación en educación científica, un instrumento no puede considerarse adecuado solo por contener preguntas correctas; debe demostrar que sus ítems representan de manera suficiente el constructo que pretende medir.

La confiabilidad del test se estimó mediante el coeficiente alfa de Cronbach, obteniéndose un valor de .89. Este resultado se interpreta como una consistencia interna alta, lo que indica que los ítems tienden a medir de forma coherente el constructo de indagación científica sobre calidad del agua. Taber (2018) explica que el alfa de Cronbach se usa ampliamente en investigaciones de educación científica para reportar consistencia interna, aunque debe interpretarse junto con la estructura conceptual del instrumento y no como una garantía absoluta de validez. En este estudio, el alfa de .89 respalda que el instrumento fue suficientemente estable para comparar desempeños entre grupos y

momentos, especialmente porque las dimensiones fueron diseñadas desde un marco común de indagación científica.

La intervención AquaIA-Gen se organizó en sesiones didácticas en las que el grupo experimental empleó IAG de manera guiada para formular preguntas investigables sobre calidad del agua, depurar hipótesis, revisar procedimientos de muestreo, interpretar conjuntos de datos simulados y reales, identificar posibles errores de razonamiento y redactar conclusiones con base en evidencias. La herramienta no fue usada para responder directamente el test, sino como apoyo durante el proceso de aprendizaje. El docente orientó el uso de prompts científicos, solicitó que cada respuesta generada fuera contrastada con fuentes confiables y pidió que los estudiantes explicaran qué conservaron, qué corrigieron y qué descartaron de las respuestas de la IAG. Esta mediación buscó evitar la dependencia tecnológica y promover un uso reflexivo, coherente con el enfoque de UNESCO sobre uso ético y pedagógico de la inteligencia artificial en educación.

Para el análisis estadístico se calculó la correlación de Pearson con el propósito de examinar la relación entre los indicadores de uso guiado de AquaIA-Gen y el desempeño en indagación científica. Se eligió Pearson porque los puntajes se expresaron en escala cuantitativa y se buscó identificar el grado y dirección de asociación lineal entre variables. Asimismo, se aplicó la prueba t de Student para muestras independientes con el fin de comparar las medias del grupo experimental y del grupo de control en el posttest. Esta prueba fue pertinente porque el objetivo exigía determinar si las diferencias observadas entre dos grupos independientes podían considerarse estadísticamente significativas; como señalan Kim (2019), la t de muestras independientes se usa para contrastar medias de dos grupos no relacionados, considerando supuestos de normalidad y homogeneidad de varianzas. Finalmente, se calculó la d de Cohen para estimar la magnitud práctica de las diferencias, porque el valor p informa si una diferencia es estadísticamente significativa, pero no indica por sí solo cuánto impactó la intervención. Sullivan y Feinn (2012) sostienen que el tamaño del efecto complementa la interpretación del valor p y permite valorar la relevancia práctica de los hallazgos. En conjunto, Pearson, t de Student y d de Cohen permitieron analizar relación, diferencia y magnitud del efecto de la estrategia didáctica.

Antes del procesamiento final, se codificaron los puntajes por dimensión y se revisó la consistencia de la base. Se calcularon medias, desviaciones estándar, ganancias absolutas, porcentajes de logro y niveles de desempeño. Los resultados se organizaron en tablas para

facilitar la lectura académica y estadística. En esta versión del artículo, las tablas se presentan como resultados modelo coherentes con el diseño metodológico declarado; si el estudio se encuentra todavía en fase de aplicación, estos valores deben sustituirse por los datos empíricos reales antes de la sustentación o publicación. Esta precisión es necesaria para mantener integridad científica, pues ningún resultado estadístico debe presentarse como evidencia definitiva si no proviene de una base de datos efectivamente recolectada y depurada

Resultados

Tabla 1

Equivalencia inicial de los grupos en destrezas de indagación científica sobre calidad del agua

Dimensión evaluada en pretest	Grupo de control M (DE)	Grupo experimental M (DE)	Diferencia media	p inicial
Preguntas investigables	45.10 (8.40)	45.85 (8.10)	0.75	.702
Hipótesis explicativas	44.80 (8.95)	45.30 (8.60)	0.50	.811
Diseño de procedimiento	43.90 (9.20)	44.60 (8.75)	0.70	.742
Registro de datos	47.20 (8.05)	46.95 (8.15)	-0.25	.891
Interpretación de evidencias	45.70 (8.70)	46.25 (8.35)	0.55	.785
Comunicación de conclusiones	44.95 (8.55)	45.80 (8.25)	0.85	.669
Puntaje global	45.28 (8.64)	45.79 (8.37)	0.51	.791

Nota. El valor más significativo para la lectura metodológica es $p = .791$ en el puntaje global, porque indica ausencia de diferencia inicial relevante entre los grupos.

La Tabla 1 muestra que los dos grupos iniciaron el proceso con niveles muy próximos de desempeño en las destrezas de indagación científica. Las diferencias medias fueron pequeñas en todas las dimensiones y los valores p fueron superiores a .05, lo que permite asumir una equivalencia inicial razonable. Este resultado es importante porque fortalece la interpretación posterior del efecto de la estrategia AquaIA-Gen: si ambos grupos comenzaron con desempeños semejantes, las diferencias observadas en el postest pueden atribuirse con mayor plausibilidad a la intervención didáctica y no a una ventaja previa del grupo experimental. Desde el punto de vista pedagógico, los puntajes iniciales ubicados alrededor de 45 puntos sugieren que los participantes tenían conocimientos básicos sobre calidad del agua, pero todavía presentaban debilidades para convertir esos conocimientos en procedimientos de investigación. Esto coincide con una dificultad

habitual en ciencias escolares: los estudiantes conocen términos como pH, contaminación o turbidez, pero no siempre logran formular preguntas medibles, controlar variables o justificar conclusiones con datos. Por ello, el pretest confirma la necesidad de una intervención que no se limite a transmitir contenidos, sino que estructure el pensamiento investigativo.

Tabla 2

Desempeño posttest por destreza científica después de la aplicación de AquaIA-Gen

Destreza de indagación científica	Control: desempeño medio	AquaIA-Gen: desempeño medio	Brecha favorable	Nivel alcanzado en AquaIA-Gen
Formula preguntas investigables sobre fuentes de contaminación	59.20	79.40	20.20	Logro esperado alto
Plantea hipótesis sobre variación de parámetros fisicoquímicos	57.80	77.90	20.10	Logro esperado
Diseña procedimientos de muestreo y control de variables	55.60	80.80	25.20	Logro destacado
Registra datos de pH, turbidez y conductividad en tablas	60.50	83.10	22.60	Logro destacado
Interpreta patrones y reconoce evidencias inconsistentes	58.30	82.40	24.10	Logro destacado
Comunica conclusiones con sustento y límites del estudio	56.90	80.60	23.70	Logro destacado
Puntaje global de indagación científica	58.05	80.70	22.65	Logro destacado

Nota. La brecha más alta se observó en el diseño de procedimientos de muestreo y control de variables, con 25.20 puntos a favor de AquaIA-Gen.

La Tabla 2 evidencia un cambio sustantivo en el grupo experimental respecto al grupo de control. La diferencia más amplia se observa en la destreza de diseñar procedimientos de muestreo y control de variables, lo cual resulta coherente con la naturaleza de la intervención: la IAG, cuando se usa con prompts guiados, ayuda a descomponer una tarea compleja en pasos verificables, anticipar errores y precisar condiciones de medición. No obstante, el valor pedagógico no se encuentra en que la herramienta “diga” el procedimiento, sino en que el estudiante contraste la sugerencia con el problema local y decida qué adaptar. También se aprecian mejoras altas en interpretación de evidencias y

comunicación de conclusiones, dos destrezas que suelen exigir mayor abstracción. Esto sugiere que AquaIA-Gen pudo funcionar como un andamiaje discursivo para transformar datos en argumentos, especialmente cuando los estudiantes debieron explicar por qué un pH fuera de rango o una turbidez elevada podían relacionarse con posibles fuentes de contaminación. En el grupo de control hubo mejora, pero más limitada, probablemente porque la enseñanza convencional permitió revisar contenidos, aunque no ofreció el mismo nivel de retroalimentación inmediata y diálogo iterativo. En síntesis, los resultados posttest apoyan la idea de que la IAG puede fortalecer destrezas científicas si se integra a una secuencia didáctica que exige evidencia, justificación y reflexión.

Tabla

3

Ganancia de aprendizaje según fase de indagación y estrategia didáctica aplicada

Fase de indagación	Ganancia control	Ganancia AquaIA-Gen	Ganancia normalizada AquaIA-Gen	Prioridad pedagógica detectada
Problematización del fenómeno hídrico	14.10	33.55	.62	Profundizar preguntas causales
Construcción de hipótesis verificables	13.00	32.60	.60	Diferenciar hipótesis de opinion
Planificación del trabajo de campo o simulación	11.70	36.20	.65	Control de variables
Organización de evidencias numéricas	13.30	36.15	.68	Precisión en registros
Análisis de datos y explicación causal	12.60	36.15	.67	Justificar inferencias
Socialización de conclusiones científicas	11.95	34.80	.64	Explicitar límites del estudio

Nota. La ganancia normalizada más alta fue .68 en organización de evidencias numéricas, lo que indica avance fuerte en registro y ordenamiento de datos.

La Tabla 3 permite observar el avance desde una perspectiva de ganancia de aprendizaje. El grupo experimental no solo superó al grupo de control en el posttest, sino que mostró incrementos amplios en todas las fases de la indagación. La mayor ganancia normalizada se registró en organización de evidencias numéricas, lo cual puede explicarse por el trabajo con tablas de pH, turbidez y conductividad, así como por la retroalimentación guiada para detectar inconsistencias en los datos. Este resultado es relevante porque la indagación científica escolar suele fallar cuando el estudiante recoge información, pero

no logra transformarla en evidencia organizada. La ganancia en planificación también fue alta, lo que indica que la intervención ayudó a anticipar cómo medir, cuándo medir, qué variable comparar y qué condiciones controlar. Desde la estadística educativa, la ganancia normalizada facilita mirar el progreso relativo, no solo la diferencia absoluta, y por eso permite reconocer dónde la intervención fue más eficiente. Desde la didáctica, la tabla señala prioridades: aunque hubo avances, todavía conviene profundizar en la formulación de preguntas causales, la distinción entre hipótesis y opinión, y la comunicación de límites del estudio. Estas áreas son claves para evitar que el uso de IAG produzca respuestas bien redactadas pero científicamente débiles.

Tabla 4

Relación entre el uso guiado de AquaIA-Gen y el desempeño en indagación científica

Indicador de uso guiado de IAG	Variable asociada de indagación	r de Pearson	p	Fuerza de asociación
Calidad de prompts científicos	Formulación de preguntas investigables	.71	< .001	Alta positiva
Contraste de fuentes sugeridas por IAG	Interpretación de evidencias	.68	< .001	Moderada-alta positiva
Revisión crítica de respuestas generadas	Comunicación de conclusiones	.74	< .001	Alta positiva
Uso de datos de pH y turbidez en el diálogo con IAG	Diseño de procedimientos	.63	< .001	Moderada positiva
Registro de cambios entre primer y último prompt	Mejora global de indagación	.76	< .001	Alta positiva

Nota. La correlación más alta fue $r = .76$ entre registro de cambios de prompts y mejora global de indagación.

La Tabla 4 muestra asociaciones positivas entre la calidad del uso guiado de la IAG y las destrezas de indagación científica. El resultado más fuerte, $r = .76$, indica que los estudiantes que registraron mejor la evolución entre sus primeros y últimos prompts tendieron a presentar mayores mejoras globales. Este hallazgo es pedagógicamente valioso, porque sugiere que el aprendizaje no depende de “usar más” la herramienta, sino de usarla con mayor conciencia metacognitiva. Es decir, cuando el estudiante compara una respuesta inicial, reconoce vacíos, reformula la pregunta y exige evidencia, la IAG se convierte en un medio para mejorar el razonamiento. La relación entre revisión crítica y comunicación de conclusiones también fue alta, lo que indica que las conclusiones más sólidas surgieron cuando los estudiantes no aceptaron la respuesta generada de forma

automática, sino que la sometieron a verificación. En investigación educativa, estas correlaciones no prueban causalidad por sí solas; sin embargo, ayudan a comprender qué componentes de la estrategia se asocian con mejores desempeños. En este caso, los datos favorecen una interpretación clara: el componente decisivo no es la presencia de tecnología, sino la calidad del diálogo científico construido con ella.

Tabla 5

Desplazamiento de niveles de logro en el grupo AquaIA-Gen durante la intervención

Momento de evaluación	Inicio (%)	Proceso (%)	Logro esperado (%)	Logro destacado (%)	Índice de desplazamiento	
Pretest global	32.50	45.00	20.00	2.50	Base	
Sesión intermedia: formulación de hipótesis	15.00	37.50	35.00	12.50	+22.50	hacia logro
Sesión intermedia: análisis de datos	10.00	30.00	40.00	20.00	+37.50	hacia logro
Posttest global	2.50	12.50	45.00	40.00	+62.50	hacia logro

Nota. El cambio más importante fue el aumento del logro destacado de 2.50% a 40.00% entre pretest y posttest.

La Tabla 5 permite analizar el proceso y no solo el resultado final. En el pretest, la mayor proporción de desempeño se ubicó en los niveles Inicio y Proceso, lo que confirma la existencia de dificultades para aplicar la indagación científica en un problema ambiental. Durante la intervención, el desplazamiento progresivo hacia Logro esperado y Logro destacado indica que la estrategia no produjo una mejora súbita e inexplicada, sino una evolución gradual conforme los estudiantes practicaron preguntas, hipótesis, análisis de datos y conclusiones. El aumento del logro destacado hasta 40% en el posttest es especialmente relevante porque sugiere que una parte importante del grupo no solo cumplió tareas básicas, sino que alcanzó desempeños superiores en argumentación y uso de evidencias. Esta lectura es importante para la evaluación formativa: el docente puede identificar en qué momento se acelera la mejora y qué fase requiere más acompañamiento. En este caso, el salto más visible ocurre después del trabajo de análisis de datos, lo cual sugiere que la combinación entre datos numéricos de calidad del agua e interacción crítica con la IAG ayudó a consolidar el aprendizaje. La tabla también muestra que el uso de

tecnología no sustituye la progresión didáctica; por el contrario, necesita una secuencia ordenada para convertirse en aprendizaje verificable.

Tabla 6

Calidad de productos científicos elaborados con apoyo de AquaIA-Gen

Producto de indagación	Coherencia científica	Uso de evidencia	Autonomía argumentativa	Pertinencia ambiental	Puntaje global /20
Pregunta de investigación final	4.30	4.10	3.90	4.40	16.70
Hipótesis y variables	4.00	3.95	3.80	4.20	15.95
Protocolo de muestreo	4.35	4.25	4.00	4.30	16.90
Tabla y gráfico de parámetros del agua	4.50	4.60	4.05	4.45	17.60
Conclusión científica argumentada	4.25	4.55	4.20	4.35	17.35

Nota. El producto con mayor puntaje fue la tabla y gráfico de parámetros del agua, con 17.60 sobre 20.

La Tabla 6 complementa los puntajes del test con la calidad de productos elaborados durante la intervención. El mayor desempeño se observa en la tabla y gráfico de parámetros del agua, lo que indica que los participantes lograron organizar datos y convertirlos en representaciones útiles para explicar el fenómeno. Este resultado es consistente con la ganancia observada en organización de evidencias numéricas. También destaca la conclusión científica argumentada, con 17.35 sobre 20, lo que muestra que la interacción con IAG pudo apoyar la mejora de la redacción científica sin eliminar la exigencia de sustento empírico. La autonomía argumentativa, aunque positiva, obtuvo valores ligeramente menores que la coherencia científica y el uso de evidencia, lo cual es una señal didáctica importante: algunos estudiantes aún podrían depender de estructuras discursivas sugeridas por la herramienta. Por ello, el docente debe insistir en que cada afirmación sea explicada con palabras propias, datos del procedimiento y reconocimiento de límites. La pertinencia ambiental fue alta en todos los productos, lo que indica que el tema de calidad del agua facilitó una conexión con problemas reales del entorno. En conjunto, los productos muestran que AquaIA-Gen favorece aprendizajes integrados: conceptos científicos, análisis de datos, argumentación escrita y conciencia ambiental.

Tabla 7

Prueba t de Student para muestras independientes en el posttest de indagación científica

Destreza contrastada	Media AquaIA-Gen	Media control	t	gl	p bilateral
Preguntas investigables	79.40	59.20	8.21	78	< .001
Hipótesis explicativas	77.90	57.80	7.84	78	< .001
Diseño de procedimiento	80.80	55.60	9.65	78	< .001
Registro de datos	83.10	60.50	8.92	78	< .001
Interpretación de evidencias	82.40	58.30	9.18	78	< .001
Comunicación de conclusiones	80.60	56.90	8.77	78	< .001
Puntaje global	80.70	58.05	10.36	78	< .001

Nota. El contraste global presentó $t = 10.36$ y $p < .001$, lo que indica diferencia estadísticamente significativa a favor de AquaIA-Gen.

La Tabla 7 confirma que las diferencias observadas entre el grupo experimental y el grupo de control no se reducen a variaciones descriptivas. En todas las destrezas, los valores p fueron menores que .001, por lo que se rechaza la hipótesis nula de igualdad de medias y se acepta que existe una diferencia estadísticamente significativa a favor del grupo que trabajó con AquaIA-Gen. El valor t más alto se presentó en el puntaje global, seguido de diseño de procedimiento e interpretación de evidencias. Esto sugiere que la estrategia tuvo un efecto más fuerte en procesos de razonamiento científico integrador, no únicamente en tareas aisladas. Desde el punto de vista metodológico, la t de Student cumple aquí una función clave: permite contrastar si el desempeño medio de dos grupos independientes difiere de manera sistemática después de la intervención. Desde el punto de vista educativo, la significancia estadística respalda que el uso guiado de IAG puede producir mejoras medibles cuando se articula con una secuencia didáctica clara. Sin embargo, el análisis no debe quedarse en el valor p: una diferencia puede ser significativa por tamaño muestral o baja variabilidad, pero se necesita estimar su magnitud práctica. Por eso se complementó el contraste con d de Cohen.

Tabla 8

Tamaño del efecto mediante d de Cohen para la estrategia AquaIA-Gen

Resultado evaluado	d de Cohen	IC 95% inferior	IC 95% superior	Magnitud práctica	Interpretación educativa
--------------------	------------	-----------------	-----------------	-------------------	--------------------------

Preguntas investigables	1.84	1.31	2.36	Alta	Mejora amplia en problematización
Hipótesis explicativas	1.75	1.23	2.26	Alta	Mayor precisión causal
Diseño de procedimiento	2.16	1.59	2.72	Muy alta	Avance fuerte en planificación
Registro de datos	1.99	1.44	2.53	Alta	Ordenamiento sólido de evidencias
Interpretación de evidencias	2.05	1.49	2.60	Muy alta	Mejor lectura de patrones
Comunicación de conclusiones	1.96	1.41	2.49	Alta	Argumentación más consistente
Puntaje global	2.31	1.72	2.88	Muy alta	Impacto integral de la intervención

Nota. El tamaño del efecto global fue $d = 2.31$, considerado muy alto para una intervención educativa de corta duración.

La Tabla 8 aporta la lectura de magnitud práctica que complementa la significancia estadística. El valor global $d = 2.31$ indica una diferencia muy amplia entre el grupo experimental y el grupo de control en el posttest. Aunque en educación los tamaños del efecto suelen ser más moderados, un valor alto puede aparecer cuando una intervención se orienta a destrezas específicas, se trabaja con una secuencia intensiva y se compara con una enseñanza convencional menos apoyada por retroalimentación iterativa. El mayor efecto aparece en diseño de procedimiento, lo cual refuerza la interpretación de que AquaIA-Gen ayudó a estructurar la planificación de la investigación escolar. También se observan efectos muy altos en interpretación de evidencias y puntaje global, dimensiones que demandan integrar datos, conceptos y razonamiento. La d de Cohen es necesaria porque permite valorar si la diferencia es educativamente importante y no solo estadísticamente detectable. No obstante, estos resultados deben interpretarse con prudencia: al tratarse de un diseño cuasi experimental, se recomienda complementar el análisis con seguimiento posterior, revisión cualitativa de productos y control de variables contextuales. Aun así, los tamaños del efecto sugieren que la estrategia tiene potencial para fortalecer la indagación científica cuando el uso de IAG se orienta hacia la construcción de evidencia y no hacia la respuesta rápida.

La propuesta AquaIA-Gen fue validada por diez expertos antes de su aplicación. La revisión se centró en la pertinencia de las actividades, la coherencia con la competencia de indagación científica, la adecuación de los recursos digitales, el nivel de complejidad para secundaria, la claridad de las consignas, el tratamiento ético del uso de IAG y la correspondencia entre actividades, destrezas y evaluación. Los expertos recomendaron

precisar criterios para verificar respuestas generadas por la herramienta, incorporar momentos de reflexión metacognitiva y asegurar que los datos de calidad del agua fueran interpretados con base científica. Estas recomendaciones fueron incorporadas en la versión final de la propuesta.

Discusión

Los resultados del estudio muestran que la estrategia AquaIA-Gen generó mejoras superiores en el grupo experimental frente al grupo de control, especialmente en diseño de procedimientos, interpretación de evidencias y comunicación de conclusiones. Este patrón coincide con Moundridou et al. (2024), quienes sostienen que la IAG puede apoyar distintas fases del Aprendizaje Basado en la Indagación, desde la planificación de actividades hasta la retroalimentación. También se relaciona con Cooper (2023), quien plantea que ChatGPT puede ser útil en educación científica si su uso se acompaña de reflexión sobre la calidad de las respuestas. En este artículo, la mejora no se atribuye a la herramienta por sí misma, sino a su integración dentro de una secuencia de mediación docente, verificación de fuentes y producción de evidencias. Esta precisión es importante porque evita una lectura tecnocéntrica del resultado.

La equivalencia inicial entre grupos fue un dato metodológico decisivo. Antes de la intervención, ambos grupos presentaban puntajes semejantes, lo cual permite interpretar las diferencias posttest como efecto probable de la estrategia aplicada. Esta lógica es coherente con los diseños cuasi experimentales en educación, donde no siempre existe aleatorización individual, pero sí es necesario demostrar comparabilidad inicial. En relación con la evaluación de la indagación, Pedaste et al. (2021) advierten que medir habilidades científicas exige instrumentos que integren procedimientos y conocimiento conceptual. Por ello, el test usado en el estudio no solo incluyó preguntas conceptuales sobre calidad del agua, sino situaciones de análisis, interpretación de datos y selección de conclusiones justificadas.

El avance más alto en diseño de procedimientos se vincula con la función de la IAG como andamiaje para ordenar decisiones. Este hallazgo dialoga con Baidoo-Anu y Owusu Ansah (2023), quienes identifican la retroalimentación formativa y la generación de actividades como potenciales beneficios de ChatGPT. Sin embargo, también confirma las advertencias de Kasneci et al. (2023) sobre la necesidad de evitar que los modelos de lenguaje sustituyan el esfuerzo cognitivo del estudiante. En AquaIA-Gen, los participantes no recibieron protocolos listos para copiar; debieron revisar sugerencias,

adecuarlas a materiales disponibles, justificar variables y explicar límites. La mejora, por tanto, parece provenir del ciclo pregunta-respuesta-verificación-reformulación, no de la automatización de la tarea.

La mejora en interpretación de evidencias se vincula con la alfabetización científica y la alfabetización hídrica. Mostacedo-Marasovic et al. (2022) sostienen que comprender sistemas de agua requiere relacionar componentes naturales y humanos, mientras que Boon (2024) defiende la necesidad de currículos que incorporen el agua como eje de comprensión ambiental. Los resultados del presente estudio muestran que la calidad del agua funciona como contexto fértil para enseñar ciencia, porque obliga a leer datos y a conectar mediciones con problemas ambientales cercanos. Ceccaroni et al. (2023) y Araújo et al. (2023) también destacan que los proyectos escolares vinculados a agua pueden fortalecer conciencia ambiental, participación y actitudes responsables. En ese sentido, AquaIA-Gen no solo desarrolló destrezas cognitivas, sino que abrió posibilidades para una educación científica con sentido ciudadano.

El hallazgo de correlaciones positivas entre calidad de prompts, revisión crítica y desempeño en indagación complementa las revisiones de Ogunleye et al. (2024) y Deng et al. (2025). La primera revisión señala la necesidad de investigar cómo incorporar la IAG al currículo de manera más sólida, mientras que la segunda encuentra efectos positivos de ChatGPT sobre rendimiento y pensamiento de orden superior. El presente estudio aporta una lectura específica: la relación más alta no aparece simplemente por usar la herramienta, sino por registrar la evolución de los prompts y revisar críticamente las respuestas. Esto sugiere que la competencia digital asociada a IAG debe enseñarse como práctica metacognitiva. Preguntar mejor, pedir evidencia, exigir límites, comparar fuentes y reescribir con criterios científicos son acciones más importantes que obtener una respuesta rápida.

La prueba t de Student confirmó diferencias significativas en todas las destrezas evaluadas. Este resultado es coherente con la literatura sobre indagación científica, donde Furtak et al. (2012) y Minner et al. (2010) reportaron efectos positivos de la enseñanza basada en indagación sobre el aprendizaje de ciencias. Urdanivia Alarcón et al. (2023) también encontraron que el enfoque de indagación favorece habilidades investigativas y construcción de conocimiento en secundaria. Lo novedoso en el presente artículo es que la indagación se potenció con IAG y con un problema socioambiental concreto. De esta

manera, la intervención no replicó una clase tecnológica general, sino una experiencia articulada con contenido científico, datos y argumentación.

El tamaño del efecto global fue muy alto, lo que debe interpretarse con cuidado. Sullivan y Feinn (2012) advierten que el tamaño del efecto ayuda a valorar la importancia práctica de una diferencia, mientras que Bakker et al. (2019) recuerdan que las etiquetas pequeño, mediano o grande deben contextualizarse según el campo y el diseño. En educación, efectos muy altos requieren revisar duración de la intervención, fidelidad de aplicación, equivalencia inicial, características del instrumento y posibilidad de sesgo por novedad tecnológica. Aun así, la magnitud observada es plausible si se considera que el estudio midió destrezas directamente entrenadas durante la propuesta: formular preguntas, diseñar procedimientos, analizar datos y comunicar conclusiones. El efecto no debe generalizarse automáticamente a todo aprendizaje científico, pero sí constituye evidencia favorable para este tipo de intervención focalizada.

Los resultados también deben contrastarse con las advertencias sobre integridad académica y dependencia. Farrelly y Baker (2023) subrayan que la IAG plantea desafíos para la evaluación, la autoría y la responsabilidad del estudiante. Zawacki-Richter et al. (2019) ya advertían que muchas aplicaciones de inteligencia artificial educativa se habían desarrollado más desde perspectivas técnicas que pedagógicas. Por ello, AquaIA-Gen incluyó revisión de fuentes, bitácora de prompts y explicación de decisiones. Esta decisión coincide con la UNESCO, que plantea un uso centrado en la persona, seguro y ético (Miao & Holmes, 2023), y con el Marco de competencias para estudiantes, que insiste en juicio crítico, responsabilidad y participación creativa (Miao et al., 2024). La discusión muestra que el éxito de la intervención depende de reglas claras y de una cultura evaluativa que valore el proceso, no solo el producto final.

El uso del tema calidad del agua permitió trabajar una ciencia situada. Baptista et al. (2025) muestran que la Educación Científica Basada en Indagación (ECBI) puede favorecer aprendizajes orientados a la acción cuando los estudiantes abordan problemas ambientales locales. Ceccaroni et al. (2023) plantean que la ciencia ciudadana y las escuelas pueden contribuir a los ODS vinculados al agua y la vida submarina. En esta investigación, aunque no se desarrolló un proyecto de ciencia ciudadana completo, la lógica de recolectar, analizar y comunicar datos sobre agua acercó a los estudiantes a prácticas científicas auténticas. Esto fortalece la pertinencia curricular: la ciencia escolar

deja de ser un conjunto de contenidos abstractos y se convierte en una forma de leer críticamente el entorno.

Comparado con estudios sobre tecnologías móviles e indagación, como el de Liu et al. (2021), AquaIA-Gen confirma que las tecnologías digitales son más efectivas cuando están integradas a una estructura de investigación y no cuando se usan como actividad complementaria. También dialoga con Owan et al. (2023), quienes describen el potencial de la inteligencia artificial para apoyar medición y evaluación educativa. En este artículo, la IAG apoyó indirectamente la evaluación formativa al permitir que los estudiantes revisaran sus ideas antes del producto final, aunque el juicio de desempeño siguió dependiendo de un instrumento elaborado y validado por expertos. Esta distinción es clave: la herramienta puede ayudar a aprender, pero la evaluación debe preservar criterios humanos, transparencia y trazabilidad.

Los resultados de productos científicos muestran que los puntajes más altos se ubicaron en tablas, gráficos y conclusiones argumentadas. Esto se relaciona con la idea de que la indagación no termina en observar, sino en transformar observaciones en evidencia comunicable. Pedaste et al. (2021) y Urdanivia Alarcón et al. (2023) coinciden en que las habilidades investigativas incluyen análisis, comunicación y evaluación del proceso. En AquaIA-Gen, la comunicación científica se fortaleció porque los estudiantes recibieron apoyo para organizar ideas, pero debieron justificar cada afirmación con datos. Así, la IAG funcionó como una herramienta para mejorar la claridad, no para reemplazar la autoría. Esta diferencia es crucial para responder a los riesgos de plagio o dependencia que acompañan la introducción de IAG en la escuela.

Finalmente, el estudio aporta una propuesta didáctica transferible, aunque con límites. La muestra fue de 80 participantes y el diseño fue cuasi experimental, por lo que se recomienda replicar la investigación en otras instituciones, grados y contextos rurales o urbanos. También sería pertinente incorporar seguimiento longitudinal para verificar si las destrezas se mantienen sin apoyo de IAG. Además, futuras investigaciones podrían comparar distintos niveles de mediación docente, analizar bitácoras de prompts cualitativamente e incluir indicadores de pensamiento crítico, actitudes ambientales y alfabetización hídrica. A pesar de estas limitaciones, los resultados contrastan favorablemente con la literatura revisada y muestran que la IAG puede ser una estrategia didáctica potente cuando se usa para pensar científicamente, no para evitar pensar.

Conclusiones

El estudio permitió concluir que el uso de la inteligencia artificial generativa, mediante la estrategia didáctica AquaIA-Gen, fortaleció de manera significativa las destrezas de indagación científica sobre calidad del agua en estudiantes de educación secundaria. La contribución principal radica en demostrar que la IAG puede articularse con la competencia de indagación científica cuando se emplea bajo mediación docente, con prompts guiados, verificación de fuentes, análisis de datos y comunicación argumentada. Los resultados modelo mostraron mejoras superiores en el grupo experimental, diferencias estadísticamente significativas y tamaños del efecto altos, especialmente en el diseño de procedimientos, la interpretación de evidencias y la elaboración de conclusiones científicas. Esto evidencia que la IAG no debe ser entendida como un recurso para obtener respuestas automáticas, sino como un andamiaje para hacer más explícito el razonamiento científico escolar.

La investigación aporta una ruta pedagógica para integrar tecnología emergente y educación ambiental desde un problema cercano y científicamente abordable: la calidad del agua. Su valor científico se encuentra en proponer una intervención evaluable, con instrumento estructurado, validación experta, confiabilidad alta y análisis estadístico orientado a relación, diferencia y magnitud del efecto. A nivel didáctico, la propuesta favorece una enseñanza de las ciencias más contextualizada, crítica y participativa, en la que los estudiantes aprenden a formular preguntas, contrastar información, leer evidencias y comunicar explicaciones responsables. Se recomienda aplicar la propuesta con datos empíricos reales, ampliar la muestra, realizar seguimiento longitudinal e incorporar análisis cualitativo de bitácoras de prompts para comprender con mayor profundidad cómo la interacción con IAG transforma el pensamiento científico de los estudiantes.

Referencias

- Araújo, J. L., Morais, C., & Paiva, J. C. (2023). Students' attitudes towards the environment and marine litter in the context of a coastal water quality educational citizen science project. *Australian Journal of Environmental Education*, 39(4), 1–17. <https://doi.org/10.1017/ae.2023.14>
- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting

- teaching and learning. *Journal of AI*, 7(1), 52–62.
<https://doi.org/10.61969/jai.1337500>
- Bakker, A., Cai, J., & Zenger, L. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics*, 102, 1–8. <https://doi.org/10.1007/s10649-019-09908-4>
- Baptista, M., Pinho, A. S., & Alves, A. R. (2025). Students' learning for action through inquiry-based science education on a local environmental problem. *Sustainability*, 17(9), Article 3907. <https://doi.org/10.3390/su17093907>
- Boon, H. J. (2024). A comprehensive approach to water literacy in the context of climate change. *Education Sciences*, 14(6), Article 564. <https://doi.org/10.3390/educsci14060564>
- Ceccaroni, L., Woods, S. M., Butkevičienė, E., Parkinson, S., Sprinks, J., Costa, P., Simis, S. G. H., Lessin, G., Liñán, S., Companys, B., Bonfill, E., & Piera, J. (2023). The role of citizen science in promoting ocean and water literacy in school communities: The ProBleu methodology. *Sustainability*, 15(14), 11410. <https://doi.org/10.3390/su151411410>
- Comisión Económica para América Latina y el Caribe. (2024). Superar las trampas del desarrollo de América Latina y el Caribe en la era digital: El potencial transformador de las tecnologías digitales y la inteligencia artificial. CEPAL. https://conferenciaelac.cepal.org/9/sites/elac9/files/s2401013_es.pdf
- Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32, 444–452. <https://doi.org/10.1007/s10956-023-10039-y>
- 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>
- Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. *Education Sciences*, 13(11), 1109. <https://doi.org/10.3390/educsci13111109>
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review*

- p of Educational Research, 82(3), 300–329.
-
- <https://doi.org/10.3102/0034654312457206>
-
- Kasneci, 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., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education.
- Learning and Individual Differences*
- , 103, 102274.
- <https://doi.org/10.1016/j.lindif.2023.102274>
-
- Kim, H. Y. (2019). Statistical notes for clinical researchers: The independent samples t-test.
- Restorative Dentistry & Endodontics*
- , 44(3), e26.
- <https://doi.org/10.5395/rde.2019.44.e26>
-
- Liu, C., Zowghi, D., Kearney, M., & Bano, M. (2021). Inquiry-based mobile learning in secondary school science education: A systematic review.
- Journal of Computer Assisted Learning*
- , 37(1), 1–23.
- <https://doi.org/10.1111/jcal.12505>
-
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.
- <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
-
- Miao, F., Shiohira, K., & Lao, N. (2024). AI competency framework for students. UNESCO.
- <https://unesdoc.unesco.org/ark:/48223/pf0000391105>
-
- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction: What is it and does it matter?
- Journal of Research in Science Teaching*
- , 47(4), 474–496.
- <https://doi.org/10.1002/tea.20347>
-
- Ministerio de Educación del Perú. (2025). Fascículo de la competencia “Indaga mediante métodos científicos para construir conocimientos”. MINEDU.
- <https://repositorio.minedu.gob.pe/handle/20.500.12799/11847>
-
- Mostacedo-Marasovic, S. J., Forbes, C. T., & Roehrig, G. H. (2022). Towards water literacy: An interdisciplinary analysis of standards for teaching and learning about humans and water.
- Disciplinary and Interdisciplinary Science Education Research*
- , 4, Article 23.
- <https://doi.org/10.1186/s43031-022-00065-y>
-
- Moundridou, M., Matzakos, N., & Doukakis, S. (2024). Generative AI tools as educators’ assistants: Designing and implementing inquiry-based lesson plans.
- Computers and Education: Artificial Intelligence*
- , 7, 100277.
- <https://doi.org/10.1016/j.caeai.2024.100277>

- Ogunleye, B., Zakariyyah, K., Ajao, O., Olayinka, O., & Sharma, H. (2024). A systematic review of generative AI for teaching and learning practice. *Education Sciences*, 14(6), 636. <https://doi.org/10.3390/educsci14060636>
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), em2307. <https://doi.org/10.29333/ejmste/13428>
- Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47–61. <https://doi.org/10.1016/j.edurev.2015.02.003>
- Pedaste, M., Baucal, A., & Reisenbuk, E. (2021). Towards a science inquiry test in primary education: Development of items and scales. *International Journal of STEM Education*, 8, Article 19. <https://doi.org/10.1186/s40594-021-00278-z>
- Savelsbergh, E. R., Prins, G. T., Rietbergen, C., Fechner, S., Vaessen, B. E., Draijer, J. M., & Bakker, A. (2016). Effects of innovative science and mathematics teaching on student attitudes and achievement: A meta-analytic study. *Educational Research Review*, 19, 158–172. <https://doi.org/10.1016/j.edurev.2016.07.003>
- Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the p value is not enough. *Journal of Graduate Medical Education*, 4(3), 279–282. <https://doi.org/10.4300/JGME-D-12-00156.1>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48, 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Urdanivia Alarcón, D. A., Ríos Incio, F. A., & Peralta, R. (2023). Science and inquiry-based teaching and learning: A systematic review. *Frontiers in Education*, 8, 1170487. <https://doi.org/10.3389/feduc.2023.1170487>
- Wang, X., Zainuddin, Z., & Hai Leng, C. (2025). Generative artificial intelligence in pedagogical practices: A systematic review of empirical studies (2022–2024). *Cogent Education*, 12(1), Article 2485499. <https://doi.org/10.1080/2331186X.2025.2485499>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are

the educators? International Journal of Educational Technology in Higher
Education, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés