

Percepción de futuros docentes de matemática sobre la integración de la inteligencia artificial en la evaluación en un curso de tecnologías digitales

Future mathematics teachers' perception of artificial intelligence integration in assessment in a digital technologies course

Alexander Borbón-Alpízar¹

Borbón-Alpízar, A. Percepción de futuros docentes de matemática sobre la integración de la inteligencia artificial en la evaluación en un curso de tecnologías digitales. *Tecnología en Marcha*. Vol. 39 N° especial sobre Inteligencia Artificial. Febrero, 2026. Pág. 123-136.

 <https://doi.org/10.18845/tm.v39i5.8517>

Palabras clave

Inteligencia artificial; evaluación del estudiante; actitud del estudiante; formación de docentes; enseñanza de las matemáticas; rol del docente.

Resumen

La integración de la Inteligencia Artificial (IA) generativa en la educación superior requiere comprender la percepción estudiantil. Este estudio analizó las valoraciones de futuros docentes de matemática (N=14) sobre el uso de ChatGPT como herramienta de evaluación en un curso universitario. Mediante un enfoque mixto, se exploró su percepción sobre la calidad de la retroalimentación y la validez de la calificación generada por IA. Los resultados revelan una dicotomía fundamental: los estudiantes valoran positivamente la IA como herramienta de retroalimentación formativa, reconociendo su utilidad para la mejora. Sin embargo, manifiestan una considerable desconfianza en su rol como agente de evaluación sumativa, al percibirla incapaz de realizar juicios holísticos. El análisis sugiere que los estudiantes ven al docente y a la IA como agentes con fortalezas complementarias, perfilando un modelo de “evaluación aumentada”. La opacidad del proceso de la IA (“caja negra”) se identificó como una barrera crítica. Aunque una mayoría (57.1%) prefiere la calificación humana, una minoría significativa (42.9%) se inclina por la IA, valorando su inmediatez y objetividad. Se concluye que el rol más prometedor para la IA en la evaluación no es el de un juez autónomo, sino el de un copiloto evaluativo que potencia las capacidades del docente.

Keywords

Artificial intelligence; student evaluation; student attitudes; teacher education; mathematics education; teacher role.

Abstract

The integration of generative Artificial Intelligence (AI) in higher education requires understanding student perceptions. This study analyzed the views of pre-service mathematics teachers (N=14) on using ChatGPT as an assessment tool in a university course. Through a mixed-methods approach, their perceptions of the quality of feedback and the validity of AI-generated grades were explored. The findings reveal a fundamental dichotomy: students positively value AI as a formative feedback tool, recognizing its utility for improvement. However, they express significant distrust in its role as a summative assessment agent, perceiving it as incapable of making holistic judgments. The analysis suggests that students view the instructor and AI as agents with complementary strengths, outlining an “augmented assessment” model. The opacity of the AI’s process (the “black box” problem) was identified as a critical barrier. Although a majority (57.1%) prefers human grading, a significant minority (42.9%) leans towards AI, valuing its immediacy and objectivity. It is concluded that the most promising role for AI in assessment is not that of an autonomous judge, but of an assessment copilot that enhances the instructor’s capabilities.

Introducción

En el panorama actual de la educación superior, la integración de las tecnologías digitales ha dejado de ser una opción pedagógica para convertirse en un imperativo estratégico. Específicamente, en la didáctica de la matemática, herramientas computacionales como LaTeX (con paquetes como Beamer y TikZ), Inkscape, y Geogebra, representan un medio para que

los futuros docentes no solo adquieran competencias técnicas, sino que también desarrollen habilidades para la creación de materiales didácticos de alta calidad y profesionalismo. Sin embargo, la reciente irrupción de la inteligencia artificial (IA), particularmente a través de modelos generativos como ChatGPT, ha introducido nuevas dinámicas en el proceso de enseñanza y aprendizaje, planteando interrogantes sobre su rol en la autonomía estudiantil y en los procesos de evaluación.

El presente estudio se centra en explorar las percepciones de los estudiantes de la carrera de “Enseñanza de la Matemática con Entornos Tecnológicos” sobre la integración de la IA en un curso que les enseña a utilizar herramientas computacionales. Específicamente, se analiza la función de la IA como una herramienta para la calificación de las guías de trabajo autónomo. A pesar del creciente interés en la IA para la educación, la literatura académica aún carece de un análisis profundo sobre cómo los estudiantes, y en particular los futuros educadores, perciben esta tecnología cuando interactúan con ella o cuando se utiliza para evaluarlos.

El propósito de este artículo es dilucidar las opiniones de los estudiantes sobre la utilidad, la equidad y la efectividad de la IA en la retroalimentación, contrastando estas percepciones con el papel tradicional del docente. Para alcanzar este fin, el estudio se guía por las siguientes preguntas de investigación:

- PI1: ¿Cómo valoran los futuros docentes de matemática la calidad y utilidad de la retroalimentación formativa generada por la IA frente a la humana?
- PI2: ¿Cuál es el grado de confianza y validez que los estudiantes otorgan a la IA en su rol de agente de evaluación sumativa?
- PI3: ¿En qué medida la opacidad de los procesos de la IA (el problema de la “caja negra”) actúa como una barrera para su aceptación en la evaluación universitaria?

Estado del arte

Introducción al Contexto de la IA en la Educación

La irrupción de la Inteligencia Artificial (IA) en el ámbito educativo no es un fenómeno reciente, sino parte de una evolución constante que busca optimizar los procesos de enseñanza y aprendizaje [1]. Sin embargo, la reciente democratización de modelos de lenguaje grandes (LLM) como ChatGPT ha marcado un punto de inflexión, transitando de herramientas especializadas a asistentes cognitivos de acceso masivo [2]. Este cambio ha generado un intenso debate académico sobre las oportunidades y desafíos que estos sistemas presentan, especialmente en lo que respecta a la evaluación, la autonomía del estudiante y la redefinición del rol docente [3], [4].

Este estudio se sitúa en dicha coyuntura, buscando aportar evidencia empírica sobre la percepción de los estudiantes universitarios en un contexto técnico-matemático, un área donde la precisión y la lógica son fundamentales.

De la Evaluación Automatizada a la Interacción Cognitiva

La investigación sobre la IA en la evaluación tiene una trayectoria considerable. Los primeros esfuerzos se centraron en sistemas de calificación automática de ensayos (Automated Essay Scoring), como el sistema e-rater®, que demostraron la viabilidad de utilizar algoritmos para evaluar textos basándose en criterios predefinidos [5]. Estos sistemas iniciales sentaron las bases para exploraciones más complejas, llevando a estudios comparativos que analizan la consistencia y fiabilidad de la calificación por máquina frente a la humana, encontrando que la IA puede alcanzar altos niveles de concordancia en tareas estructuradas [6].

La llegada de ChatGPT ha revolucionado este campo. A diferencia de los sistemas anteriores, diseñados para una única tarea evaluativa, los LLM ofrecen una interacción dialógica y multifacética. La literatura actual ha comenzado a explorar este potencial, identificando oportunidades para una retroalimentación formativa instantánea y personalizada [7]. No obstante, también ha alertado sobre los desafíos inherentes, como la generación de información incorrecta (“alucinaciones”), los sesgos algorítmicos y el riesgo de socavar el pensamiento crítico si su uso no es mediado pedagógicamente [2], [4].

En el ámbito específico de la educación matemática, la discusión se ha centrado en cómo estas herramientas pueden transformar no solo el aprendizaje del estudiante, sino la formación misma de los docentes, quienes deben estar preparados para guiar a los alumnos en un nuevo ecosistema digital [8]. La literatura actual, por tanto, ha pasado de preguntar si la IA puede evaluar, a cuestionar cómo y para qué fines debe utilizarse, identificando una brecha en la comprensión profunda de la percepción y aceptación de los estudiantes, quienes son los usuarios finales de esta tecnología en el proceso de aprendizaje.

Investigaciones recientes subrayan que la integración de la IA generativa en la educación superior ha transitado de una mera automatización de tareas a un modelo de asociación cognitiva o ‘co-piloto’. No obstante, estudios del 2024 y 2025 como [9], [10] han comenzado a documentar una brecha crítica en la confianza estudiantil: mientras que los alumnos valoran la inmediatez y el volumen de la retroalimentación de la IA para procesos formativos, persiste una preferencia significativa por el juicio humano en instancias sumativas debido a la percepción de que los docentes ofrecen una evaluación más justa, creíble y contextualmente situada. Esta literatura refuerza la idea de que la IA es percibida como un recurso técnico eficiente, pero cuya legitimidad como juez autónomo sigue siendo precaria frente a la figura del educador.

Marco teórico

Para analizar e interpretar los datos de este estudio, se adopta un marco teórico que se sustenta en tres pilares conceptuales derivados de la literatura reciente.

Redefinición de los Roles Pedagógicos en la Era Digital

La integración de la IA en el aula obliga a una redefinición del rol tradicional del docente. De ser la principal fuente de conocimiento, el docente transita hacia un rol de curador, facilitador y guía crítico [3]. Su función se vuelve menos informativa y más transformadora, centrándose en enseñar a los estudiantes a interactuar eficazmente con la tecnología, a evaluar críticamente la información y a aplicar el conocimiento en contextos complejos [11]. Este marco conceptual es fundamental para interpretar los datos relacionados con la preferencia de los estudiantes por la intervención humana en tareas específicas tales como la evaluación sumativa.

El Modelo de Sinergia Humano-IA (“Human-in-the-Loop”)

Frente a la dicotomía entre la inteligencia humana y la artificial, emerge con fuerza el modelo de sinergia o “Human-in-the-Loop”. Este paradigma, apoyado por organismos como la UNESCO [11], postula que los resultados óptimos en entornos complejos se logran no por la sustitución del humano por la máquina, sino por su colaboración estratégica. En educación, esto se traduce en un modelo donde la IA se encarga de tareas de andamiaje, repetitivas o de retroalimentación a gran escala, mientras que el educador humano se reserva para el juicio crítico, la empatía, la contextualización y la inspiración [1]. Este estudio se enmarca directamente en la validación de este modelo, utilizando las percepciones de los estudiantes para determinar si prefieren un modelo de sustitución o uno de colaboración sinérgica en su experiencia de aprendizaje, en particular, en la evaluación del curso.

Marco metodológico

Este estudio adoptó un enfoque de investigación mixto, con un diseño descriptivo-exploratorio. El componente cuantitativo se centró en describir las percepciones estudiantiles mediante estadística descriptiva, mientras que el componente cualitativo, de carácter secundario, emerge del análisis de contenido de las respuestas abiertas para contextualizar los hallazgos numéricos.

4.1 Población y Muestra

La población objeto de estudio fue la totalidad de estudiantes (N=14) del curso “Tecnologías Digitales Aplicadas a la Matemática Educativa I” de la carrera de “Enseñanza de la Matemática con Entornos Tecnológicos” durante el II Semestre de 2025 del “Instituto Tecnológico de Costa Rica”, todos futuros docentes de matemática. Se trabajó con la población completa del curso, que a su vez constituye una muestra por conveniencia de los estudiantes de la carrera.

4.2 Instrumento de Recolección de Datos

Se diseñó un cuestionario *ad hoc*, administrado en formato digital, para evaluar las percepciones estudiantiles sobre el uso de la Inteligencia Artificial (IA) en procesos de evaluación. El instrumento se estructuró en tres secciones, si bien este artículo reporta únicamente los resultados de la primera:

1. Percepción sobre la evaluación realizada por IA: 14 ítems tipo Likert y una pregunta de elección forzada para medir constructos como satisfacción, transparencia y justicia.
2. Impacto de la IA como apoyo en guías de trabajo: 13 ítems Likert y una pregunta de elección forzada sobre motivación y autoestudio.
3. Comentarios adicionales: una única pregunta abierta.
4. Todos los ítems de escala utilizaron un formato Likert de cinco puntos (1 = Totalmente en desacuerdo; 5 = Totalmente de acuerdo).

4.3 Procedimiento

El estudio se desarrolló en el curso mencionado, donde los estudiantes completaron 16 guías de trabajo autónomo (e.g., LaTeX, Inkscape) con un doble apoyo de IA. Para la resolución de las guías, los estudiantes emplearon libremente modelos generativos como ChatGPT y Gemini. Para la evaluación, el equipo docente utilizó la API de ChatGPT (GPT-4o para las guías 1-4 y GPT-5 para las guías 5-16) mediante un *prompt* estandarizado que asignaba a la IA el rol de evaluador, generando retroalimentación detallada basada en una rúbrica.

La recolección de datos se realizó a mitad del semestre, una vez que los estudiantes completaron la totalidad de las guías. Se obtuvo el consentimiento informado de los participantes, garantizando el anonimato y la voluntariedad de su participación.

4.4 Análisis de Datos

Se aplicó un análisis de métodos mixtos. Primero, se evaluó la consistencia interna de los ítems Likert mediante el Alfa de Cronbach, lo que guió su agrupación en factores o su análisis individual. Pese a la limitación del coeficiente por el tamaño muestral (N=14), se empleó como criterio para estructurar los resultados. Posteriormente, los datos cuantitativos se analizaron con estadística descriptiva (medias, desviación estándar y frecuencias). Finalmente, las respuestas abiertas fueron objeto de un análisis de contenido cualitativo para identificar temas emergentes. Los hallazgos se presentan de forma integrada, triangulando los datos cuantitativos y cualitativos para una mayor robustez.

Limitaciones del estudio

Como toda investigación de carácter descriptivo-exploratorio, este estudio presenta ciertas limitaciones que deben ser consideradas al interpretar los hallazgos. La principal limitación metodológica radica en el tamaño de la muestra ($N=14$), la cual corresponde a la totalidad de los estudiantes de un curso específico en el Instituto Tecnológico de Costa Rica.

Al tratarse de una muestra por conveniencia, los resultados obtenidos no pretenden ser generalizables a otras poblaciones estudiantiles o áreas disciplinares fuera del ámbito técnico-matemático. No obstante, este enfoque permitió un análisis profundo y contextualizado de las percepciones de futuros docentes en un entorno de aprendizaje especializado. Asimismo, el carácter incipiente de la tecnología evaluada (modelos como GPT-4o y GPT-5) implica que las percepciones aquí reportadas podrían evolucionar rápidamente conforme aumente la transparencia de los sistemas y la alfabetización digital de los participantes.

A pesar de estas limitaciones, la homogeneidad de la muestra (estudiantes de una carrera altamente técnica) otorga una validez interna significativa para comprender la adopción tecnológica en la formación de docentes de matemática.

Resultados

A continuación, se presentan los hallazgos sobre la percepción estudiantil de la evaluación realizada por IA. Para el análisis, los ítems enunciados en negativo (A5, A6, A11, A12, A13) fueron recodificados. La nomenclatura y texto completo de cada ítem (A1-A14) se detallan en el cuadro 1.

Cuadro 1. Nomenclatura para los ítems del estudio.

Nomenclatura	Texto completo
A1	La calificación de ChatGPT fue coherente entre estudiantes con trabajos similares.
A2	La retroalimentación de ChatGPT fue clara y fácil de entender.
A3	La retroalimentación de ChatGPT me indicó acciones concretas para mejorar mi guía.
A4	ChatGPT identificó con precisión errores técnicos (p. ej., sintaxis LaTeX, estructura TikZ, capas en Inkscape).
A5	La calificación de ChatGPT no reflejó el esfuerzo o complejidad de mi trabajo.
A6	ChatGPT pasó por alto detalles importantes del formato o de los criterios de la rúbrica.
A7	Prefiero la claridad de la retroalimentación de ChatGPT frente a la de un/a calificador/a humano/a.
A8	ChatGPT es más consistente que un/a humano/a al aplicar la misma rúbrica.
A9	ChatGPT detecta mejor problemas técnicos puntuales (compilación, paquetes, estilos) que un/a humano/a.
A10	La percepción de imparcialidad (sin sesgos personales) es mayor con ChatGPT que con un/a humano/a.
A11	Un/a calificador/a humano/a es más justo/a que ChatGPT en mis guías.
A12	Para valorar la calidad estética (diagramación Beamer, diseño en Inkscape), un/a humano/a lo hace mejor que ChatGPT.
A13	La forma en que ChatGPT califica es opaca; no entiendo por qué asigna ciertas observaciones.
A14	La calificación de ChatGPT estuvo alineada con la rúbrica del curso (criterios y ponderaciones).

A modo de panorama general, la Figura 1 presenta la distribución de las respuestas originales para cada ítem, antes de la recodificación mencionada.

5. Sobre la calificación que realiza el chatGPT

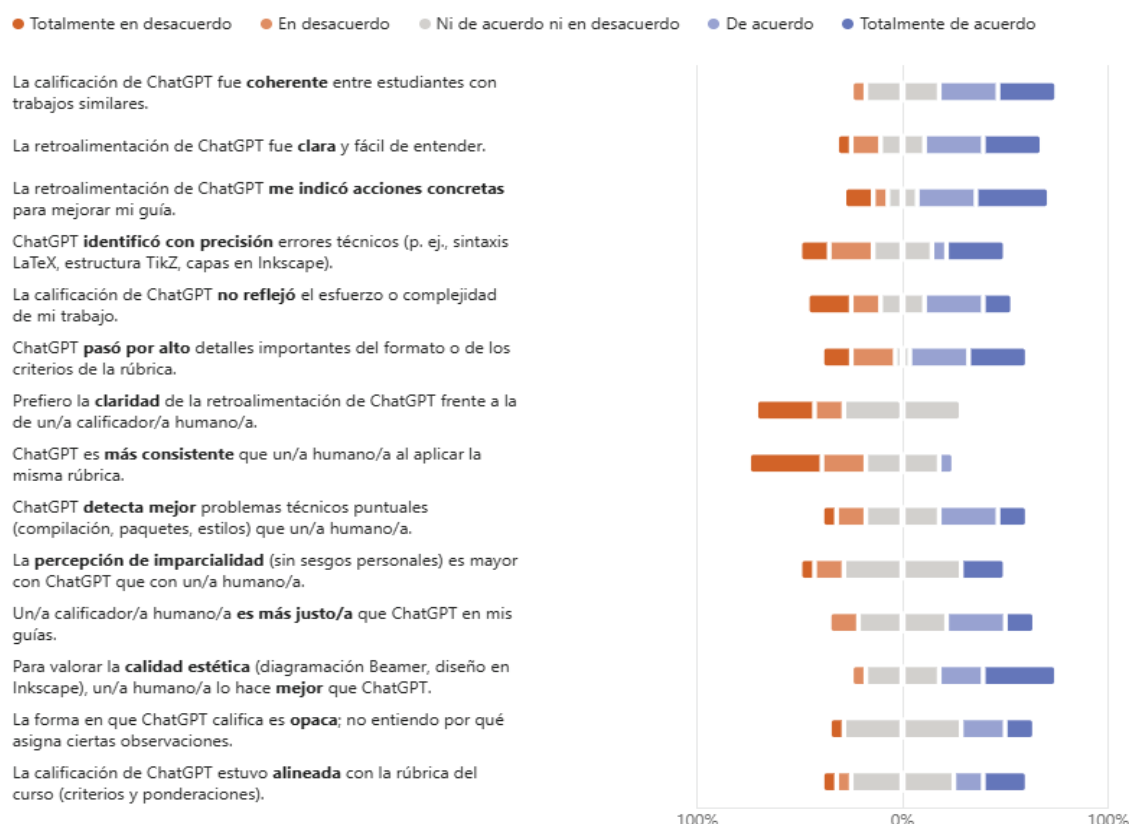


Figura 1. Distribución de frecuencias de las respuestas para los 14 ítems de la primera sección. Nota: Los resultados se muestran con los valores originales, sin invertir la escala de los ítems redactados en negativo (A5, A6, A11, A12, A13).

A continuación, se procede con el análisis detallado de estos resultados, agrupados según los constructos psicométricos definidos en la construcción del instrumento.

Dicotomía entre Utilidad Formativa y Validez Sumativa

Los ítems A1-A6, diseñados para medir la percepción general de la evaluación, mostraron una fiabilidad inicial marginal ($\alpha = 0.670$). Un Análisis de Componentes Principales (PCA) reveló una estructura bidimensional que explica un 71.3% de la varianza, distinguiendo claramente entre dos constructos:

1. Utilidad Formativa (A1-A4; $\alpha = 0.785$): Este componente fue valorado favorablemente ($M=3.54$, $DE=1.29$), indicando que los estudiantes consideran la retroalimentación de la IA coherente, clara y útil para la mejora.
2. Limitaciones Sumativas (A5-A6; $\alpha = 0.815$): En contraste, este componente arrojó una percepción negativa ($M=2.82$, $DE=1.44$). Los estudiantes se mostraron escépticos sobre la capacidad de la IA para reflejar el esfuerzo (A5) u observar detalles importantes de la rúbrica (A6).

Estos resultados exponen una dicotomía fundamental: los estudiantes valoran a la IA como una eficaz herramienta de retroalimentación formativa, pero desconfían de su validez para ejercer un juicio sumativo final, el cual perciben como carente de una comprensión holística.

Comparación entre Evaluación por IA y Humana: Un Modelo Aumentado

El análisis comparativo (ítems A7-A12) arrojó una consistencia interna muy baja ($\alpha = 0.367$), lo que confirmó que no miden un constructo único. Un PCA posterior reveló una percepción matizada donde los estudiantes atribuyen fortalezas complementarias a cada tipo de evaluador.

- **Preferencia por el Juicio Humano:** Se identificó una fuerte preferencia por el evaluador humano en las dimensiones nucleares de la evaluación. Los estudiantes lo consideran significativamente superior en claridad y consistencia (ítems A7-A8; $M=2.21$, $DE=0.96$; $\alpha = 0.868$) y en la capacidad de valorar la calidad estética (A12, recodificado; $M=2.14$, $DE=1.03$).
- **Ventajas y Ambivalencia de la IA:** La percepción de que la IA es más imparcial (A10) fue ambivalente ($M=3.14$, $DE=1.17$), sin un consenso claro. Su única ventaja percibida, aunque moderada, fue en la detección de errores técnicos puntuales (A9).

Estos hallazgos no abogan por la sustitución del docente, sino por un modelo de evaluación aumentada. El alumnado percibe un flujo de trabajo ideal donde la IA ofrece una primera capa de feedback objetivo y técnico, mientras que el juicio complejo, cualitativo y final permanece como dominio del evaluador humano.

Transparencia como Barrera Crítica

El constructo sobre la Transparencia y Validez del Proceso (A13-A14; $\alpha = 0.732$) identificó la raíz de la desconfianza estudiantil. Aunque los estudiantes se mostraron ambivalentes sobre si la calificación de la IA se alineaba con la rúbrica (A14: $M=3.36$, $DE=1.15$), fueron claros al señalar que su proceso de calificación es opaco (A13, recodificado; $M=2.64$, $DE=1.01$). Esta percepción del problema de la “caja negra” (black box) se confirma como un obstáculo fundamental para la aceptación de la IA en roles evaluativos de alto impacto.

Preferencia General de Calificación

Para culminar el análisis, se planteó a los estudiantes una pregunta de elección forzada que buscaba sintetizar su postura global sobre la evaluación: “*Con respecto a la calificación de los trabajos del curso: Prefiero que me califique ChatGPT / Prefiero que me califique el docente*”. Los resultados se muestran en la Figura 2.

6. Con respecto a la calificación de los trabajos el curso



Figura 2. Resultado de la pregunta cerrada de la primera parte.

Este resultado encapsula la tensión que se muestra en el estudio. Los resultados mostraron una división: un 57.1% (8 estudiantes) prefirió al docente, mientras que un significativo 42.9% (6 estudiantes) se inclinó por ChatGPT.

Esta división no es contradictoria. La mayoría que prefiere al docente prioriza la confianza, el juicio matizado y la transparencia. Por otro lado, la considerable minoría que opta por la IA probablemente valora atributos como la inmediatez del feedback, la disponibilidad 24/7 y la objetividad en tareas técnicas. Este resultado refuerza la conclusión de que los modelos pedagógicos más efectivos no serán de sustitución, sino de integración híbrida, capitalizando las fortalezas únicas tanto del docente como de la inteligencia artificial.

Conclusiones y recomendaciones

El presente estudio exploró las percepciones de futuros docentes de matemática sobre la integración de la IA generativa en la evaluación técnica. Los hallazgos revelan una relación compleja que se aleja de una aceptación monolítica, permitiendo derivar conclusiones sobre el estado actual de esta tecnología y plantear recomendaciones prospectivas para su evolución pedagógica.

Conclusiones derivadas del estudio

A partir del análisis de los datos recolectados, se presentan las siguientes conclusiones fundamentales:

1. Dicotomía Funcional entre Retroalimentación y Calificación:

La conclusión más robusta de este estudio es la clara distinción conceptual que los estudiantes realizan entre el rol formativo y el rol sumativo de la evaluación. La IA fue valorada positivamente como una herramienta de retroalimentación formativa, destacando su capacidad para ofrecer comentarios claros, coherentes y accionables que facilitan el andamiaje cognitivo y la mejora continua. Sin embargo, esta percepción favorable se invierte cuando la IA asume el papel de agente evaluativo sumativo. En este dominio, emerge una considerable desconfianza, fundamentada en la percepción de que la IA es incapaz de realizar un juicio holístico que valore adecuadamente el esfuerzo, la complejidad o las sutilezas estéticas y conceptuales de un trabajo. Los estudiantes no la consideran un juez final válido.

2. Hacia un Modelo de Evaluación Aumentada por IA:

Los resultados refutan la idea de una competencia entre el evaluador humano y la IA. En su lugar, perfilan un modelo de simbiosis o “Evaluación Aumentada por IA”. Los estudiantes atribuyen fortalezas complementarias y casi excluyentes a cada uno: mientras que el docente es percibido como el estándar de oro en claridad, consistencia, justicia y juicio cualitativo, la IA es reconocida (aunque no de forma unánime) por su potencial en imparcialidad y en la detección de errores técnicos específicos. Este panorama sugiere que el futuro de la evaluación no yace en la automatización total, sino en la creación de flujos de trabajo híbridos donde la IA actúe como un asistente que potencie las capacidades del docente, liberándolo de tareas repetitivas para que pueda centrarse en la retroalimentación de alto nivel.

3. La Transparencia como Barrera Crítica para la Aceptación:

El escepticismo estudiantil hacia la evaluación sumativa de la IA no es un prejuicio abstracto, sino que se ancla en un problema fundamental: la opacidad de su proceso de razonamiento. La opacidad del proceso algorítmico no es solo un fallo técnico, sino un impedimento pedagógico que limita la capacidad del estudiante para aprender del error a través de la lógica de la calificación. La percepción de que la IA opera como una “caja negra” (black box), cuyo juicio no es enteramente comprensible, se erige como la

principal barrera para su aceptación en roles de alto impacto. Sin una comprensión clara de por qué la herramienta asigna una calificación o emite una observación, la confianza necesaria para validar su juicio se ve socavada. Esta falta de transparencia es, por tanto, un desafío pedagógico y técnico de primer orden que debe ser abordado para cualquier implementación exitosa.

Recomendaciones y prospectiva

Basándose en las conclusiones anteriores, se proponen las siguientes líneas de acción para la práctica y la investigación futura:

Para la Práctica Docente:

- Implementar un Enfoque de “IA como Copiloto”: Se recomienda utilizar la IA generativa principalmente para tareas de retroalimentación formativa de bajo impacto. Por ejemplo, para revisiones de borradores, chequeos de sintaxis de código (como en LaTeX), o para generar ejemplos y sugerencias de mejora que los estudiantes puedan aceptar o rechazar críticamente.
- Reservar el Juicio Sumativo al Docente: La calificación final, especialmente en tareas complejas que involucren creatividad, argumentación o diseño, debe permanecer como una prerrogativa del evaluador humano. El docente puede, no obstante, utilizar los informes de la IA como un insumo más para su propia evaluación, validando, corrigiendo o enriqueciendo sus observaciones.
- Fomentar la Transparencia en el Uso: Cuando se utilice la IA en la evaluación, es crucial explicar a los estudiantes cómo funciona el sistema, qué *prompts* se utilizan y cuáles son sus limitaciones. Hacer explícito que la IA es una herramienta de apoyo y no el juez final puede mitigar la desconfianza y la percepción de opacidad.

Para el Diseño Instruccional y Curricular:

- Desarrollar la Competencia de “Alfabetización en IA”: Los programas de formación de docentes, y de la educación superior en general, deben incluir explícitamente el desarrollo de competencias para interactuar críticamente con la IA. Esto incluye no solo saber usar la herramienta, sino también aprender a diseñar *prompts* efectivos (*prompt engineering*), a evaluar la calidad y fiabilidad de sus resultados, y a comprender sus sesgos y limitaciones éticas.
- Diseñar Evaluaciones Resistentes a la IA y que la Integren: Se debe repensar el diseño de las tareas evaluativas. En lugar de prohibir el uso de la IA, se pueden diseñar actividades que requieran que los estudiantes la utilicen como herramienta para luego analizar, criticar o mejorar el producto generado, enfocando la evaluación en estos procesos cognitivos de orden superior.

Para Futuras Investigaciones:

- Ampliar y Diversificar la Muestra: Es imperativo replicar este estudio con muestras más grandes y diversas, incluyendo estudiantes de diferentes disciplinas (humanidades, ciencias sociales, etc.) para determinar si estas percepciones son específicas del ámbito técnico-matemático o si son generalizables.
- Realizar Estudios Longitudinales: Investigar cómo evolucionan las percepciones de los estudiantes a lo largo del tiempo, con una exposición más prolongada y variada a diferentes modelos de IA, es un paso necesario para comprender los procesos de adaptación y aceptación tecnológica.

- Analizar la Calidad Objetiva del Feedback: Mientras este estudio se centró en la *percepción* del feedback, futuras investigaciones deberían analizar objetivamente la calidad, precisión y utilidad del feedback generado por diferentes modelos de IA (ej. GPT-4o vs. GPT-5 vs. Gemini) en comparación con el feedback de expertos humanos.

En síntesis, este estudio concluye que la Inteligencia Artificial se perfila no como un sustituto del docente evaluador, sino como un potente copiloto evaluativo. Su valor reside en su capacidad para potenciar el proceso formativo, ofreciendo retroalimentación inmediata y escalable. Sin embargo, su legitimidad como agente de juicio final es aún precaria y depende de que se resuelvan los desafíos fundamentales de transparencia y comprensión holística. El camino por seguir es uno de integración crítica y reflexiva, donde la tecnología sirva para aumentar, y no para reemplazar, la insustituible labor pedagógica del ser humano.

Agradecimientos

A la red académica MAT-IA de la Escuela de Matemática del Instituto Tecnológico de Costa Rica.

Referencias

- [1] I. Roll and R. Wylie, "Evolution and revolution in artificial intelligence in education," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 582-599, 2016. doi: 10.1007/s40593-016-0110-3.
- [2] E. Kasneci *et al.*, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, Art. no. 102274, 2023. doi: 10.1016/j.lindif.2023.102274.
- [3] R. Rondon *et al.*, "El Impacto de la Inteligencia Artificial en la Formación Docente," *Revista Tecnológica-Educativa Docentes 2.0*, vol. 17, no. 2, pp. 368-375, 2024. doi: 10.37843/rted.v17i2.566.
- [4] J. Rudolph, S. Tan, and S. Tan, "ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?," *Journal of Applied Learning and Teaching*, vol. 6, no. 1, 2023. doi: 10.37074/jalt.2023.6.1.9.
- [5] Y. Attali and J. Burstein, "Automated essay scoring with e-rater® V.2," *The Journal of Technology, Learning, and Assessment*, vol. 4, no. 3, pp. 1-30, 2006.
- [6] K. Bouziane and A. Bouziane, "AI versus human effectiveness in essay evaluation," *Discovery Education*, vol. 3, no. 201, 2024. doi: 10.1007/s44217-024-00320-6.
- [7] X. Zhai, "ChatGPT for next generation science learning," *ACM*, vol. 29, no. 3, pp. 42-46, 2023. doi: 10.1145/3589649.
- [8] A. Clark-Wilson, O. Robutti, and N. Sinclair, *The Mathematics Teacher in the Digital Era: International Research on Professional Learning and Practice*. Springer Nature, 2023.
- [9] M. Bobula, "Generative artificial intelligence (AI) in higher education: a comprehensive review of challenges, opportunities, and implications," *Journal of Learning Development in Higher Education*, no. 30, pp. 1-27, 2024. doi: 10.47408/jldhe.vi30.1137.
- [10] M. Henderson *et al.*, "Comparing Generative AI and teacher feedback: student perceptions of usefulness and trustworthiness," *Assessment & Evaluation in Higher Education*, vol. 50, no. 1, pp. 1-16, 2025. doi: 10.1080/02602938.2025.2502582.
- [11] F. Pedro, M. Subosa, A. Rivas, and P. Valverde, *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*. París, Francia: UNESCO, 2019.

Anexos

Anexo A: Ejemplo de una guía de trabajo con apoyo de la IA

Guía de trabajo – Clase 3: Escritura de contenido matemático en LaTeX

Curso: Tecnologías Digitales Aplicadas a la Matemática Educativa I

Duración: 2 horas

Modalidad: Presencial / con uso de computadora

Objetivos de la clase

- Escribir expresiones matemáticas en LaTeX usando entornos en línea y en bloque.
- Usar el paquete amsmath para mejorar la presentación de fórmulas.
- Consultar a una IA para explorar la notación matemática avanzada y resolver dudas de sintaxis.

Herramientas requeridas

- TeXstudio (o un navegador con acceso a Overleaf <https://www.overleaf.com/>)
- Acceso a IA (ChatGPT o similar).
- Acceso a internet.
- Esta guía impresa o digital.

Actividades prácticas

• **Actividad 1 – Matemáticas en línea y en bloque (35 min)**

1. Abre el documento anterior o realiza un documento nuevo en TeXstudio (o tu proyecto en Overleaf) y crea una nueva sección llamada 'Contenido matemático'.
2. Consulta a la IA cómo se escriben expresiones matemáticas en línea y en bloque en LaTeX.
3. Escribe al menos tres expresiones en línea, usando $\$...\$$ (ej. $E = mc^2$).
4. Escribe al menos tres expresiones en bloque usando $\[... \]$ o el entorno equation.
5. Incluye etiquetas y referencias con $\backslash label\{$ y $\backslash ref\{$ en al menos una ecuación.
6. Documenta una de las preguntas hechas a la IA, el código sugerido y el resultado aplicado.

• **Actividad 2 – Uso del paquete amsmath (30 min)**

1. Pregunta a la IA cómo usar el paquete amsmath y sus entornos más comunes.
2. Incluye $\backslash usepackage\{amsmath\}$ en el preámbulo.
3. Escribe una fórmula larga utilizando align:

```
\begin{align}
f(x) &= x^2 + 3x + 2 \\
&= (x+1)(x+2)
\end{align}
```

4. Crea un sistema de ecuaciones usando cases:

```
f(x) = \begin{cases}
x^2 & \text{si } x \geq 0 \\
-x & \text{si } x < 0
\end{cases}
```

5. Documenta una de las interacciones con la IA.

• **Actividad 3 – Consultas adicionales con IA (30 min)**

1. Utiliza una IA para:

- Consultar símbolos especiales como `\therefore`, `\angle`, `\infty`, `\approx`, `\nabla`.
- Preguntar cómo numerar solo algunas ecuaciones o cómo alinear sin numerar.
- Explorar cómo insertar matrices con el entorno matrix.

2. Escribe en el documento ejemplos del uso de los comandos investigados con la IA.

Producto esperado

- Archivo **.tex** generado en TeXstudio con al menos:

- 3 expresiones en línea
- 3 expresiones en bloque
- Una fórmula alineada
- Un sistema con cases

- Registro breve de interacción con IA.

El archivo se debe subir en el tecDigital en Evaluaciones-Edición Texto-S2C1.

Evaluación

Criterio	Nivel logrado		
Expresiones matemáticas en línea y bloque			
Uso de amsmath (align, cases)			
Interacción significativa con IA			
Claridad del documento y organización			

Anexo B: Ejemplo de una evaluación realizada por IA

Revisión según rúbrica – Clase 3

3. Expresiones matemáticas en línea y en bloque 

Incluye más de **tres expresiones en línea** ($2x$, x^3 , $18i$, etc.) y varias en bloque ($[\dots]$, equation). Usa `\label` y `\ref` correctamente en la ecuación de Euler. Cumple lo solicitado.

Nivel:  (25/25)

4. Uso de amsmath (align, cases) 

El preámbulo contiene `\usepackage{amsmath}`. Hay un ejemplo correcto en align (transformada discreta de Fourier), pero el sistema de ecuaciones no está bien implementado: usa $\begin{cases} \dots \end{cases}$ con `&=` dentro, lo que no es válido en cases (ese entorno no acepta `=` alineado así). Debió usar `\begin{cases} \dots \end{cases}` dentro de equation o $[\dots]$ y con formato de funciones a trozos o ecuaciones sin `&=`.

Nivel:  (20/25)

5. Interacción significativa con IA 

Menciona consultas a IA (Gémini) sobre fórmulas importantes, uso de align y cases, y símbolos especiales. Sin embargo, no presenta el código sugerido por la IA ni describe el resultado aplicado paso a paso, como sí lo pide la guía.

Nivel: ● (20/25)

6. Claridad del documento y organización ●

Documento limpio, con secciones y subsecciones, fórmulas bien separadas y código ordenado. Uso adecuado de etiquetas y referencias.

Nivel: ● (25/25)

Nota final: 90 / 100

Observación general: Buen trabajo con variedad de expresiones y ejemplos. Cumple lo esencial, pero el entorno cases no se implementó correctamente y la interacción con IA podría ser más detallada, incluyendo el código sugerido y el análisis del resultado.

Declaración sobre uso de Inteligencia Artificial (IA)

Los autores declaramos que hemos utilizado una herramienta de inteligencia artificial *Gemini* para asistirnos en la redacción de este artículo. Esta herramienta nos ayudó a mejorar la estructura y la claridad del texto. Los contenidos generados por la IA fueron revisados minuciosamente por nosotros para asegurar su precisión y coherencia con el objetivo del estudio. Además, empleamos la herramienta de inteligencia artificial para analizar los datos recopilados. Esta herramienta nos permitió identificar patrones y tendencias que de otro modo podrían haber pasado desapercibidos. Nos aseguramos de validar los resultados obtenidos con otros métodos de análisis para evitar sesgos. Utilizamos la herramienta de inteligencia artificial para traducir el resumen de este artículo del español al inglés. La herramienta nos ayudó a agilizar el proceso de traducción, pero realizamos una revisión exhaustiva para asegurar la calidad y precisión de las traducciones. Por último, se utilizó para la revisión gramatical y ortográfica de este artículo. Esta nos permitió identificar errores y mejorar la fluidez del texto. No obstante, realizamos una revisión final para garantizar que el artículo cumpliera con los estándares de calidad de la revista.