

PETICIONES Y VALORACIONES DE LA FORMULACIÓN DE PROBLEMAS ASISTIDA POR INTELIGENCIA ARTIFICIAL GENERATIVA

Sara Embid ¹ (sembidso@ull.edu.es)

Josefa Perdomo-Díaz¹ (jperdomd@ull.edu.es)

¹Universidad de La Laguna

Resumen

Este estudio analiza la interacción de investigadores en didáctica de la matemática con Perplexity, una herramienta basada en inteligencia artificial generativa (IAG). Se persigue identificar las características que los investigadores incluyen en sus peticiones a la IAG, su grado de satisfacción con las respuestas obtenidas y los criterios sobre los que se apoya su valoración. Participaron once parejas de investigadores que usaron Perplexity para formular problemas matemáticos abiertos para primaria, en cuatro fases: formulación, valoración, reformulación y nueva valoración. Las características incluidas fueron: apertura del problema, área de conocimiento, público objetivo, realismo del contexto y redacción. La satisfacción media pasó de 6,45 a 7,27 (en una escala de 1 a 10) tras la reformulación, con mejoras en la adecuación al área de conocimiento y la redacción, y deficiencias en el realismo del contexto y la apertura del problema. Los resultados indican que la IAG puede actuar como un “espejo didáctico”, ayudando a generar problemas alineados con criterios de calidad matemática y pedagógica.

Palabras clave: Formulación de problemas, Inteligencia artificial, Evaluación, Problemas matemáticos abiertos.

Abstract

This study analyzes the interaction of researchers in mathematics education with Perplexity, a tool based on generative artificial intelligence (GenAI). The aim is to identify the characteristics that researchers include in their requests to the GenAI, their degree of satisfaction with the responses obtained, and the criteria on which their evaluation is based. Eleven pairs of researchers participated, using Perplexity to pose open mathematics problems for primary education in four phases: formulation, evaluation, reformulation, and re-evaluation. The included characteristics were: openness of the problem, area of knowledge, target audience, realism of the context, and wording. The average satisfaction increased from 6.45 to 7.27 (on a scale from 1 to 10) after the reformulation, with improvements in the adequacy of the area of knowledge and the wording, and deficiencies in the realism of the context and the openness of the problem. The results indicate that GAI can act as a “didactic mirror,” helping to generate problems aligned with mathematical and pedagogical quality criteria.

Keywords: Problem Posing, Artificial intelligence, Evaluation, Open mathematical problems.

Introducción

La formulación y reformulación de problemas matemáticos constituye un eje central en la educación matemática, con profundas repercusiones sociales, políticas y educativas (Cai y Hwang, 2020). Desde la perspectiva social, esta actividad permite formar ciudadanos críticos, capaces de analizar y transformar su entorno (NCTM, 2014). En el ámbito político, su presencia en los currículos responde a la necesidad de una educación matemática actualizada (Cai y Hwang, 2020; NCTM, 2014). A nivel educativo, formular problemas matemáticos abre nuevas vías para potenciar habilidades como la creatividad, el razonamiento y la resolución de problemas (Silver, 1994; Singer et al., 2013).

Numerosos docentes recurren a los libros de texto como fuente de problemas bien formulados (Son y Diletti, 2017); sin embargo, la irrupción reciente de asistentes conversacionales basados en inteligencia artificial generativa como ChatGPT, Perplexity y NotebookLM (Torres-Salinas y Arroyo Machado, 2025) está empezando a transformar esta práctica. Estos asistentes no solo permiten generar enunciados con características específicas, sino que también facilitan la evaluación y reformulación de problemas, enriqueciendo así los enfoques tradicionales apoyados en libros de texto (Embid y Perdomo-Díaz, 2024; Pochulu y Font, 2024). Además, su uso puede contribuir a la personalización del aprendizaje, al ofrecer retroalimentación inmediata y favorecer entornos educativos más inclusivos (Looi y Jia, 2025).

El potencial que ofrecen estas herramientas es tanto y la velocidad a la que se producen cambios es de tal magnitud, que continuamente emergen nuevos interrogantes que requieren ser abordados desde la investigación. En esta propuesta, nos centramos en el estudio de la interacción entre los asistentes conversacionales y sus usuarios. En concreto, el objetivo general es explorar cómo investigadores en didáctica de la matemática interactúan con la herramienta Perplexity para formular problemas matemáticos abiertos dirigidos a estudiantes de educación primaria. Esto permitirá reflexionar sobre qué aprendizajes pueden extraerse de las experiencias de los investigadores para enriquecer la formación docente y también para incorporar la formulación de problemas como una actividad matemática más en el aula.

Marco conceptual

Formulación de problemas abiertos

La formulación de problemas es una actividad matemática que puede partir de situaciones diversas que van desde modificar un problema dado a crear un problema desde el inicio, sin basarse en otro problema previo (Cai y Hwang, 2020; Silver, 1994). Como actividad matemática, la formulación de problemas puede ser realizada por cualquier individuo. En el caso del profesorado, la formulación de problemas constituye, además de una actividad matemática, una práctica profesional inherente a su labor docente.

Abordar la formulación de problemas exige, en primer lugar, concretar qué se entiende por “problema”. Como advierten Baumanns y Rott (2022), en la literatura científica se encuentran estudios que consideran cualquier actividad matemática como un problema y otros que ligan este término al hecho de que la actividad se presente en un contexto no matemático. En otra dirección van autores como Schoenfeld (1985), quien reserva el término problema para aquella actividad cuyo esquema de resolución no es evidente a primera vista, o De Guzmán (1993) quien sostiene que los problemas se dan cuando no se conoce el camino que puede llevarnos de la situación dada a su resolución. Tal y como Blanco y Pino (2015) advierten, una misma tarea puede considerarse un problema en un momento determinado, pero dejar de serlo una vez comprendido su procedimiento de resolución.

Los problemas matemáticos pueden ser de distintos tipos y en la literatura especializada se han propuesto diversas tipologías para clasificarlos. Entre ellas, la de Fan y Zhu (2006) resulta especialmente amplia, distinguiendo entre problemas rutinarios y no rutinarios, tradicionales y no tradicionales, aplicados y no aplicados, de un paso y de múltiples pasos, con datos suficientes, superfluos o insuficientes, problemas verbales, visuales, simbólicos o combinados, así como problemas abiertos y cerrados.

Los problemas cerrados se caracterizan por tener una única solución correcta (situación final cerrada) y, generalmente, un camino de resolución bien definido (a partir de una situación inicial cerrada). En cambio, un problema abierto, en el sentido de Pehkonen (1997), puede presentar apertura tanto en la situación inicial como en la final. Los problemas con situación inicial abierta son aquellos en los que el punto de partida no está completamente definido, lo que permite al estudiante elegir datos, condiciones o incluso formular el enunciado. Ejemplos de este tipo son los proyectos de investigación, la modelización matemática o la propia formulación de problemas, donde el planteamiento inicial queda abierto a la interpretación del alumno. Por su parte, los problemas con situación final abierta (o de final abierto) son tareas que admiten múltiples soluciones o distintos caminos de resolución. Un ejemplo sería pedir a los estudiantes que encuentren todas las formas posibles de construir un rectángulo con un perímetro dado, en lugar de calcular el perímetro de un rectángulo concreto.

La calidad de un problema depende de diversos factores. Por ejemplo, Leavy y Hourigan (2022) consideran aspectos como la motivación del contexto, la claridad del lenguaje y accesibilidad cultural, la coherencia curricular, el número de pasos para su resolución, el número de soluciones, la demanda cognitiva, la variedad de estrategias de resolución y las oportunidades de aprendizaje. Cankoy y Özder (2017) tienen en cuenta otros indicadores, como la razonabilidad (es decir, si los datos tienen sentido en el contexto planteado) y la estructura matemática subyacente en los procesos de resolución.

Nuestro estudio se centra en los problemas de final abierto y en el análisis de la interacción entre investigadores en educación matemática y herramientas basadas en inteligencia artificial generativa durante una actividad de formulación de este tipo de problemas.

Herramientas basadas en inteligencia artificial generativa

La inteligencia artificial, entendida como la capacidad de una máquina para realizar tareas habitualmente asociadas a la inteligencia humana, tiene una presencia cada vez mayor en el ámbito educativo (Wardat et al., 2024). En particular, destaca la inteligencia artificial generativa (IAG), caracterizada por su capacidad para producir contenido.

Una de las herramientas más extendida en este ámbito son los chatbots, asistentes conversacionales diseñados para mantener diálogos de forma intuitiva (Kuhail et al., 2023). Se puede distinguir entre dos tipos de chatbots: los tradicionales, que funcionan a partir de reglas predefinidas y solo pueden responder a preguntas programadas previamente; y los chatbots basados en IA generativa que utilizan modelos avanzados de lenguaje para adaptarse a la conversación y generar respuestas personalizadas para cada usuario e interacción (Terblanche, 2024). En ambos casos, los usuarios formulan peticiones o prompts, en los que especifican la tarea que desean que realice la herramienta, y emiten valoraciones (no necesariamente públicas), mediante las cuales juzgan hasta qué punto las respuestas se ajustan a sus expectativas personales o profesionales.

En relación con los chatbots basados en IAG, encontramos asistentes con distintas especificaciones (Area-Moreira et al., 2024; García-Peñalvo et al., 2024). Algunos, como Perplexity, tienen integrado un motor de búsqueda basado en IA y conexión con internet, otros,

en cambio, usan modelos desconectados o locales basados en fuentes proporcionadas exclusivamente por el usuario, como Notebook LM (Torres-Salinas y Arroyo Machado, 2025). Además, el contenido que generan puede ser exportado en distintos formatos de audio, video, imagen o texto; si bien, nuestro interés son las respuestas generadas en formato simple; es decir, texto plano o, en su defecto, mínimamente formateado con markdown o similar.

En educación matemática, la investigación sobre chatbots basados en IAG se ha centrado principalmente en la asistencia para la resolución de problemas, donde se han identificado pros (proporcionar soluciones detalladas) y contras (errores e imprecisiones, falta de comprensión profunda) (Noster et al., 2024; Parra et al., 2024; Schorcht et al., 2024). Esta asistencia se ha visto condicionada, entre otros factores, por la manera de hacer consultas al sistema, los datos de entrenamiento y la cercanía de estos problemas a modelos tipo de los que comúnmente se encuentran resoluciones en repositorios digitales. También, el hecho de proporcionar o no ejemplos suficientes para que el sistema logre emular el comportamiento deseado, lo que se refleja en el uso de técnicas como la “zero-shot” (sin ejemplos), “one-shot” (un único ejemplo) y “few-shot” (unos pocos ejemplos) (Schorcht et al., 2024).

En el ámbito de la formulación de problemas, la investigación se ha centrado en el uso de herramientas de IA como asistentes de escritura para generar problemas con características específicas (Berryhill et al., 2024; Embid y Perdomo-Díaz, 2024) o como oponentes didácticos que emulan modos de argumentación para indicar por qué un problema no cumple determinados criterios (Pochulu y Font, 2025). Asimismo, se han desarrollado estudios sobre la reformulación de problemas para su aplicación en contextos distintos al original (Einarsson et al., 2024) y sobre las percepciones docentes asociadas a la práctica de formular problemas matemáticos (Embid et al., aceptado). Comprender cómo docentes en formación y ejercicio, investigadores y formadores de docentes operan con estas herramientas, qué principios tecnológicos las sustentan y de qué manera pueden aprovecharse en prácticas como la formulación de problemas matemáticos resulta clave para su uso en la formación inicial docente.

Objetivos del estudio

El objetivo general de esta investigación se aborda a partir de dos objetivos específicos (OE):

OE1. Identificar qué características de las actividades están presentes en las peticiones de formulación de problemas abiertos que los investigadores realizan a la herramienta de IAG y en su valoración de las respuestas obtenidas.

OE2. Analizar cómo cambia la valoración de los investigadores al modificar las peticiones para la formulación de problemas matemáticos abiertos en la herramienta de IAG.

Metodología

Muestra

El estudio se realizó con los veintidós participantes de la Reunión Intermedia del Grupo de Conocimiento y Desarrollo Profesional del Profesor (CDPP) de la SEIEM, celebrada en marzo de 2025. Se trata de una muestra intencional, donde todos los integrantes tienen en común su interés por la investigación en educación matemática. La muestra está formada por individuos con diversa situación académica: un estudiante de máster, cuatro estudiantes de doctorado y diecisiete doctores con distintas trayectorias, uno con menos de cinco años de antigüedad en esta situación y el resto con más de cinco años.

Para el desarrollo de la experimentación, los participantes se agruparon en once parejas, de manera autónoma. En diez de estas parejas, al menos uno de los integrantes era doctor con más de cinco años de experiencia profesional. Algo más de la tercera parte de las parejas (36,36 %) estaba integrada por dos participantes con poca experiencia previa en investigación sobre formulación de problemas matemáticos. En relación con el uso de chatbots, casi la mitad de las parejas indicó que ninguno de sus miembros tenía experiencia (45,45 %).

Diseño experimental

La experiencia diseñada para este estudio consistió en un taller con una primera parte dedicada a la introducción y discusión en torno a conceptos como formulación de problemas, problema y problema de final abierto y una segunda parte dedicada a realizar una actividad de formulación de problemas (Figura 1) usando la herramienta Perplexity (<https://www.perplexity.ai/>). Esta

herramienta se seleccionó por su facilidad de uso, acceso sin registro y actuación como motor de búsqueda con IAG integrada que facilita las fuentes de información consultadas. Además, permite consultar modelos como Claude Sonnet 4, GPT-4.1, Gemini 2.5 Pro y Grok 3.

Actividad por parejas

1. Busca Perplexity en Google.
2. Accede sin necesidad de registro en la web.
3. Escribe **UNA** petición para que Perplexity formule problemas de final abierto para primaria.
4. **NO** cierras la web y formula una segunda petición para tratar de mejorar la respuesta generada.



<https://www.perplexity.ai/>

Figura 1. Enunciado de la actividad diseñada para la experimentación (Elaboración propia).

La actividad consta de dos tareas: a) redactar una petición para que Perplexity genere problemas de final abierto dirigidos a estudiantes de educación primaria; b) formular una nueva solicitud con el objetivo de mejorar la respuesta obtenida.

Procedimiento de recogida y análisis de datos

Para la recogida de datos se diseñó un cuestionario en formato Google Forms, organizado en tres bloques. El primero de los bloques consta de un conjunto de preguntas cuyo objetivo es la identificación del perfil académico y experiencia previa de los participantes en formulación de problemas y uso de IAG. Esta información se utilizó para la descripción de la muestra, realizando un análisis de frecuencia. En los otros dos bloques se recopila la información correspondiente a la interacción de las parejas con la herramienta (Figura 2). Se distingue entre: (a) la petición inicial con la herramienta Perplexity, evaluación numérica en una escala del 1 al 10 del grado de satisfacción ante la respuesta generada y justificación de dicha valoración; y (b) reformulación de la petición, segunda evaluación numérica del grado de satisfacción y justificación correspondiente.

The image shows a digital questionnaire interface with two main sections. The left section contains three input areas: a text box for a question about Perplexity, a 1-10 star rating scale, and a justification text box. The right section contains two input areas: a text box for a question about proposed modifications, a 1-10 star rating scale, and a justification text box.

Figura 2. Segundo y tercer bloque del cuestionario (Elaboración propia).

El proceso de análisis se dividió en dos partes, atendiendo a la naturaleza de los datos y de los objetivos. Por un lado, con los datos correspondientes a los requerimientos formulados a Perplexity y las justificaciones de las valoraciones de las respuestas de la herramienta, se realizó un análisis de contenido, empleando el método de categorías emergentes. Esto permite atender al primero de los objetivos específicos.

El estudio de la valoración que los participantes hacen de las respuestas de la herramienta se hizo a partir de un análisis de frecuencias de la satisfacción indicada para cada una de las dos respuestas obtenidas. Para ello, se definieron cinco niveles siguiendo los rangos habituales del sistema de calificación universitario español: Suspenso (SUS: 1, 2, 3 ó 4), Aprobado (AP: 5 ó 6), Notable (NOT: 7 u 8), Sobresaliente (SB: 9) y Matrícula de Honor (MH: 10). Se analizó la evolución en el grado de satisfacción indicado y se exploraron las relaciones entre el grado de satisfacción indicado y las categorías que emergieron en el análisis de las peticiones y las justificaciones de las valoraciones.

Resultados

La presentación de los resultados se estructura en torno a cuatro momentos secuenciales que reflejan las etapas de interacción de los participantes con el asistente de inteligencia artificial:

1. Formulación de problemas (FP): Las parejas indican que solicitaron generar a Perplexity, especificando las características deseadas en el problema a formular.
2. Primera valoración (V1): Tras recibir la primera respuesta del asistente, los participantes indican su satisfacción, asignando una puntuación cuantitativa discreta en una escala de 1 a 10, y la justifican mediante comentarios cualitativos.
3. Reformulación (RP): En función de la primera respuesta de Perplexity, los participantes solicitan ajustes o modificaciones para adaptar el problema a sus expectativas o necesidades.
4. Segunda valoración (V2): Los participantes realizan una nueva evaluación de satisfacción, asignando una nueva puntuación y justificando cualitativamente.

Características de las peticiones a la IAG

A partir del análisis de las solicitudes que los participantes realizaron en los momentos FP y RP y de las justificaciones presentadas en los momentos V1 y V2, se identificaron cinco categorías a partir de las cuáles se realizó la codificación de las peticiones y valoraciones: (1) apertura del problema, que considera si el enunciado permite múltiples enfoques o soluciones; (2) área de conocimiento, referida al dominio matemático implicado (por ejemplo, geometría o aritmética); (3) público objetivo, relacionado con el alumnado al que se dirige; (4) realismo del contexto, que evalúa la conexión del problema con situaciones verosímiles; y (5) redacción, centrada en la claridad y accesibilidad del enunciado. Para indicar el grado de presencia de cada categoría, en cada uno de los cuatro momentos, se establecieron tres niveles de codificación:

- Sin evidencia (SE): No se menciona ninguna característica asociada a la categoría.
- No suficiente (NSUF): Se identifica la ausencia o insuficiencia de alguna característica asociada a la categoría.
- Suficiente (SUF): Se reconoce la presencia y pertinencia de alguna característica asociada a la categoría.

La Tabla 1 muestra la frecuencia con que emergieron cada una de las cinco categorías a lo largo de las cuatro etapas del proceso de interacción con el asistente de IAG.

Tabla 1. Número de parejas con evidencia de consideración de cada categoría

	FP	V1	RP	V2	Total
Apertura del problema	9	7 (3)	3	8 (4)	27(7)
Área de conocimiento	10	7 (4)	5	7	29(4)
Público objetivo	9	2 (1)	2	2	15(1)
Realismo del contexto	0	4 (3)	2	4 (3)	10(6)
Redacción	1	3 (2)	4	7 (3)	15(5)

Nota. Los valores entre paréntesis indican los casos en que las parejas identifican la ausencia o insuficiencia de alguna característica asociada a la categoría (NSUF).

Las características que aparecen con mayor frecuencia en las peticiones a la IAG y la valoración de sus respuestas son el área de conocimiento y la apertura del problema.

En la etapa de formulación de problemas (FP), se observa una alta frecuencia en la consideración del área de conocimiento (10), la apertura del problema (9) y el público objetivo (9). Sin embargo, el realismo del contexto estuvo ausente en esta fase y la redacción tuvo una presencia minoritaria (1).

Durante la primera valoración (V1), tras recibir la primera propuesta generada por la IA, se evidencian insuficiencias respecto al área de conocimiento (4 casos insuficientes), la apertura del problema (3 casos insuficientes) y el público objetivo (1 caso insuficiente). Además, se identificaron por primera vez comentarios sobre el realismo del contexto (4, con 3 de ellos de insuficiencia) y la redacción (3, con 2 de ellos de insuficiencia).

En la etapa de reformulación (RP), los participantes solicitaron ajustes principalmente relacionados con la redacción (5) y el realismo del contexto (4). El área de conocimiento (2) y la apertura del problema (3) recibieron menor atención en esta fase, donde las principales demandas de mejora se desplazaron hacia la claridad contextual de los enunciados.

Finalmente, en la segunda valoración (V2), se observa una mayor presencia y adecuación respecto al área de conocimiento (7), la apertura del problema (8) y la redacción (7), aunque persisten insuficiencias, especialmente en realismo del contexto (4, con 3 insuficientes) y la apertura del problema (8, con 4 insuficientes).

Valoración de las respuestas de la IAG y su relación con las categorías emergentes

En relación con la puntuación numérica asignada por cada pareja a la respuesta dada por Perplexity, lo que hemos considerado un indicador de la satisfacción de los participantes, los datos muestran una gran variedad (Tabla 2). Un análisis global de esos datos muestra que tres parejas (P1, P4, P11) asignaron una puntuación inferior a cinco, es decir, un suspenso, a la respuesta de la IAG, en alguno de los dos momentos de valoración V1 y V2. En el lado opuesto, tres parejas puntuaron a Perplexity con un 10 en alguna de sus respuestas (P1, P6, P9), con P6 señalando satisfacción máxima en los dos momentos (P6).

Tabla 2. Puntuaciones asignadas por cada pareja a las respuestas dadas por Perplexity

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
V1	3	8	5	5	8	10	7	7	6	7	5
V2	10	6	5	4	8	10	8	9	10	6	4

Si se comparan las puntuaciones en los dos momentos, el análisis de las medias revela una mejora global: la media en V1 fue de 6,45 (DT = 1,91) y en V2 se incrementó en 0,82 puntos hasta 7,27 (DT = 2,37). Se identifican tres tendencias en relación con cambios en la satisfacción: ascenso, mantenimiento y descenso (Tabla 3).

Tabla 3. Evaluación de la satisfacción en V1 (filas) y V2 (columnas)

	SUS (3-4)	AP (5-6)	NOT (7-8)	SB (9)	MH (10)
SUS (3-4)					P1
AP (5-6)	P4, P11	P3			P9
NOT (7-8)		P2, P10	P5, P7	P8	
SB (9)					
MH (10)					P6

La diagonal de la tabla 3 correspondería a aquellas parejas que muestran el mismo grado de satisfacción en los dos momentos, la parte superior serían las parejas que muestran un ascenso en su nivel de satisfacción y la parte inferior las parejas en las que hay un descenso.

En tres de las parejas hubo un aumento de satisfacción (P1, P8, P9). En este grupo destaca la pareja P1, cuyas respuestas en cada uno de los cuatro momentos fueron las siguientes:

FP: *Un problema abierto*

V1: *No era de matemáticas*

RP: *Que sea de matemáticas*

V2: *Se podría llevar directamente al aula*

Esta pareja pasó de asignar un Suspenso (3) a puntuar la respuesta de la IA con Matrícula de Honor (10). Atendiendo a sus respuestas, el aumento de satisfacción está asociado a un cambio en la categoría área de conocimiento.

En el caso de la pareja P9, que pasó de un aprobado a una MH, el aumento también está asociado a la categoría área de conocimiento, aunque de forma diferente a P1. Las respuestas de P9 fueron:

FP: *Pedimos un problema aritmético. Suponíamos un problema de suma, restas y multiplicaciones sencillas de números naturales. Nuestro conocimiento DM nos motivó a acotar el contenido a la aritmética. También le dijimos que fuera de 2º de primaria.*

V1: *El contexto, los contenidos matemáticos y el lenguaje eran adecuados para la edad. Cumplía el requisito de ser abierto y el resto de condiciones. Solo tenía problemas con el conjunto numérico involucrado.*

RP: *Evitar números decimales.*

V2: *Creemos que se logró un problema con las características solicitadas.*

Cuatro parejas mantuvieron su categoría en ambas mediciones (P3, P5, P6, P7). Por ejemplo, las respuestas de P6 muestran una propuesta ya bien orientada en V1, que se mantuvo en V2 con algunos cambios en la redacción y especificidad del contenido. Estas fueron las respuestas de P6:

FP: *Me gustaría un problema de final abierto de geometría para el tercero de la primaria*

V1: *Es un problema abierto y trabaja con ideas sofisticadas de geometría*

RP: *Que se incluya otras características más allá de ángulos y lados, y mejore las instrucciones del problema.*

V2: *Las instrucciones fueron ampliadas sobremanera, incluyendo figuras planas y tridimensionales.*

Finalmente, se identificaron cuatro casos en los que la satisfacción indicada en V2 fue inferior a la señalada en V1 (P2, P4, P10, P11). En el caso de P4, por ejemplo, aunque se intentó mejorar el realismo del contexto, la segunda versión (V2) no logró consolidar una propuesta

verdaderamente abierta ni verosímil, forzando un contexto artificial que no resolvió las limitaciones señaladas en V1. Estas fueron las respuestas de P4:

FP: *Le dimos nuestro perfil, maestras de primaria y le pedimos un problema abierto para 5 primaria sobre fracciones*

V1: *No era realista en nuestro contexto y las preguntas que planteaba proponer hacía que no fuese abierto. Además, la estrategia siempre sería la misma*

RP: *Cambio a un contexto más real*

V2: *Sigue sin ser un problema abierto tanto en estrategia como en resultado. Ha forzado un contexto discreto distinto que tampoco es muy real.*

Los participantes que mejoraron su categoría presentaron la mayor satisfacción final (media = 9,67; DT = 0,47), seguidos por quienes mantuvieron su valoración (media = 8,00; DT = 1,41). Por su parte, quienes descendieron de categoría mostraron los niveles más bajos de satisfacción (media = 5,00; DT = 1,00).

En base a estos resultados, se formuló una regla para predecir la variación en el nivel de satisfacción entre la valoración inicial (V1) y la segunda valoración (V2), basada en la presencia de elementos no suficientemente fundamentados (NSUF) en V2. Se hipotetizó que la persistencia de insuficiencias en alguno de los aspectos evaluados en V2 debería conllevar una disminución en la satisfacción percibida. Se contrastaron los resultados esperados con los observados, obteniéndose una tasa de coincidencia del 72,7% (8 de 11 casos). Las excepciones se tienen en las parejas P3, P6 y P7, quienes, a pesar de no satisfacer alguno de los aspectos evaluados, mantienen o incluso incrementan su nivel de satisfacción en V2.

Asimismo, se estableció otra regla de decisión según la cual, si en la valoración persiste al menos un NSUF, es necesario seguir refinando el prompt y valorar la nueva respuesta. Esta pauta permite identificar propuestas aún deficitarias en alguna de las dimensiones consideradas, incluso si han mejorado en otras. En este sentido, se conjetura que el aumento de ciclos de valoración puede contribuir a una mejora progresiva en la calidad de las propuestas, reduciendo la presencia de aspectos no suficientemente desarrollados. Esta tendencia se refleja en los resultados: mientras que en la primera valoración (V1) un 82 % de los participantes presentaban al menos un código NSUF (siendo P6 y P8 las únicas excepciones), en la segunda valoración (V2) este porcentaje se redujo al 64 %, sumándose P1 y P9 al grupo de excepciones.

Discusión final y conclusiones

Este estudio piloto se enmarca en una investigación predoctoral más amplia orientada a explorar el potencial de la inteligencia artificial generativa para apoyar los procesos de formulación y reformulación de problemas matemáticos en la formación inicial del profesorado de educación primaria (Embid y Perdomo-Díaz, 2024; Embid et al., aceptado). En relación con el objetivo específico 1, los resultados revelan que los participantes se centraron inicialmente en tres características concretas de la tarea (área de conocimiento, apertura del problema y público objetivo). La interacción con Perplexity hizo visibles otras dimensiones relevantes, como el realismo del contexto y la redacción. En este sentido, consideramos que Perplexity actuó como un “espejo didáctico” que permitió a los investigadores detectar aspectos no explicitados en sus peticiones iniciales y ajustar sus expectativas y requerimientos. Este término viene a complementar otros descritos en la literatura, como el de “oponente didáctico” de Pochulu y Font (2025), según el cual la IA no sustituye al profesor, sino que respalda e intensifica su análisis reflexivo.

Respecto al objetivo específico 2 (OE2), se observó una satisfacción general alta tras el uso de la IAG, con la mayoría de las parejas manteniendo o incrementando su valoración. Destacan casos en los que la satisfacción pasó de niveles bajos (3 o 6) a la puntuación máxima (10), lo que podría sugerir que la IAG puede ser especialmente útil para parejas de usuarios que, en un inicio, tienen menos claridad en aquello que pueden solicitar a un sistema generativo con opción a formular problemas. Trabajos previos advierten sobre la necesidad de una interacción guiada y reflexiva para aprovechar plenamente el potencial de la IAG en este contexto (Pochulu y Font, 2025). Los factores críticos que afectaron negativamente la satisfacción de los usuarios estuvieron relacionados principalmente con el realismo del contexto, la claridad en la redacción del enunciado y el grado de apertura del problema, aspectos ya señalados como desafíos en trabajos previos de formulación de problemas (Embid et al., 2024). Cankoy y Özder (2017) ya alertaban de dificultades con el realismo del contexto o razonabilidad entre el propio alumnado de primaria. Estos resultados refuerzan la necesidad de que herramientas basadas en IAG sean entrenadas más allá de los problemas ordinarios de las aulas de primaria, de modo que los

problemas que generen consideren las características didácticas y contextuales relevantes para el profesorado en formación.

En líneas generales, el estudio muestra que la reiteración en el uso de inteligencia artificial generativa puede aumentar la satisfacción de los usuarios, aunque este incremento es, en promedio, bajo. El uso de un asistente como Perplexity para la formulación y posterior reformulación de problemas abiertos de matemáticas para primaria podría resultar beneficioso para ciertos perfiles de investigadores que buscan un “gemelo” digital capaz de aportar información adicional, aunque con una calidad de las respuestas variable (Embid et al., 2024; Noster et al., 2024). La satisfacción mejora especialmente cuando las respuestas de la IAG se ajustan a las necesidades específicas de contenido y público objetivo (indicador de coherencia curricular de Leavy y Hourigan, 2022), mientras que disminuye si no se cumplen adecuadamente aspectos como el realismo del contexto (Cankoy y Özder, 2017) o la apertura del problema (Pehkonen, 1997). Sin embargo, el tamaño reducido y la especificidad de la muestra (investigadores en didáctica de la matemática), junto con la falta de ciertos datos (como el registro detallado de prompts y respuestas), limitan la generalización de los resultados y la validez de los análisis. A pesar de estas limitaciones, el uso de IAG abre oportunidades para mejorar la formación docente, inspirar nuevas propuestas formativas y promover la reflexión sobre la integración de IA en la educación matemática.

Agradecimientos

Este trabajo ha sido realizado con el apoyo del proyecto PID2022-139007NB-I00, financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE. Además, S.E. cuenta con un contrato predoctoral PREP2022-000959 financiado por MCIN/AEI/10.13039/501100011033 y FSE+.

Referencias bibliográficas

Area-Moreira, M., Del Prete, A., Sanabria-Mesa, A. L. y Sannicolás-Santos, M. B. (2024). No todas las herramientas de IA son iguales. Análisis de aplicaciones inteligentes para la enseñanza universitaria. *Digital Education Review*, 45, 141-149. <https://doi.org/10.1344/der.2024.45.141-149>

- Baumanns, L. y Rott, B. (2022). Developing a framework for characterising problem-posing activities: A review. *Research in Mathematics Education*, 24(1), 28-50. <https://doi.org/10.1080/14794802.2021.1897036>
- Berryhill, A., Chandler, L., Bondurant, L. y Sapkota, B. (2024). Using ChatGPT As A Thought Partner in Writing Relevant Proportional Reasoning Word Problems. *Connections*, 33 (4).
- Blanco, L.J. y Pino, J. (2015). ¿Qué entendemos por problema de matemáticas? En Manuales UEX (Ed.), *La resolución de problemas de Matemáticas en la formación inicial de profesores de Primaria* (pp. 81-92). Universidad de Extremadura
- Cai, J. y Hwang, S. (2020). Learning to teach through mathematical problem posing: Theoretical considerations, methodology, and directions for future research. *International Journal of Educational Research*, 102, 101391. <https://doi.org/10.1016/j.ijer.2019.01.001>
- Cankoy, O. y Özder, H. (2017). Generalizability Theory Research on Developing a Scoring Rubric to Assess Primary School Students' Problem Posing Skills. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(6), 2423-2439. <https://doi.org/10.12973/eurasia.2017.01233a>
- De Guzmán, M. (1993). *Tendencias Innovadoras en educación matemática*, Ediciones OEA
- Einarsson, H., Lund, S. H. y Jónsdóttir, A. H. (2024). Application of ChatGPT for automated problem reframing across academic domains. *Computers and Education: Artificial Intelligence*, 6, 100194. <https://doi.org/10.1016/j.caeai.2023.100194>
- Embid, S. y Perdomo Díaz, J. (2024). Evaluación del potencial de distintos chatbots para la formulación de problemas matemáticos. *FPIEM: Formación del Profesorado e Investigación en Educación Matemática*, 16, 41-58.
- Embid, S., Perdomo-Díaz, J. y Giaconi, V. (aceptado). Formulación de problemas y uso de IA: Creencias motivacionales en futuros docentes de primaria. *Avances De Investigación En Educación Matemática*
- García-Peñalvo, F. J., Llorens-Largo, F. y Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1). <https://doi.org/10.5944/ried.27.1.37716>
- Leavy, A. y Hourigan, M. (2022). The Framework for Posing Elementary Mathematics Problems (F-PosE): Supporting Teachers to Evaluate and Select Problems for Use in Elementary Mathematics. *Educational Studies in Mathematics*, 111(1), 147-176. <https://doi.org/10.1007/s10649-022-10155-3>
- Looi, C.-K. y Jia, F. (2025). Personalization capabilities of current technology chatbots in a learning environment: An analysis of student-tutor bot interactions. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-025-13369-z>

Kantowski, M. G. (1981). Problem solving. En: E. Fennema (Ed.), *Mathematics education research: Implications for the 80's* (pp.111–126). Reston, VA: National Council of Teachers of Mathematics

Kuhail, M. A., Alturki, N., Alramlawi, S. y Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973-1018. <https://doi.org/10.1007/s10639-022-11177-3>

National Council of Teachers of Mathematics. NCTM (2014). *Principles to actions: Ensuring mathematical success for all*. Reston, VA: Author.

Noster, N., Gerber, S. y Siller, H.-S. (2024). Pre-Service Teachers' Approaches in Solving Mathematics Tasks with ChatGPT. *Digital Experiences in Mathematics Education*, 10, 543–567. <https://doi.org/10.1007/s40751-024-00155-8>

Parra, V., Sureda, P., Corica, A., Schiaffino, S. y Godoy, D. (2024). Can Generative AI Solve Geometry Problems? Strengths and Weaknesses of LLMs for Geometric Reasoning in Spanish. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 65. <https://doi.org/10.9781/ijimai.2024.02.009>

Pehkonen, E. (1997). The state-of-art in mathematical creativity. *Zentralblatt Für Didaktik Der Mathematik*, 29(3), 63-67. <https://doi.org/10.1007/s11858-997-0001-z>

Perdomo-Díaz, J. y de Armas-González, P. (2025). La resolución de problemas de final abierto como herramienta para la atención a la diversidad y el desarrollo de la creatividad matemática. En *Desarrollo de habilidades en el aula de matemáticas* (pp. 6–11). Universidad de O'Higgins.

Perdomo-Díaz, J. y Felmer, P. (2017). El taller RPaula: activando la resolución de problemas en las aulas. *Profesorado. Revista de currículum y formación del profesorado*, 21 (2), 425-444. <https://doi.org/10.30827/profesorado.v21i2.10343>

Pochulu, M. D. y Font, V. (2025). Idoneidad didáctica de tareas de matemáticas reformuladas con inteligencia artificial. *PARADIGMA*, e2025027. <https://doi.org/10.37618/PARADIGMA.1011-2251.2025.e2025027.id1621>

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press

Schorcht, S., Buchholtz, N. y Baumanns, L. (2024). Prompt the problem – investigating the mathematics educational quality of AI-supported problem solving by comparing prompt techniques. *Frontiers in Education*, 9. <https://doi.org/10.3389/educ.2024.1386075>

Silver, E. A. (1994). On Mathematical Problem Posing. *For the Learning of Mathematics*, 14(1),19-28.

Singer, F. M., Ellerton, N. y Cai, J. (2013). Problem-posing research in mathematics education: New questions and directions. *Educational Studies in Mathematics*, 83(1), 1-7.

Son, J. W. y Diletti, J. (2017). What Can We Learn from Textbook Analysis? En J.-W. Son, T. Watanabe y J.-J. Lo (Eds.), *What Matters? Research Trends in International Comparative Studies in Mathematics Education* (pp. 3-32). Springer International Publishing. https://doi.org/10.1007/978-3-319-51187-0_1

Terblanche, N. H. D. (2024). Smooth Talking: Generative Versus Scripted Coaching Chatbot Adoption and Efficacy Comparison. *International Journal of Human-Computer Interaction*, 0(0), 1-14. <https://doi.org/10.1080/10447318.2024.2423125>

Torres-Salinas, D. y Arroyo-Machado, W. (2025). Flujos de trabajo inteligentes con ChatGPT, Perplexity y NotebookLM. <https://doi.org/10.5281/zenodo.14981646>

Wardat, Y., Tashtoush, M., Alali, R. y Saleh, S. (2024). Artificial Intelligence in Education: Mathematics Teachers' Perspectives, Practices and Challenges. *Iraqi Journal For Computer Science and Mathematics*, 5(1), 60-77. <https://doi.org/10.52866/ijcsm.2024.05.01.004>

Zhu, Y. y Fan, L. (2006). Focus on the Representation of Problem Types in Intended Curriculum: A Comparison of Selected Mathematics Textbooks from Mainland China and the United States. *International Journal of Science and Mathematics Education*, 4(4), 609-626. <https://doi.org/10.1007/s10763-006-9036-9>