

Análisis de los diálogos con ChatGPT del profesorado en formación durante el diseño de actividades didácticas de Química

Analysis of the Dialogues with ChatGPT by Pre-service Teachers during the Design of Chemistry Didactic Activities

Victor Lopez-Simó ^a, Mikael Frank Rezende Junior ^b

^a Departament de Didàctica de la Matemàtica i les Ciències Experimentals, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Catalunya, España; ^b Instituto de Física e Química, Universidade Federal de Itajubá, Itajubá, Brasil.

Resumen. Este estudio analiza los cambios en la estructura y el contenido de los diálogos entre estudiantes del grado en Maestro y ChatGPT durante el diseño de actividades didácticas de química. Se examinaron 223 prompt procedentes de 28 conversaciones (14 PRE y 14 POS) realizadas por estudiantes de tercer curso, observándose un enriquecimiento progresivo en los contenidos científicos y didácticos, así como una mayor sofisticación en las interacciones. Estas incluyeron preguntas más complejas, un uso adecuado de la terminología técnica y una mayor consideración de las ideas previas de los futuros maestros. También se identificaron transformaciones cualitativas en las dinámicas conversacionales, con mayor complejidad, intencionalidad pedagógica y presencia de prompt expansivos, inexistentes en la fase inicial y prominentes en la final. Los prompts fueron categorizados según su contenido (científico, didáctico, contextual y práctico) y función comunicativa (adición, expansión, modulación o refutación), lo que permitió una comparación entre fases. El estudio adoptó un enfoque mixto, combinando estadística descriptiva y análisis cualitativo de contenido. Los resultados sugieren que actividades estructuradas para el uso pedagógico de ChatGPT pueden fortalecer competencias vinculadas al conocimiento didáctico del contenido (PCK) y al uso crítico de tecnologías de Inteligencia Artificial Generativa en la enseñanza de las ciencias.

Palabras-clave:
ChatGPT, Diseño Didáctico, Educación Científica, Inteligencia Artificial Generativa, Química.

Submetido em
07/11/2024

Aceito em
22/08/2025

Publicado em
25/08/2025

Abstract. This study analyzes changes in the structure and content of dialogues between pre-service primary teachers and ChatGPT during the design of chemistry teaching activities. A total of 223 prompts from 28 conversations (14 PRE and 14 POS) carried out by third-year students were examined, revealing a progressive enrichment in scientific and pedagogical content, as well as increased sophistication in the interactions. These included more complex questions, appropriate use of technical terminology, and consideration of students' prior conceptions. Qualitative transformations in the conversational dynamics were also identified, characterized by greater complexity, pedagogical intentionality, and the emergence of expansive prompts, absent in the initial phase and prominent in the final one. Prompts were categorized according to their thematic content (scientific, pedagogical, contextual, and practical) and communicative function within the dialogue (addition, expansion, modulation, or refutation), allowing for a comparative analysis between phases. The study adopted a mixed-methods approach, combining descriptive statistics and qualitative content analysis. Findings suggest that structured activities designed to support pedagogical uses of ChatGPT can enhance pre-service teachers' competencies related to pedagogical content knowledge (PCK) and foster a critical and reflective approach to the use of generative artificial intelligence technologies in science education.

Keywords: ChatGPT, Chemistry, Generative Artificial Intelligence, Instructional Design, Science Education.

Introducción

Aunque la historia de la Inteligencia Artificial (IA) se remonta a muchas décadas atrás, es a partir de los años 2000 donde la IA en general comenzó a ser utilizada en la vida cotidiana, en sectores como finanzas, salud, transporte o entretenimiento, teniendo cada vez mayor presencia en nuestras vidas. No obstante, su gran popularización se ha asociado recientemente con las IA conversacionales basadas en Large Language Model (LLM), siendo ChatGPT de OpenAI la más icónica y popular. Estas herramientas han demostrado capacidades promisorias en tareas tradicionalmente desempeñadas por profesionales de distintas áreas, incluyendo la educación (Bommarito et al., 2023). Desde entonces, otras IA de grandes empresas de tecnología como Gemini (Google), Copilot (Microsoft) o Claude (Anthropic) compiten por ofrecer la tecnología más avanzada y parecida a interacción conversacional humana.

Pero, desde que ChatGPT se lanzó en noviembre de 2022 y pronto se hizo conocido en todo el mundo y utilizado por millones de personas, el tema de la Inteligencia Artificial Generativa (IAG) está en el centro de los debates en el ámbito educativo y genera movimientos interpretativos, predictivos y adaptativos que van desde los micro contextos de cada institución educativa hasta los grandes organismos internacionales, como la UNESCO (Unesco, 2023; Liu et al., 2023). En este contexto, son cada vez más numerosas las dudas vinculadas a los efectos sociales, políticos y económicos de estas tecnologías en nuestras vidas, ya que alteran nuestra relación con el mundo, con los demás y con el conocimiento. En el ámbito específico de la educación, también existe evidencia de las inseguridades que provoca la aparición de cada nueva tecnología. Un ejemplo de ello fue la expansión de la herramienta de búsqueda digital Google en sus inicios. Sin embargo, en la actualidad, este tipo de búsqueda se ha trivializado y, de hecho, ha tenido escaso impacto en las prácticas escolares tradicionales de enseñanza y aprendizaje. Como señalan Serrano Sánchez y Sánchez Vera (2024), la historia de la incorporación de tecnologías en la educación se ha caracterizado por un uso anecdótico o meramente ilustrativo de herramientas, sin una integración didáctica consistente.

Entonces, ¿por qué nosotros, en la educación y en otros ámbitos, nos hacemos en la actualidad tantas preguntas sobre las IAG? En general, muchos de los interrogantes que surgen con estas tecnologías se basan en puntos ligados al su funcionamiento, por ejemplo, el hecho de que el interlocutor robótico accede y depura la información, la contextualiza y aprende de ella, diferenciándose de los tradicionales chatbots que se basan en reglas previamente establecidas (Eraña, 2024). Sustituir así un interlocutor humano por uno robótico basado en un modelo de lenguaje artificial y con acceso rápido a un volumen gigantesco de datos (Big Data) e información, capacitado no sólo para “dar” respuestas, sino para “generar” respuestas diferentes con cada pregunta, reaviva puntos de debate en el campo educativo, que van desde las posibilidades y limitaciones del uso pedagógico de las tecnologías digitales hasta cuestiones éticas asociadas a este uso; responsabilidad y necesidad de lectura crítica; preguntas de evaluación; formación de docentes; autoría y plagio; relación profesor-alumno; mediación pedagógica, etc.

¿Qué es el Prompt?

En el funcionamiento de las IAG, las conversaciones que realizamos se basan en los denominados “prompt”. A grandes rasgos, podemos definir un prompt como una indicación o estímulo que se utiliza para solicitar una respuesta, acción o entrada específica de parte de un usuario. En el contexto de la programación o la interacción con sistemas informáticos, un prompt es una solicitud de entrada de datos dirigida a un usuario o a un sistema, y este mecanismo se utiliza para obtener información e interactuar de manera dinámica. Por lo general, un prompt puede manifestarse como un mensaje de texto o una interfaz gráfica que presenta una pregunta o una instrucción específica. En la generación actual de IAG, los prompt en forma de pequeñas unidades de texto permiten substituir complejos comandos de lenguaje computacional que se usaban hasta hace poco. Esta nueva forma de relacionarnos a través de frases cortas similares a las que dos personas podrían tener a través de un chat entre humanos es una de las claves del éxito de las IA actuales. La notoriedad de la idea del prompt puede verse reflejada en la Figura 1, donde se constata que las búsquedas en Internet que incluían el término “prompt”, según la Google Search, se incrementaron bruscamente a raíz de la aparición de ChatGPT, y actualmente el número de búsquedas a escala global son mucho mayores que antes de 2022.

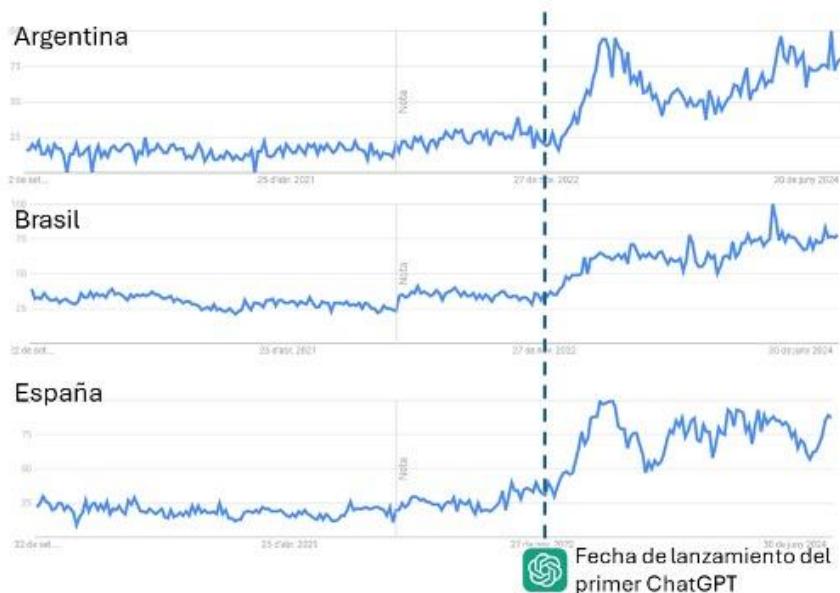


Figura 1. Gráfico generado con la herramienta Google Search del número de búsquedas que incluyen el término “prompt” en tres países distintos: Argentina, Brasil y España. En todos los casos se identifica claramente un antes y un después del lanzamiento de ChatGPT.

Así pues, en el contexto actual, aprender a hacer buenos prompt es una tarea clave. En la interacción con LLM, los prompt desempeñan un papel crucial en la orientación de las respuestas generadas, siendo esenciales para obtener resultados pertinentes. El prompting es el acto de crear una instrucción (prompt) para guiar a un modelo de IA hacia la respuesta deseada. Trabajos como Korzynski et al. (2023) y Baidoo-Anu y Owusu Ansah (2023) han

enfatizado que las habilidades de prompting con IA son habilidades esenciales en la era digital. Elaborar prompt pertinentes para cada conversación implica, por ejemplo, aportar la claridad y la exactitud de las instrucciones proporcionadas, minimizando ambigüedades y garantizando un enfoque específico en la tarea o cuestión propuesta. La especificidad es igualmente importante, ya que delimita el tema y el alcance de la respuesta deseada, permitiendo que el modelo aborde el asunto con el nivel adecuado de detalle. El contexto aportado en los prompt es otra característica vital, proporcionando información adicional que sitúa la cuestión dentro de un escenario particular, lo que facilita la generación de respuestas más relevantes. Además, el formato del prompt, incluyendo su estructura y presentación, puede influir significativamente en la calidad de la respuesta, debiendo indicar explícitamente aspectos como la extensión deseada de la respuesta, el estilo y cualquier requisito específico. Considerar estas características al elaborar prompt no solo mejora la eficacia de las interacciones con modelos de lenguaje, sino que también contribuye a la producción de respuestas que atienden de manera más adecuada las necesidades y expectativas.

En el contexto educativo, estas habilidades de prompting han cobrado mucha relevancia con la aparición de estas herramientas. Diferentes autores han propuesto guías y manuales de cómo hacer un buen uso de la IA en las aulas (Muñoz et al., 2024, Generalitat de Catalunya, 2024), y dentro de las recomendaciones aportadas destaca la importancia de elaborar buenos prompt no solo para hacer el diseño de materiales didácticos más eficaz en términos de calidad (Tena & Couso, 2023), sino en término de tiempo que se dedica a la preparación de los materiales (tanto del alumnado como materiales de apoyo: rúbricas, objetivos de aprendizaje, narrativas, identificación de ideas alternativas del alumnado, clarificación conceptual, etc.). Cabe pensar, por lo tanto, que la habilidad de prompting debe ser una pieza clave en la formación inicial del profesorado.

Así, esta investigación se propone analizar la evolución de las habilidades de prompting de futuros maestros durante un proceso formativo en la asignatura Science Education. Se examinarán los cambios estructurales y funcionales en los prompt generados antes y después de una intervención orientada al diseño de propuestas didácticas apoyadas en IAG.

Marco Teórico

Como destacamos al inicio, la IA ha estado presente en educación mucho antes de la llegada de ChatGPT (Giró Gràcia & Sancho-Gil, 2022), pero el lanzamiento de este popular LLM fue todo un revulsivo, y los debates entorno a los aspectos metodológicos, logísticos, organizativos o éticos no tardaron en llegar. A los pocos meses de su irrupción algunos autores señalaron un impacto positivo en el rendimiento educativo de los estudiantes, gracias a la retroalimentación inmediata proporcionada por la IAG y la promoción del aprendizaje colaborativo (Altarawneh, 2023), y también como un apoyo valioso para el aprendizaje autónomo y la tutoría digital (Fuchs & Aguilos, 2023; Liang et al., 2024), y permitiendo técnicas más sofisticadas que puedan evaluar no sólo la retención de contenido, sino también las habilidades de pensamiento crítico y resolución de problemas (Bordt & von Luxburg, 2023; Gilson et al., 2023).

No obstante, autores como Čavojský et al. (2023) destacaron riesgos como los aspectos éticos del uso de estas IAG y las posibles implicaciones en la sobrecarga cognitiva y la integridad académica. Además, hay que tener en cuenta que, aunque ChatGPT (al igual que otros LLM) puede proporcionar respuestas a preguntas y simular diálogos, sus sugerencias no siempre son precisas ni están libres de sesgos. En este sentido, Avraamidou (2024) sostiene que la integración de la inteligencia artificial en la educación está reproduciendo patrones coloniales de explotación y homogeneización, mientras que sería necesario un enfoque feminista y centrado en las personas que priorice la justicia y la diversidad en el aprendizaje. Finalmente, si nos centramos más concretamente en el ámbito de la educación científica, también han surgido debates más específicos. Diferentes trabajos han analizado el desempeño del ChatGPT en tareas como la resolución de cuestiones escolares de física (Bitzenbauer, 2023; Kortemeyer, 2023; MacIsaac, 2023; López-Simó & Rezende, 2024), señalando importantes desempeños, pero también errores y limitaciones.

En el marco de toda esta discusión, cabe pensar que el dominio de la IAG como herramienta para su uso educativo debe estar dentro de las competencias profesionales de los docentes, que incluyen un amplio abanico de habilidades (selección de contenidos curriculares, adaptación al contexto educativo, identificación de las ideas y las visiones del alumnado, gestión del aula, uso de recursos, evaluación etc.). Para caracterizar dichas competencias, uno de los modelos más conocidos es el TPACK, un planteamiento idealizado por Mishra y Koehler (2006), que es una revisión del clásico modelo PCK de Shulman (1986). Para estos autores, el profesorado debe integrar conocimientos de distinta índole. Así, por un lado, tendríamos el conocimiento del contenido disciplinar (CK), que en nuestro caso es el conocimiento nuclear de las disciplinas científicas como la física, la química, la biología o la geología. A su vez, habría un conocimiento pedagógico-didáctico (PK), que incluye los saberes sobre los procesos de enseñanza y aprendizaje, las estrategias y metodologías; y también un conocimiento tecnológico (TK) centrado en el dominio de las tecnologías digitales existentes en el ámbito educativo.

Tal como se muestra en la Figura 2, estos diferentes tipos de conocimientos pueden combinarse, apareciendo el PCK (es decir, el conocimiento sobre como enseñar y promover el aprendizaje de un contenido disciplinar específico) y el TPACK (el conocimiento sobre cómo integrar tecnologías educativas concretas en los procesos de enseñanza y aprendizaje de estos contenidos disciplinares específicos). En el caso de las ciencias, el TPACK de un docente le debería permitirle, por ejemplo: (a) seleccionar cual es la mejor simulación o laboratorio virtual para aprender un concepto concreto de química; (b) decidir cuándo y cómo usar dicha simulación dentro de una secuencia didáctica y cómo combinarla con experimentos reales; (c) definir qué espera que el alumnado aprenda con dicha herramienta digital; (d) prever qué posibles complicaciones tendrá su uso tanto a nivel logístico como didáctico; (e) planificar cómo será la regulación de los aprendizajes antes, durante y después de su uso, etc. (López-Simó et al., 2017).

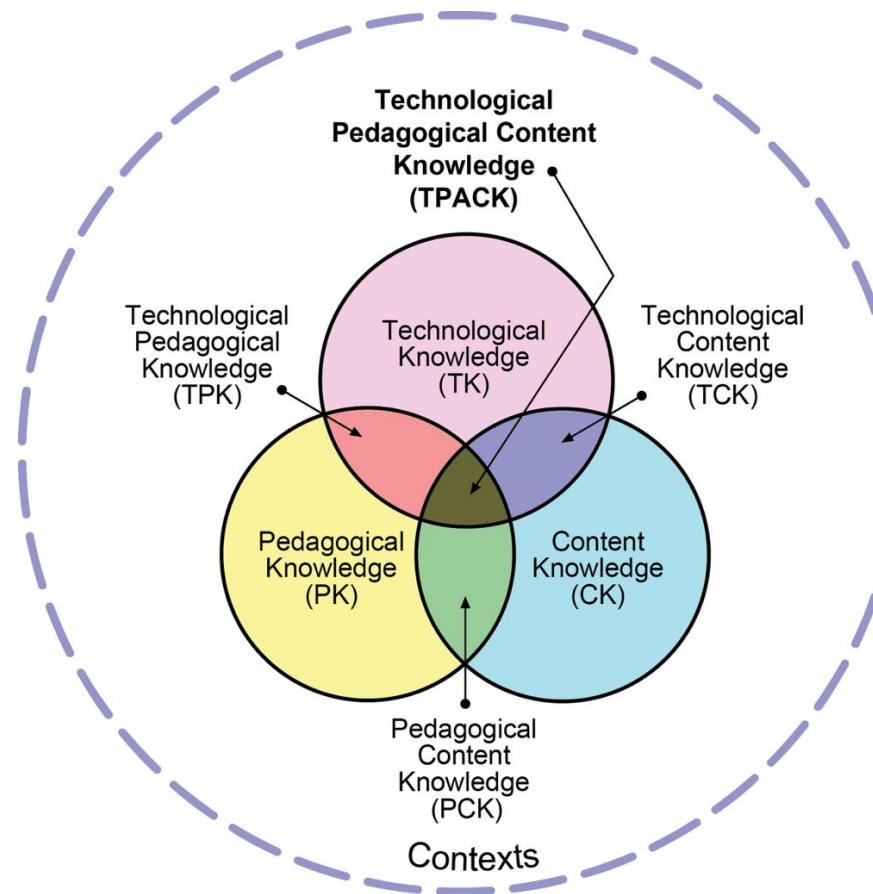


Figura 2. Diagrama propuesto por Mishra y Koehler (2006) para representar la Integración de los distintos conocimientos escolares: el conocimiento del contenido disciplinar, el conocimiento pedagógico y el conocimiento tecnológico.

Con la irrupción de la IA en las escuelas, diferentes autores han propuesto abordar la adopción de esta tecnología desde la perspectiva del TPACK (Castiglione, 2023; Leahy & Mishra, 2023; Goldman et al., 2024), discutiendo cómo las habilidades de prompting previamente discutidas se plasman en forma de TK. Además, Mishra et al. (2023) pone sobre la mesa la necesidad del desarrollo no solo de los vectores anteriores, sino también del Conocimiento Contextual (XK).

Finalmente, dada la componente altamente conversacional de las interacciones escritas entre docentes y ChatGPT u otras IAG, consideramos que también puede ser de gran utilidad los marcos interpretativos de las interacciones estudiante – docente que se dan en las aulas, que han definido patrones de interacción como los conocidos IRE (Iniciación-Respuesta-Evaluación), y que se han usado desde hace décadas en una amplia variedad de contextos (Candela, 1998; Rojas-Drummond et al., 2013; Ruiz-Primo & Furtak, 2007). A grandes rasgos, una conversación puede concebirse como una combinación de “talk moves” donde intervienen dinámicas de ‘Back’, que permiten recuperar y reflexionar sobre conceptos ya abordados previamente en la conversación, y dinámicas de ‘Forward’, que impulsan la formulación de nuevas preguntas y la introducción de temas inéditos. Además, toda conversación puede también combinar fases de ‘Apertura’, donde se fomenta la exploración de nuevas ideas, y ‘Cierre’, enfocadas en la conclusión y el resumen de los aprendizajes.

(Mortimer & Scott, 2010). Este marco útil para comprender como se dan las interacciones contiene elementos que podría también servir para ejercitar la comprensión de los diálogos entre personas y ChatGPT, aunque se trata de un campo aún poco explorado académicamente.

En un escenario de investigación distinto, Severo Prodanov y Serrano De Andrade Neto (2023) señalaron que la continuidad de las investigaciones sustentadas en el modelo TPACK puede favorecer prácticas innovadoras en la formación del profesorado en química, dado que, tras participar en programas de capacitación, los futuros docentes evidenciaron mejoras en su autopercepción respecto a los distintos dominios del conocimiento profesional docente, especialmente en relación con el contenido y su didáctica. Este potencial del modelo TPACK se vuelve aún más significativo ante la creciente incorporación de las IAG en los contextos educativos.

Karataş y Ataç (2025), por ejemplo, proponen una modelización estructural del denominado AI-TPACK, subrayando que la integración ética y pedagógica de tecnologías basadas en IAG exige no solo competencia técnica, sino también sensibilidad crítica hacia sus implicaciones educativas. En este sentido, Feldman-Maggot et al. (2025) destacan que, en la enseñanza de la química, las IAG pueden facilitar la construcción de visualizaciones mentales de fenómenos microscópicos, elemento clave en esta disciplina, siempre que se inserten dentro de marcos pedagógicos sólidos como los que ofrece el TPACK. De manera complementaria, Celik (2023) introduce el concepto de “Intelligent-TPACK” como una evolución necesaria del modelo, en la medida en que articula saberes tecnológicos, pedagógicos, disciplinares y éticos para una mediación reflexiva con IAG. Así, en la formación docente en ciencias, y en particular en química, la incorporación crítica de la IAG constituye tanto una oportunidad para enriquecer el diseño didáctico como un desafío formativo, al exigir una revisión del rol que la tecnología desempeña en la construcción del conocimiento científico en contextos escolares.

Metodología

Contextos y participantes

A partir de este marco conceptual, se diseñó una investigación centrada en la asignatura "Science Education", dirigida a estudiantes de tercer año del grado de profesorado en la Universidad Autónoma de Barcelona (UAB), en Cataluña (España). El idioma oficial de la asignatura es el inglés, pero la convivencia lingüística incluye el castellano y el catalán. Por esto, mantuvimos las palabras escritas de los estudiantes en el lenguaje original, sin traducciones.

La asignatura se enfoca en la integración del conocimiento del contenido (CK) y el conocimiento didáctico del contenido (PCK), aunque incluye algunas actividades puntuales sobre conocimiento tecnológico-pedagógico del contenido (TPACK). El grado de profesorado que se ofrece en la UAB tiene una estructura para formación superior de maestros (educación para críos de la educación primaria), sin embargo, estos tienen la obligatoriedad de completar su educación con el máster para ejercer la profesión de docente en las escuelas básicas de

España. El sistema educativo español abarca la educación infantil, primaria, secundaria obligatoria (ESO), bachillerato y, posteriormente, la educación superior. Los estudios de grado universitario inician generalmente a los 18 años y duran aproximadamente cuatro años. Para ejercer como docente en educación primaria, el título de grado ofrece una habilitación parcial, siendo necesario completar un máster para la habilitación profesional plena.

Los participantes de esta asignatura son futuros docentes con un nivel de familiaridad con las ciencias que varía significativamente; algunos pocos provienen de un bachillerato científico, mientras que la mayoría ha dejado de estudiar ciencias a los 15 años, y con un limitado vínculo tanto en el aspecto cognitivo como emocional con la asignatura y su propuesta. En este estudio han participado 64 estudiantes organizados de 4 o 5 personas, formando 15 grupos en total y designados por G1 hasta G15. La mayoría de los estudiantes en esta asignatura son de la Catalunya, una comunidad autónoma de España que tiene la ciudad de Barcelona como su capital, pero también han participado estudiantes de otros países, principalmente de la comunidad europea debido a los incentivos del programa de movilidad Erasmus+. Este programa apoya la educación, la formación y el deporte en Europa. Entre sus principales acciones destaca la posibilidad de realizar intercambios académicos en instituciones extranjeras, promoviendo así el aprendizaje intercultural, el desarrollo académico y la movilidad estudiantil en el espacio europeo.

Actividades con el ChatGPT

Dentro de esta asignatura, se desarrolló una secuencia de actividades enfocadas en el uso de ChatGPT para el diseño didáctico en educación química. La secuencia comenzó con un seminario de dos horas en octubre de 2023, donde tras discutir conceptos clave de la química, como ejemplo: - La materia está compuesta por partículas microscópicas. - Las partículas se mueven a mayor o menor velocidad, y se enlazan más fuerte o más flojo entre ellas según la temperatura y el estado de agregación (sólido, líquido y gas).

A los estudiantes se les pidió a diseñar una actividad para niños y niñas de primaria que les ayudara a comprender que el aire y el agua están compuestos por partículas. Encima de la mesa del laboratorio tenían disponibles distintos objetos, como globos, jeringas, cañitas de plástico y vasos. Para diseñar esta primera actividad (fase PRE), los estudiantes podían usar el ChatGPT, pero sin instrucciones previas ni ningún tipo de andamiaje. Esta actividad posteriormente la tenían que implementar con otro grupo de estudiantes. Cada grupo de estudiantes envió su conversación al docente de la asignatura, donde quedaron registrados todos los prompt.

La tarea inicial consistió en redactar un prompt para diseñar una actividad de enseñanza de química a nivel primario, sin guía ni ejemplos previos. En la fase POS, realizada tras sesiones formativas sobre diseño didáctico y prompting educativo, se solicitó a los mismos grupos que reformularan su propuesta incorporando los aprendizajes adquiridos. Las instrucciones se entregaron digitalmente por medio del campus virtual, junto con ejemplos comentados y rúbricas orientativas.

Posteriormente, se hizo una revisión de algunas de las conversaciones, señalando aspectos de mejora. Después, se hizo una actividad del tipo *role model*, donde el docente de la asignatura sirve como un ejemplo a los estudiantes y realiza una conversación con ChatGPT para diseñar una actividad cualquier, en este caso sobre gravitación para estudiantes de primaria. Los estudiantes veían en directo las preguntas y las respuestas, y se iban comentando *in situ*. La actividad duró aproximadamente 1 hora. La asistencia fue del aproximadamente 60% de estudiantes, pero la conversación final fue también enviada a los estudiantes.

Más adelante en el curso se pidió de nuevo a los estudiantes que diseñaran con ChatGPT una actividad didáctica sobre química de no más de 30 minutos para niños (actividad POS), centrada en temas como los cambios de estado o las reacciones químicas, y usar materiales simples. Cada grupo se le daba un tema, como la expansión de los sólidos, los cambios de temperatura, los cambios de estado, las mezclas y disoluciones, las reacciones químicas, etc. El resultado debía ser una infografía que resumiera las ideas científicas y didácticas del experimento, acompañada de un enlace a la interacción con ChatGPT que ayudó a su diseño, que fue enviada al docente del curso.

En relación con el proceso de construcción de los prompt, cabe señalar que este se abordó pedagógicamente en dos etapas diferenciadas. El proceso pedagógico se desarrolló en dos etapas. En la fase PRE, los estudiantes formularon sus preguntas libremente, sin guía metodológica ni estructura definida, lo que permitió explorar sus concepciones iniciales sobre el uso de IAG en el diseño didáctico. Luego, se llevó a cabo una intervención formativa basada en la estrategia de *role model*, en la que el docente ejemplificó, en tiempo real, una conversación didáctica con ChatGPT, destacando aspectos clave como la definición de objetivos, la anticipación de errores conceptuales y el uso del andamiaje.

Sistema de recogida de los datos

Los datos fueron recogidos de dos fuentes distintas: las conversaciones que los estudiantes tuvieron con el ChatGPT en las fases inicial (PRE) y final (POS), y en los registros de la actividad del tipo *role model* hecha por el docente de la asignatura, y que tuvo lugar entre las fases PRE y POS. Estas conversaciones se pasaron a un archivo de Excel, donde se asignó a cada prompt a una fila individual y un código identificativo, resultando en 223 prompts analizados.

Los grupos de estudiantes almacenaron sus primeras interacciones con el chatbot en documentos compartidos, que incluían tanto los prompt como las respuestas generadas por la IAG. Posteriormente, el equipo docente analizó estas interacciones mediante una rúbrica centrada en la claridad, relevancia didáctica y adecuación de las respuestas.

En el curso había 15 grupos de trabajo, donde se compilaron inicialmente un total de 29 conversaciones, 14 de la actividad PRE y 15 de la actividad POS. Como uno de los grupos no envió su enlace de la actividad PRE en tiempo, todos los datos de este grupo fueron borrados. Consideramos entonces 28 enlaces a las conversaciones que los estudiantes habían tenido con ChatGPT, (14 PRE y 14 POS),

Las actividades desarrolladas por los estudiantes fueron entregadas exclusivamente en formato digital mediante el campus virtual institucional (Moodle), junto con los registros de

sus interacciones con el chatbot. No se contemplaron entregas impresas. Los temas asignados a los grupos correspondían a contenidos fundamentales del currículo de química para educación primaria en España. A continuación, se enumeran su frecuencia de aparición en los prompt estudiantiles: a) Propiedades de los materiales (39), b) Estados de la materia (25), c) Mezclas y disoluciones (8), d) Reacciones químicas simples (7), Cambios físicos y químicos (5), Ácidos y bases en la vida cotidiana (3)

Para la organización de los datos fueron creados códigos de identificación. Un ejemplo es G5PRE01, donde los dos primeros dígitos (G5) es una referencia al grupo, los tres próximos (PRE) que se refiere a la primera actividad de construcción del prompt realizada en el momento previo, y los dos últimos dígitos (01) que es el primero prompt creado por el equipo. Cada uno de estos prompt se consideró una unidad de análisis independiente, donde se hace un análisis literal del contenido. Es importante destacar que el estudio se centró en las características de las interacciones del prompt, omitiendo el análisis de la calidad de las respuestas proporcionadas por el ChatGPT.

Durante la sesión de *role model*, el docente ejemplificó el uso reflexivo de ChatGPT en el diseño de una actividad didáctica de química, mientras los estudiantes observaban y analizaban el proceso. Se capturaron imágenes de las interacciones, que luego fueron codificadas con base en una matriz de categorías previamente definida en contenidos del prompt y función comunicativa. La codificación fue realizada de forma independiente por dos investigadores, con sesiones posteriores de consenso.

Estrategia del análisis de los datos

Para este estudio exploratorio utilizaremos de un enfoque analítico mixto, en una perspectiva cuali-cuantitativa comparada. La adopción de este enfoque se justifica por el convencimiento de que ambos enfoques pueden complementarse y beneficiarse mutuamente (Albert Gómez, 2007; Creswell & Creswell, 2018).

Pasados al fichero de Excel, en primer lugar, se realizó un recuento del número de caracteres por prompt y el número de prompt por conversación, y los datos fueron cuantificados por medio del test de Wilcoxon, que es una prueba no paramétrica de comparación del rango medio de dos muestras relacionadas (en nuestro caso, PRE y POS) con el objetivo de determinar si hay diferencias entre ellas (Lawson et al., 2019).

En segundo lugar, se hizo un análisis cualitativo de las conversaciones mediante dos dimensiones independientes: el contenido que trataba cada prompt y la funcionalidad del prompt en la conversación, tal se muestra en la tabla 1.

Para asegurar la fiabilidad del proceso, los dos investigadores realizaron codificaciones de forma independiente. Para facilitar el proceso de análisis, se usaron diagramas de colores como el de Figura 3, que permitía tener una visión global de toda conversación. Antes de empezar el proceso de análisis, las codificaciones de ambos investigadores fueron comparadas y los casos donde había diferencias fueran discutidos y ajustados.

Tabla 1. Sistema de categorías usado para el análisis de los datos.

Dimensión	Categoría	Descripción de la categoría
Contenido del prompt	Contenido científico (CC)	El prompt incluye temática científica, disciplina, fenómeno científico, ideas abstractas (vectores, partículas, fuerzas, células, etc.).
	Contenido didáctico (CD)	El prompt incluye ideas de los estudiantes, preguntas, estrategia de evaluación, andamiaje, dificultades, objetivos de aprendizaje, etc.
	Contexto educativo (CE)	El prompt incluye preguntas o comentarios sobre la edad de los estudiantes, lugar, duración, actividades previas y posteriores.
	Aspectos prácticos (AP)	El prompt incluye preguntas o comentarios sobre materiales de laboratorio, instrumentos y dispositivos, limitaciones técnicas, organización de los grupos, temporización, etc.
	Otros (O)	El prompt incluye preguntas o comentarios sobre aspectos formales de la conversación, Cambio de idioma, resumen, tablas, etc.
Funcionalidad del prompt dentro del diálogo	Adición (ADI)	El prompt implica la adición de nuevas informaciones sin hacer referencia explícita a la conversación previa.
	Expansión (EXP)	El prompt implica la expansión o evolución en una información que ya existe en la conversación previa.
	Modulación y refinamiento (MOD)	El prompt implica cambios o refinamiento a una información anterior.
	Refutación y rechazo (REF)	El prompt implica la refutación o impugnación total de una información generada, a menudo con el objetivo de generar otra completamente nueva.

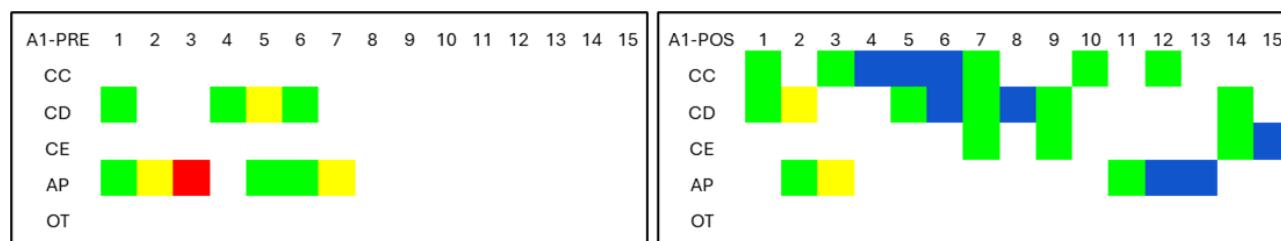


Figura 3. Ejemplos de diagrama usados en el grupo G1 para codificar las conversaciones del grupo.

Como ejemplo de nuestra clasificación de datos, en el diagrama de la Figura 3, a la izquierda aparece la conversación PRE, y a la derecha la POS. Los números de un a quince son, en orden, el número secuencial de prompt utilizado por los estudiantes para interacción con el ChatGPT. Los colores utilizados son, el verde corresponde a la categoría Adición, el azul a Expansión, el Amarillo a Modulación, y el rojo a Refutación. El eje horizontal representa el orden de los prompt dentro de la conversación. Así, a la izquierda del diagrama se puede observar como el primer prompt incluye una pregunta de adición (verde) relacionada con Contenido Didáctico (CD) y también sobre Aspectos Prácticos (AP). Reforzamos que un mismo prompt podía tener varios contenidos, tal como se muestra en la Figura 3, por ejemplo, en el segundo prompt de la actividad POS (G1POS02), donde hay el contenido didáctico (CD) en amarillo, y los aspectos prácticos (AP) en verde.

Aun se observa en este ejemplo de la Figura 3 que la conversación PRE tuvo 7 prompt y 10 contenidos identificados, y la POS tuvo 15 prompt y 25 funcionalidades. Así, es fundamental aclarar aunque existe una diferencia importante entre las dimensiones, pues, mientras la dimensión “Contenido del prompt” no es excluyente, es decir, una misma unidad de análisis puede tener diferentes categorías asociadas ya que en un mismo prompt puede haber diferentes informaciones, la dimensión “Funcionalidad del prompt dentro del dialogo” es

excluyente, ya que se considera que cada prompt sirve para una única funcionalidad en la estructura de la conversación, o bien que existe una funcionalidad que predomina sobre las demás.

Otra consideración metodológica importante es que la secuencia en que aparece la información puede alterar su clasificación, ya que un prompt con un mismo contenido puede tener diferentes funcionalidades. Por ejemplo, en la solicitud “para niños de 5 i 6 curso” (G2PRE03), si se sitúa al principio de la conversación, se consideraría una funcionalidad de Adición (ADI), mientras que, si el usuario ya ha estado solicitado al ChatGPT diferentes actividades y luego incluye este prompt, se considera una actividad de refinamiento (REF). Es decir, que el análisis de la funcionalidad dependerá del contexto y del orden en que aparece durante la conversación.

Resultados y Discusión

Los datos recolectados permitieron responder a las preguntas planteadas mediante un análisis dividido en cuatro focos principales. Empezamos nuestra discusión presentando los resultados en la Tabla 2 de un recuento de la extensión de los prompt a través del número de caracteres en cada uno de ellos, la frecuencia de identificación del contenido de los prompt, donde cada interacción fuera clasificada en relación con el contenido del prompt (CC, CD, CE y AP) y en relación con la funcionalidad del prompt dentro del diálogo (ADI, EXP, MOD, REF)

Tabla 2. Recuentos de características de los prompt.

Características de los grupos			Cantidades		Promedio		Contenido de los prompt				Funcionalidad de los prompt			
Id.	Fase	Cambio de lenguaje	De Prompt	Caracteres en los Prompt	Caracteres en cada prompt	Contenidos en cada prompt	CC	CD	CE	AP	ADI	ESP	MOD	REF
G1	PRE	No	7	883	126	1,43	0	4	0	6	6	0	3	1
	POS	Catalán	15	1995	133	1,67	8	8	4	5	14	8	2	0
G2	PRE	Castellano	6	796	133	1,17	3	2	1	1	6	0	1	0
	POS	No	5	1238	248	1,40	1	2	1	3	5	0	2	0
G3	PRE	No	8	1474	184	1,25	4	4	2	0	6	0	4	0
	POS	No	7	537	77	1,28	1	1	3	4	6	1	2	0
G4	PRE	No	8	986	123	1,50	2	0	5	5	8	0	4	0
	POS	No	4	1914	479	3,25	3	3	3	4	8	1	3	1
G5	PRE	No	3	516	172	1,67	2	0	1	2	3	0	2	0
	POS	No	5	1264	253	2,40	3	3	3	3	8	4	0	0
G6	PRE	No	3	373	124	1,33	0	1	1	2	3	0	1	0
	POS	No	5	625	125	1,40	5	0	2	0	4	3	0	0
	PRE	No	3	524	174	1,33	1	0	2	1	3	0	1	0

Características de los grupos			Cantidades		Promedio		Contenido de los prompt				Funcionalidad de los prompt			
Id.	Fase	Cambio de lenguaje	De Prompt	Caracteres en los Prompt	Caracteres en cada prompt	Contenidos en cada prompt	CC	CD	CE	AP	ADI	ESP	MOD	REF
G7	POS	No	10	546	55	1,50	4	1	4	6	6	1	6	2
G8	PRE	No	10	1291	129	1,50	1	1	6	7	11	0	2	2
	POS	No	15	1092	73	1,40	8	1	5	7	13	5	3	0
G9	PRE	No	5	630	126	1,40	2	1	2	2	7	0	0	0
	POS	No	3	383	128	3,33	3	2	3	2	4	3	3	0
G10	PRE	No	5	4232	846	2,60	3	4	4	2	11	0	2	0
	POS	No	6	806	134	1,33	3	3	1	1	5	1	2	0
G11	PRE	Castellano	9	550	61	1,33	1	1	2	8	7	0	3	2
	POS	No	11	1946	177	1,45	2	7	1	6	11	2	2	1
G12	PRE	No	12	739	62	1,33	0	0	7	9	9	0	6	1
	POS	No	9	1517	169	1,22	3	7	1	0	4	6	1	0
G13	PRE	Catalán	15	1137	76	1	7	3	2	3	12	0	0	3
	POS	No	15	2995	200	1,33	7	5	6	2	12	7	1	0
G14	PRE	No	10	2233	223	1	1	2	3	4	7	0	3	0
	POS	No	9	894	99	0,89	4	2	1	1	7	1	0	0

A partir de la clasificación inicial de estos datos derivados de los prompt, proponemos una discusión basada en cuatro puntos principales: ¿Qué observamos?, ¿Qué cambia entre PRE y POS?, ¿Qué patrón general encontramos? y ¿Estas diferencias son significativas?

¿Qué observamos?

Primero, que en las fases PRE o POS hay una estabilidad numérica con relación a los prompt creados por los grupos, y que casi no hay cambios en el lenguaje de la interacción, a pesar de esta asignatura ocurrir en lengua inglesa y de la lengua materna de la mayoría de los grupos no ser el inglés.

Para una visualización de las características de los prompt generados, proponemos una simplificación de los datos de la tabla 2, considerando así que cada prompt tiene un cierto “tamaño”, o sea, una longitud que indica la cantidad de sus caracteres y promedios (columnas 5 y 6), y un espesor que indica la frecuencia y variación de contenidos (columna 8 hasta 11). En esta propuesta, se observó que tanto al respecto de la longitud de los prompt y de la conversación, y también su espesor, que hay también una estabilidad en sus recuentos y promedios.

En cuanto a las complejidades de los prompt, o sea, sus contenidos y sus funcionalidades, observamos una estabilidad en la frecuencia de los temas CE (38/37), una disminución parcial en AP (52/ 44), y un marcado aumento en la concentración en CC (27/52) y CD (23/46). Sobre el contenido de la temática de los prompt CE y AP, en general observamos que

estos incluyan preguntas o comentarios sobre la edad de los estudiantes, lugar, duración, materiales de laboratorio, organización de los grupos, temporización, o sea, muy relacionadas con la logística de las actividades. En cambio, los prompt clasificados como CC y CD abordan temas más complejos que tratan de las temáticas científicas, fenómenos científicos, ideas abstractas, ideas de los estudiantes, estrategias de evaluación y andamiaje.

Con el cruzamiento de las observaciones entendemos que el aumento en la concentración de CC y CD de casi el doble entre las fases PRE y POS está relacionado a un cambio en las interacciones de los estudiantes por medio de los prompt, ahora más profundizado y enfocado en la comprensión y enseñanza de conceptos clave que en la logística de las actividades. Así, indicamos que más que duplicar el número de interacciones de los temas más complejos CC y CD, hay un marcado cambio en la calidad del tipo de las interacciones.

En general, sobre la diversidad de estilos de prompt observamos que no hay un padrón identificado, pues algunos son más cortos y otros son mucho más largos; otros son claramente más simples (números en la columna 7 más pequeño) y otros más complejos (números en la columna 7 más grande) pues existen referencias a más de un contenido. Pero, aunque exista variaciones entre los grupos y entre las fases PRE y POS, no se identificaron cambios significativos atribuibles directamente a la actividad, aunque sí se observaron mejoras en la calidad estructural y funcional de los prompt en sus características, pero en sus complejidades hubo cambios significativos tanto en sus contenidos como en sus funcionalidades.

¿Qué cambia entre PRE y POS?

Para visualizar que pasó entre las fases PRE y POS, proponemos en la Figura 4 una clasificación secuencial de cada uno de los prompt en acuerdo al ejemplo de la Figura 3, pues así lograremos mirar los principales cambios entre los tipos de interacciones de cada uno de los grupos.

La idea central de este tipo de propuesta de visualización es tener un cuadro general temporal, secuencial y comparativo de las tipologías de acciones de los grupos con futuros maestros cuando son solicitados a interactuar con una IAG para la ejecución de una actividad práctica con relación directa a su formación y su futura actuación profesional.

Nuestra propuesta así es identificar los cambios y variaciones en general de todos los grupos juntos sin mirar datos o especificidades cada uno de ellos.

Primero, hay una prevalencia de los colores verde (ADI) en ambos escenarios PRE y POS, y la presencia del color rojo (REF) es más evidente de forma concentrada en pocos grupos, principalmente en PRE. También es muy visual que la fase POS son más largas, más llenas y más colorida que en PRE, indicando cambios en las interacciones entre las dos fases.

Hay una presencia más del azul (EXP) en la fase POS, generando una mayor variación de colores. También hay un desplazamiento de zonas coloradas hacia CC y CD (en la parte superior de cada recuadro). De una manera un poco más tímida, se constata una mayor

distribución de interacciones hacia los recuadros situados a la derecha en las actividades de la fase POS.



Figura 4. Categorías utilizadas para el análisis de datos. Los rectángulos en gris indican prompt no clasificados por presentar ambigüedad o irrelevancia respecto a las categorías analíticas definidas

Esta visualización indica que hay cambios individuales y generales en las interacciones de los grupos entre PRE y POS, principalmente con una presencia más distribuidas de la funcionalidad de expansión (EXP) y una mejor ocupación de los recuadros CC y CD en las interacciones POS.

Aunque este estudio se centró en los prompt generados por los estudiantes, observamos que las respuestas ofrecidas por ChatGPT en la fase POS fueron más pertinentes y estructuradas, incluyendo descripciones elaboradas de actividades, sugerencias de evaluación y referencias a ideas previas del alumnado. Esta retroalimentación pareció favorecer interacciones más orientadas en términos pedagógicos. No obstante, es preciso también matizar estos resultados desde una perspectiva crítica, pues, como advierte Feenberg (2002, 2018), las tecnologías no son neutras: incorporan valores y pueden reconfigurar las prácticas educativas según lógicas externas al campo pedagógico. Así, si bien las respuestas de la IAG pueden contribuir a una mejora formal del discurso instruccional, existe el riesgo de que orienten la enseñanza hacia formatos estandarizados y tecnocráticos, priorizando la eficiencia sobre la reflexión crítica. Así, el uso acrítico de herramientas como ChatGPT podría debilitar procesos clave de coautoría, problematización y construcción colectiva de significado en el aula.

¿Qué patrón general encontramos?

Para obtener una visión aún más sintética y global de los cambios entre las fases PRE y POS relacionadas a las funcionalidades del prompt (ADI, EXP, MOD, REF) en cada uno de los contenidos (CC, CD, CE, AP) en los diálogos de los grupos, hicimos una sobreposición de los datos utilizando el código de colores adoptados (Figura 3), donde todos los datos de la Figura 4 han sido sobrepuertos en sus dimensiones, y que nos resultó en la Figura 5.

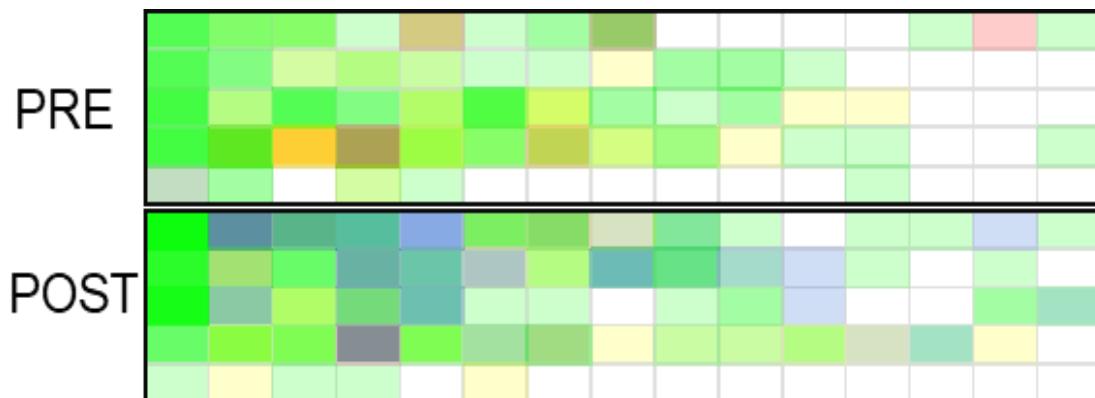


Figura 5. Sobreposición de las dimensiones en colores

La mezcla de cuadros reafirma los hallazgos anteriores. En particular, el cuadro POS está mucho más lleno y con mayor variación de colores que en cuadro PRE, reafirmando que, en general, hubo algún avance (cuál y cuánto) en las interacciones de los grupos, con más y mejores tipos de interacciones.

En relación a la mezcla de colores hemos detectado, en un primero escenario, una presencia del verde (ADI) en ambos PRE y POS. En las interacciones PRE, debido la presencia de amarillo (MOD) y algunos rojos (REF), hay una prevalencia del amarillo y marrón (combinación de verde, amarillo y rojo). Por su parte, en las acciones POS, debido a una destacada presencia del azul (EXP), hay una prevalencia de variaciones del azul y violeta en la combinación de colores. Esta visualización nos permitió comparar el patrón general

encontrado en las fases PRE y POS, pero también, sus matices de colores han indicado variaciones más restrictivas en PRE que en POS.

En términos generales, si bien el análisis cuantitativo indicó un leve aumento en la longitud media de los prompt en la fase POS, dicho incremento fue contextualizado por una evaluación cualitativa centrada en el contenido y la funcionalidad de las interacciones. No se asumió que una mayor extensión implicara automáticamente mejor calidad; al contrario, se identificaron indicios de un uso más estratégico y pedagógico de la herramienta, evidenciado en el empleo explícito de conceptos científicos clave, formulación precisa de objetivos y atención a variables didácticas como ideas previas del alumnado, evaluación y organización temporal de la secuencia didáctica. En este sentido, aunque los prompt fueron más extensos en la fase POS, lo relevante fue la mejora cualitativa en términos científicos y didácticos.

Uno de los indicadores más reveladores fue el incremento de la funcionalidad de expansión (EXP), ausente en la fase PRE, lo que sugiere una mayor capacidad por parte de los estudiantes para profundizar, refinar y contextualizar sus preguntas y el diálogo con la IAG. De este modo, esta evolución estuvo directamente asociada a la intervención formativa, que ejemplificó prácticas de prompting pedagógico eficaces, como la especificación del nivel educativo, la secuenciación de preguntas, la anticipación de errores conceptuales y la solicitud de salidas organizadas, tales como rúbricas o tablas.

No obstante, estos avances en la calidad y diversidad funcional de los prompt no deben ser interpretados de forma exclusivamente positiva o lineal. Tal como advierten Feenberg (2002, 2018) y Avraamidou (2024), la incorporación de tecnologías emergentes como la IAG en contextos educativos puede generar tensiones entre la estandarización algorítmica del conocimiento y la diversidad de perspectivas epistémicas y pedagógicas. El aumento de prompt de tipo “expansivo” y la integración de más contenidos científicos y didácticos no garantizan, por sí solos, una mejora crítica en los procesos de enseñanza-aprendizaje. Existe el riesgo de que los estudiantes adopten estrategias de prompting cada vez más eficaces desde el punto de vista técnico, pero desconectadas de una reflexión profunda sobre el papel del conocimiento escolar, la equidad cognitiva o las finalidades educativas.

¿Estas diferencias son significativas?

Para verificar si las diferencias son significativas, empezamos verificando estadísticamente nuestras Dimensiones y Categorías, proponiendo en la tabla 3 una compilación con el test de Wilcoxon.

En observación de la tabla 3 verificamos que en la mayoría de las categorías no hay diferencias significativas, excepto en tres casos:

- 1) en la dimensión “contenido del prompt”, la categoría del contenido científico (CC) aumenta.

Es importante destacar que este aumento fue numéricamente significativo, tanto en su generalidad cuando hay un marcado aumento de 27 (PRE) para 55 (POS), como en la individualidad de los grupos. Por ejemplo, en la fase PRE hay tres grupos que no proponen

nada en compatibilidad a la categoría CC, pero en la fase POS todos los grupos proponen algo en esta categoría.

Tabla 3. Recuento y estadística de las características de los prompt

Dimensión y categoría		PRE		POS		Wilcoxon	
Dimensión	Categoría	Total	Promedio	Total	Promedio	Estadístico	valor P
Aspectos descriptivos	# prompt	104	7,4	119	8,5	39,0	0,426
	# caracteres	2560	171	2347	167	33,0	0,379
Contenido del prompt	CC	27	1,8	55	3,9	12,5	0,036
	CD	23	1,5	46	3,3	11,0	0,049
	CE	38	2,5	37	2,6	44,0	0,916
	AP	52	3,5	44	3,1	36,5	0,526
Funcionalidad del prompt	ADI	101	7,2	106	7,6	30,5	0,824
	MOD	32	2,3	29	2,1	43,5	0,626
	REF	9	0,6	4	0,3	8,0	0,301
	EXP	0	0	45	3,2	0,0	0,002

Cualitativamente también hay un marcado cambio en la estructura de los Prompt propuestos. Por ejemplo, el G5 empiezan preguntando a la IAG: “*can you design an activity to explain that air is composed by small parts?*” (A5PRE02), y en la actividad POS evolucionan para:

“Imagine you are a competent primary school teacher expert in experimental sciences didactic. You've been tasked in creating an experiment in the properties of liquids and gases (diffusion, expansion, temperature) of (10-15 minutes maximum). Give me a list of proposals that seem appropriate for ages 9-11, take into account the typical misconceptions and previous ideas that children this age can have. The detailed description of the activity. The scaffolding provided by teacher (questions, explanations, examples, etc.). The scientific key ideas to be constructed with the activity, and a justification of why this activity is valuable to construct those ideas.” (G5POS01).

En suma, los datos evidencian una evolución entre los prompt PRE y POS, tanto en la cantidad como en la calidad de las temáticas y fenómenos científicos abordados, conceptos científicos, ideas abstractas etc.

2) en la dimensión “contenido del prompt” la categoría del contenido didáctico (CD) aumenta.

En esta categoría CD hay un aumento del doble (23 a 46) en relación con la frecuencia general de los grupos, y en sus individualidades, miramos que en la fase PRE hay cuatro grupos que no tienen Prompt clasificados en la categoría CD, pero en la fase POS, solamente uno de los grupos no tiene Prompt clasificados en esta categoría.

También hay un cambio en las estructuras de los prompt propuestos, por ejemplo, en el grupo G10, cuando empiezan preguntando en su prompt “*Give me ideas of activities that I can do with my 3rd grade students in order to work the properties of solids in science class*” (G10PRE01), y después evolucionan para:

“I'm a primary school teacher, and I want to develop some teaching scenarios related physical changes in chemistry. First of all, I need to clarify what and why teaching physical changes in

chemistry in primary school. I'm interested in topic [...] from the list of teaching scenarios. Could you develop it a bit more? I would like to identify what does my students think about the topic [...]. Which could be the expected misconceptions and confusions from my students?" (G10POSO102)

En este ejemplo de marcado cambio, hay elementos numéricos y cualitativos a partir de los datos que indican la presencia del contenido didáctico (CD), donde identificamos principalmente un aumento de temática sobre las ideas de los estudiantes, de andamiaje y estrategia de evaluación, de los objetivos de aprendizaje, etc.

3) en la dimensión “funcionalidad del prompt” la categoría Expandir (EXP) aumenta.

En la categoría expansión (EXP) de la dimensión sobre las funcionalidades del Prompt hay un destacado aumento en criterios individuales y general de los grupos, ya que en estos casos no hay ningún prompt en la fase PRE y pasan a existir 45 prompt categorizados en la fase POS, y cuando solo uno de los grupos no ha elaborado prompt en la categoría EXP.

Por ejemplo, el grupo G11 empieza su conversación con el ChatGPT:

“Hi I'm a primary school teacher and next week I will hold a class about weak changes in matter, so change of states. First of all I would like to truly understand why should we teachers teach change of states, second i would like to know the common believing of kids about this topic, and their common mistakes or misconceptions” (G11POSO1).

Hasta que la conversación se expande, después de la respuesta dada por la IAG para

“OK thank you, but could you explain more in depth the points 3 and 4? After that if you can, could you do a table with misconceptions and next to it a simple way to make them understand that what they are thinking is wrong.” (G11POSO2).

Después de un avance en la conversación, y también con el análisis del G11 sobre las respuestas propuestas por el ChatGPT, hay un nuevo movimiento de expansión del prompt de este grupo para:

“going back to the table that you made for me before... I would like that you add one more line where you explain, in a few sentences, how would you correct their misconception. I don't want any more experiments, i want sentences to correct them, and make them understand easily.” (G11POSO8).

Así, de forma muy contundente, identificamos un destacado aumento de expansión (EXP) en las funcionalidades de los prompt, cuando también hubo una clara evolución en las informaciones que ya existían en la conversación previa.

Esta evolución en la elaboración de prompt permite establecer una conexión con los componentes del modelo TPACK. De forma empírica, se observa cómo los futuros docentes comienzan a integrar progresivamente los tres núcleos de conocimiento: el conocimiento del contenido disciplinar (CK), evidenciado en los prompt centrados en fenómenos científicos y conceptos abstractos; el conocimiento pedagógico (PK), presente en aquellos prompt que exploran objetivos de aprendizaje, estrategias de evaluación o ideas previas del alumnado; y el conocimiento tecnológico (TK), implícito en el uso eficaz de la IAG como herramienta para el diseño didáctico. Esta intersección de saberes sugiere el desarrollo de una competencia

situada que no es meramente instrumental, sino que articula la comprensión disciplinar, la intención pedagógica y el aprovechamiento crítico de la tecnología.

Más allá de los resultados específicos observados en los contenidos y funciones de los prompt, este estudio aporta una contribución al campo de la investigación educativa al proponer el análisis de los prompt generados en interacciones con IAG como unidades significativas de indagación pedagógica. Esta elección metodológica permite no solo observar qué se pregunta y cómo, sino también reconstruir los procesos de pensamiento didáctico y disciplinar activados por los futuros docentes.

Conclusiones

Los resultados obtenidos evidencian que el uso formativo de herramientas de IAG como ChatGPT puede enriquecer el diseño didáctico de futuros maestros. A partir de los resultados presentados, pueden identificarse tres implicaciones principales derivadas del estudio: formativas, investigativas y tecnológicas.

En el plano formativo, esta investigación refuerza la necesidad de integrar prácticas de prompting pedagógico en los programas de formación inicial del profesorado, articulando los saberes didácticos, científicos y tecnológicos en el marco del modelo TPACK. Desde una perspectiva investigativa, los datos obtenidos subrayan la importancia de profundizar en el estudio de los procesos cognitivos y metacognitivos que se activan durante la interacción con tecnologías de IAG. Futuras investigaciones podrían explorar, por ejemplo, el papel de las emociones académicas, las creencias epistemológicas o las estrategias autorregulatorias en el uso educativo de estas herramientas.

Finalmente, en el ámbito tecnológico y ético-pedagógico, la creciente presencia de las IAG en los contextos escolares exige una incorporación que no se limite a usos instrumentales, sino que esté mediada por marcos críticos que contemplen los riesgos de tecnocratización, estandarización y desdibujamiento de la autoría docente. La construcción de competencias para el análisis crítico de la mediación algorítmica se convierte, así, en una dimensión indispensable de la formación docente contemporánea.

Aunque este estudio utilizó una herramienta específica (ChatGPT 3.5), su alcance no se restringe a esta plataforma ni a sus procesos internos. Más bien, se centra en las interacciones entre futuros maestros y diferentes tecnologías de IAG a través de sus prompt, considerando que muchas otras herramientas ya existen o serán desarrolladas en un futuro próximo. En este escenario, los procesos de interacción, cada vez más semejantes a las formas conversacionales humanas, requerirán investigaciones más específicas y profundas.

Con el objetivo de analizar el impacto de la incorporación de IAG en la formación de maestros de ciencias en la educación primaria, se examinaron 223 prompt generados en 28 conversaciones con ChatGPT. Aunque la mayoría de los parámetros estructurales se mantuvieron constantes, se observó una transformación significativa en los contenidos científicos y didácticos, lo que indica una mejora cualitativa sin necesidad de incrementar el volumen de interacción. Este hallazgo sugiere que los estudiantes lograron emplear de

manera más eficaz la IAG, concentrando sus esfuerzos en profundizar y expandir los conceptos clave en lugar de aumentar la cantidad de prompt.

Estos resultados revelan no solo un cambio técnico en el uso del chatbot, sino también un proceso formativo más amplio, en el cual los futuros docentes reflexionan sobre el diseño didáctico, la selección de contenidos y la adaptación de recursos digitales. El uso tutorizado de la IAG favoreció el desarrollo de una competencia metadidáctica, al permitir que los estudiantes pensaran críticamente sobre qué enseñar, cómo enseñar y cómo anticipar errores conceptuales frecuentes. Tal como muestran otros estudios (Tenà & Couso, 2023; Mishra et al., 2023), este enfoque se alinea con una perspectiva crítica que aboga por la integración reflexiva de las tecnologías emergentes en la formación docente.

La evolución observada en la construcción de prompt refleja tanto un avance técnico como una transformación en las concepciones pedagógicas subyacentes. Estos resultados dialogan con investigaciones previas (Holmes et al., 2019; Tondeur et al., 2017) que subrayan la importancia de una apropiación crítica de la tecnología en la formación inicial del profesorado. En el contexto de la enseñanza de las ciencias, esto implica preparar al profesorado no solo para utilizar herramientas tecnológicas, sino también para reflexionar sobre su papel en la construcción del conocimiento.

En este sentido, la inclusión crítica de la IAG en la formación de maestros, tanto en términos generales como específicamente en la didáctica de las ciencias, puede promover interacciones más ricas, que ayuden a superar barreras persistentes en el desarrollo del conocimiento didáctico del contenido (PCK), como la clarificación de conceptos clave, la identificación de ideas previas del alumnado y el uso del andamiaje.

A pesar de que solo se identificaron diferencias estadísticamente significativas en tres categorías analíticas, dos en la dimensión de contenido del prompt (CC y CD) y una en la dimensión funcional (EXP), estas dimensiones son especialmente relevantes para el desarrollo de habilidades complejas de diseño didáctico. Además, el incremento observado en CC, CD y EXP ocurrió sin disminución de otras categorías, lo que sugiere que la evolución cualitativa de los prompt no compromete dimensiones de soporte, sino que las enriquece.

A la luz de los resultados obtenidos, los hallazgos de este estudio permiten vislumbrar implicaciones relevantes para el diseño curricular de la formación docente en ciencias. La mejora cualitativa observada en las interacciones con la IAG, en términos tanto científicos como didácticos, sugiere la necesidad de incorporar de manera estructural en los programas de formación inicial espacios formativos específicos sobre el uso crítico y pedagógico de tecnologías generativas, no solo el desarrollo de habilidades técnicas de prompting, sino también una reflexión situada sobre las finalidades educativas, la autoría, la mediación y los efectos epistémicos de estas herramientas en contextos escolares diversos.

Este estudio, sin embargo, presenta algunas limitaciones. Aunque la muestra fue numéricamente representativa, todos los participantes pertenecían a una misma universidad, lo que restringe la generalización de los hallazgos. Asimismo, la duración limitada de la intervención pudo incidir en la profundidad de los cambios observados. Además, no se analizaron de manera sistemática las respuestas generadas por la IAG, ni se recopilaron datos

sobre los procesos cognitivos o emocionales de los estudiantes durante las interacciones. Tampoco se abordaron posibles diferencias vinculadas al género, idioma o experiencia tecnológica previa. Para mitigar estas limitaciones, se aplicó una triangulación metodológica basada en el análisis detallado de los productos generados. Futuras investigaciones podrían incluir entrevistas, cuestionarios u otros enfoques multimodales que amplíen la comprensión crítica del papel que estas tecnologías pueden desempeñar en la construcción de una práctica docente reflexiva, ética y situada.

Agradecimientos

Esta investigación se ha desarrollado en el marco del proyecto PID2022-138166NB-C22b financiado por el Ministerio de Ciencia, Innovación y Universidades de España, con el Apoyo del grupo de investigación ACELEC 2021SGR00647, y del grupo de investigación Tecnologias e Cultura Digital na Educação em Ciências (TeCDEC/Unifei).

Referencias

- Albert Gómez, M. J. (2007). *La investigación educativa: Claves teóricas*. McGraw-Hill.
- Altarawneh, H. (2023). ChatGpt impact on Student Educational Performance: A conceptual analysis. *EAI Endorsed Transactions on e-Learning*, 9. <https://doi.org/10.4108/eetel.4574>
- Avraamidou, L. (2024). Can we disrupt the momentum of the AI colonization of science education? *Journal of Research in Science Teaching*, 61(10), 2570–2574. <https://doi.org/10.1002/tea.21961>
- 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>
- Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *Contemporary Educational Technology*, 15(3), ep430. <https://doi.org/10.30935/cedtech/13176>
- Bommarito, J., Bommarito, M. J., Katz, J., & Katz, D. M. (2023). Gpt as knowledge worker: A zero-shot evaluation of (Ai)cpa capabilities. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4322372>
- Bordt, S., & Luxburg, U. von. (2023). From shapley values to generalized additive models and back. *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, 709–745. <https://proceedings.mlr.press/v206/bordt23a.html>
- Candela, A. (1998). Students' power in classroom discourse. *Linguistics and Education*, 10(2), 139–163. [https://doi.org/10.1016/S0898-5898\(99\)80107-7](https://doi.org/10.1016/S0898-5898(99)80107-7)
- Castiglione, A. (2023). Per una Pedagogia della Singolarità: Intelligenze artificiali e tecnologie digitali a supporto dell'educazione alla scrittura, un quasi-esperimento con il modello linguistico GPT-3. *Graphos. Rivista internazionale di pedagogia e didattica della scrittura*, 3, 87–105. <https://doi.org/10.4454/graphos.38>
- Čavojský, M., Bugár, G., Kormaník, T., & Hasin, M. (2023). Exploring the capabilities and possible applications of large language models for education. *2023 21st International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 91–98. <https://doi.org/10.1109/ICETA61311.2023.10344166>
- Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (Ai)-based tools into education. *Computers in Human Behavior*, 138, 107468. <https://doi.org/10.1016/j.chb.2022.107468>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (Fifth edition). SAGE.

- Eraña, A. B. (2024) Entonces, ¿qué necesito saber cómo docente sobre IA? In: A. A. Sagasta. (Coord.). *Inteligencia artificial y educación: construyendo puentes*. (pp. 29-40). Graó.
- Feenberg, A. (2002). *Transforming technology: A critical theory revisited* (2nd ed). Oxford University Press USA - OSO.
- Feenberg, A. (2018). *Technosystem: The social life of reason*. Harvard University Press.
- Feldman-Maggot, Y., Blonder, R., & Alexandron, G. (2025). Perspectives of generative ai in chemistry education within the tpack framework. *Journal of Science Education and Technology*, 34(1), 1–12. <https://doi.org/10.1007/s10956-024-10147-3>
- Fuchs, K., & Aguilos, V. (2023). Integrating artificial intelligence in higher education: Empirical insights from students about using chatgpt. *International Journal of Information and Education Technology*, 13(9), 1365–1371. <https://doi.org/10.18178/ijiet.2023.13.9.1939>
- Generalitat de Catalunya, Departament d'Educació. (2024). *La intel·ligència artificial en l'educació: Orientacions i recomanacions per al seu ús als centres*. <https://educacio.gencat.cat/web/.content/home/departament/publicacions/monografies/mon-digital/inteligencia-artificial-educacio/ia-educacio.pdf>
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does chatgpt perform on the united states medical licensing examination (Usmle)? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9, e45312. <https://doi.org/10.2196/45312>
- Giró-Gracia, X., & Sancho-Gil, J. M. (2022). La Inteligencia Artificial en la educación: Big data, cajas negras y solucionismo tecnológico. *Revista Latinoamericana de Tecnología Educativa - RELATEC*, 21(1), 129–145. <https://doi.org/10.17398/1695-288X.21.1.129>
- Goldman, S., Carreon, A., & Smith, S. (2024). Exploring the integration of artificial intelligence into special education teacher preparation through the tpack framework. *Journal of Special Education Preparation*, 4(2), 52–64. <https://doi.org/10.33043/6zx26bb2>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. The Center for Curriculum Redesign.
- Karataş, F., & Ataç, B. A. (2025). When TPACK meets artificial intelligence: Analyzing TPACK and AI-TPACK components through structural equation modelling. *Education and Information Technologies*, 30(7), 8979–9004. <https://doi.org/10.1007/s10639-024-13164-2>
- Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course? *Physical Review Physics Education Research*, 19(1), 010132. <https://doi.org/10.1103/PhysRevPhysEducRes.19.010132>
- Korzyński, P., Mazurek, G., Krzypkowska, P., & Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review*, 11(3), 25–37. <https://doi.org/10.15678/EBER.2023.110302>
- Lawson, T. R., Faul, A. C., & Verbist, A. N. (2019). *Research and statistics for social workers* (1º ed). Routledge. <https://doi.org/10.4324/9781315640495>
- Leahy, S. & Mishra, P. (2023). TPACK and the Cambrian explosion of AI. In E. Langran, P. Christensen & J. Sanson (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 2465-2469). New Orleans, United States: Association for the Advancement of Computing in Education (AACE). <https://www.learntechlib.org/primary/p/222145/>
- Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D. Y., Yang, X., Vodrahalli, K., He, S., Smith, D. S., Yin, Y., McFarland, D. A., & Zou, J. (2024). Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI*, 1(8). <https://doi.org/10.1056/AIoa2400196>
- Liu, B. L., Morales, D., Roser Chinchilla, J. F., Sabzalieva, E., Valentini, A., Vieira do Nascimento, D. M., & Yerovi Verano, C. A. (2023). *Oportunidades y desafíos de la era de la inteligencia artificial para la educación superior: Una introducción para los actores de la educación superior*. UNESCO IESALC. https://unesdoc.unesco.org/ark:/48223/pf0000386670_sp

- López Simó, V., Couso, D., Simarro Rodríguez, C., Garrido Espeja, A., Grimalt-Álvaro, C., Hernández Rodríguez, M. I., & Pintó Casulleras, R. (2017). El papel de las TIC en la enseñanza de las ciencias en secundaria desde la perspectiva de la práctica científica. *Enseñanza de las Ciencias: revista de investigación y experiencias didácticas*, Extra, 691-698.
<https://raco.cat/index.php/Ensenanza/article/view/334748>
- López-Simó, V., & Rezende, M. F. (2024). Challenging chatgpt with different types of physics education questions. *The Physics Teacher*, 62(4), 290–294. <https://doi.org/10.1119/5.0160160>
- MacIsaac, D. (2023). Chatbots attempt physics homework—chatgpt: Chat generative pre-trained transformer. *The Physics Teacher*, 61(4), 318–318. <https://doi.org/10.1119/10.0017700>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record: The Voice of Scholarship in Education*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Mishra, P., Warr, M., & Islam, R. (2023). Tpack in the age of chatgpt and generative ai. *Journal of Digital Learning in Teacher Education*, 39(4), 235–251. <https://doi.org/10.1080/21532974.2023.2247480>
- Mortimer, E., & Scott, P. (2010). Meaning making in secondary science classrooms (1. publ., reprinted). Open Univ. Press.
- Muñoz, J. M., Lorenzo, N., & Suñé, X. (Eds.). (2024). Intel·ligència artificial a la microeducació: Transformant l'aula del futur. ODITE–Espiral. <https://ciberespiral.org/inteligencia-artificial-a-la-la-microeducaciotransformant-laula-del-futur/>
- Rojas-Drummond, S., Torreblanca, O., Pedraza, H., Vélez, M., & Guzmán, K. (2013). ‘Dialogic scaffolding’: Enhancing learning and understanding in collaborative contexts. *Learning, Culture and Social Interaction*, 2(1), 11–21. <https://doi.org/10.1016/j.lcsi.2012.12.003>
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers’ informal formative assessment practices and students’ understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57–84. <https://doi.org/10.1002/tea.20163>
- Serrano Sánchez J. L. & Sánchez Vera, M. M. (2024) ¿A qué promesas y desafíos me enfrento como docente con la IA? In: A. A. Sagasta (Coord.). Inteligencia artificial y educación: construyendo puentes. (pp. 29-40). Ed. Graó.
- Severo Prodanov, T., & Serrano De Andrade Neto, A. (2023). Conhecimento tecnológico pedagógico do conteúdo (Tpack) de licenciandos de química: Resultados de um programa de formação de professores. *Investigações em Ensino de Ciências*, 28(2), 122–148. <https://doi.org/10.22600/1518-8795.ienci2023v28n2p122>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Tena, È., & Couso, D. (2023). ¿Cómo sé que mi secuencia didáctica es de calidad? Propuesta de un marco de evaluación desde la perspectiva de Investigación Basada en Diseño. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 20(2).
https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2023.v20.i2.2801
- Tondeur, J., Van Braak, J., Ertmer, P. A., & Ottenbreit-Leftwich, A. (2017). Understanding the relationship between teachers’ pedagogical beliefs and technology use in education: A systematic review of qualitative evidence. *Educational Technology Research and Development*, 65(3), 555–575. <https://doi.org/10.1007/s11423-016-9481-2>
- Unesco (2023). Currículos de IA para la enseñanza preescolar, primaria y secundaria: un mapeo de los currículos de IA aprobados por los gobiernos.
https://unesdoc.unesco.org/ark:/48223/pf0000380602_spa