

## No es magia, es *prompting*: el diseño de prompts como competencia emergente en la formación docente. Un estudio desde el modelo CRETA+R

It's Not Magic, It's Prompting: Prompt Design as an Emerging Competence in Teacher Education. A Study Based on the CRETA+R Model No es magia, es prompting: el diseño de prompts como competencia emergente en la formación docente. Un estudio desde el modelo CRETA+R

  Dra. Eva García-Beltrán

Universidad Tecnológica del Atlántico-Mediterráneo – UTAMED. España.

Recibido: 2025-04-22; Revisado: 2025-12-01; Aceptado: 2025-12-02; Publicado: 2026-01-01

### RESUMEN

La irrupción de la inteligencia artificial generativa en la educación plantea desafíos y oportunidades sin precedentes para la formación inicial docente. En este contexto, el diseño de *prompts* emerge como una competencia clave que articula saberes pedagógicos, lingüísticos, digitales y éticos. Este estudio analiza el desempeño de 481 estudiantes del Máster de Profesorado de Secundaria en una actividad centrada en la elaboración de *prompts* educativos, guiados por el modelo didáctico CRETA+R (Contexto, Rol, Ejemplos, Tarea, Ajustar, Refinar). Se aplicó una metodología mixta que combinó análisis cuantitativo (estadísticas descriptivas, correlaciones de Spearman y visualización de datos) con análisis cualitativo de ejemplos representativos. La evaluación se realizó mediante una rúbrica analítica aplicada por el profesorado, y los datos fueron procesados con el software JASP 0.19.3. Los resultados indican un buen dominio en componentes estructurales como "Contexto" y "Tarea", y mayores dificultades en los aspectos metacognitivos, como "Ajustar" y "Refinar". Aunque no se hallaron diferencias significativas entre especialidades, el análisis visual y cualitativo muestra patrones diferenciados por área. El modelo CRETA+R se consolida como un andamiaje eficaz para guiar el desarrollo progresivo de esta competencia emergente en contextos de formación docente.

### ABSTRACT

The emergence of generative artificial intelligence in education poses unprecedented challenges and opportunities for initial teacher education. In this context, prompt design is becoming a key competence that integrates pedagogical, linguistic, digital, and ethical knowledge. This study analyzes the performance of 481 students from the Master's Degree in Secondary Education Teaching in a task focused on creating educational prompts, guided by the instructional model CRETA+R (Context, Role, Examples, Task, Adjust, Refine). A mixed-methods approach was applied, combining quantitative analysis (descriptive statistics, Spearman correlations, and data visualizations) with a qualitative review of representative examples. The prompts were evaluated using an analytical rubric applied by instructors, and the data were processed with JASP software version 0.19.3. The results indicate stronger performance in structural components such as "Context" and "Task," while more metacognitive aspects like "Adjust" and "Refine" proved more challenging. Although no statistically significant differences were found across specializations, visual and qualitative analyses revealed discipline-specific patterns. The CRETA+R model is validated as an effective scaffold to support the progressive development of this emerging competence in teacher education.

### PALABRAS CLAVES · KEYWORDS

Educación de Profesores; Inteligencia Artificial; Enseñanza Asistida por Ordenador; Pensamiento Crítico; Formación Profesional.

Teacher Education; Artificial Intelligence; Computer Assisted Instruction; Critical Thinking; Vocational Training.

## 1. Introducción

En los últimos años, la inteligencia artificial generativa (IAg) ha transformado radicalmente las posibilidades tecnológicas en múltiples sectores, y la educación no ha sido la excepción. A diferencia de formas anteriores de IA centradas en análisis predictivo o automatización, los modelos generativos —como los grandes modelos de lenguaje (LLM) o los sistemas de generación visual y multimedia— introducen una capacidad de diálogo, creación de contenido y adaptación contextual que redefine las formas tradicionales de enseñar, aprender y evaluar.

Como señalan Bearman et al. (2023), la educación superior está atrapada entre dos discursos emergentes respecto a la IA: el de la transformación imperativa —que asume que la IA es inevitable y debe ser integrada urgentemente— y el de la alteración de la autoridad —que cuestiona cómo las relaciones de poder en la enseñanza se modifican al incorporar estas tecnologías. Desde esta perspectiva, la IAg no es simplemente una herramienta más, sino una tecnología que altera profundamente las dinámicas cognitivas, pedagógicas y sociales en el aula.

El desarrollo de la IAg ha traído consigo el surgimiento de una nueva competencia educativa: la capacidad de diseñar prompts efectivos. Un prompt es mucho más que una instrucción textual; es una forma de estructurar el conocimiento, anticipar respuestas, contextualizar intenciones y modular el comportamiento de la IA. Investigaciones recientes han subrayado que el diseño de prompts implica una combinación de habilidades lingüísticas, cognitivas, tecnológicas y pedagógicas (Bozkurt & Sharma, 2023; Korzynski et al., 2023). El proceso de escribir un prompt exige al docente tomar decisiones sobre el tono, el rol que asumirá la IA, los ejemplos que ofrecerá, el tipo de respuesta que espera y cómo refinará la interacción en función del resultado obtenido. De ahí que autores como Lo (2023) y Zamfirescu-Pereira et al. (2023) propongan considerar la escritura de prompts como una forma avanzada de alfabetización digital que debe formar parte del repertorio profesional del profesorado. Esta competencia se convierte en especialmente relevante en el contexto educativo actual, donde la IA no solo ofrece apoyo técnico, sino que se convierte en un agente activo del proceso de enseñanza-aprendizaje. Dominar la escritura de prompts permite a los docentes no solo controlar mejor el funcionamiento de las herramientas generativas, sino también diseñar experiencias de aprendizaje personalizadas, creativas y centradas en el estudiante.

La incorporación efectiva de la IA generativa en entornos educativos exige una transformación profunda en la formación inicial del profesorado. La alfabetización digital docente ya no puede centrarse únicamente en competencias instrumentales; debe incluir la comprensión crítica de los algoritmos, la ética del uso de datos, la interacción humano-máquina y, especialmente, el diseño de interacciones a través del lenguaje. En este sentido, autores como Knoth et al. (2024) han propuesto el concepto de “AI literacy” para referirse a una forma extendida de alfabetización digital que incluye la capacidad de interactuar, evaluar y tomar decisiones pedagógicas con y sobre tecnologías basadas en inteligencia artificial. La alfabetización digital crítica implica, por tanto, desarrollar en los futuros docentes una mirada reflexiva sobre los algoritmos, los sesgos que pueden contener, las estructuras de poder que reproducen y los datos que procesan. Como señalan Bearman et al. (2023), es necesario formar docentes que no solo sean usuarios informados de la tecnología, sino también mediadores éticos capaces de tomar decisiones responsables en contextos educativos mediados por IA. En este contexto, el diseño de prompts se presenta como una práctica emergente que permite materializar esta alfabetización en situaciones

reales de diseño didáctico, ya que obliga a los docentes en formación a comprender cómo piensa, responde y aprende un sistema basado en modelos de lenguaje.

Zamfirescu-Pereira et al. (2023) advierten que incluso usuarios con formación avanzada pueden fallar en la formulación de prompts eficaces, lo que pone de relieve la necesidad de una enseñanza explícita y sistematizada sobre esta práctica. Lejos de ser una habilidad técnica menor, el diseño de prompts implica tomar decisiones sobre tono, rol, formato, ejemplos y claridad de propósito. La literatura reciente subraya además que el diseño de prompts puede servir como puerta de entrada para la reflexión crítica sobre la IA en el aula. Por ejemplo, Bearman et al. (2023) insisten en que la investigación educativa sobre IA no puede reducirse a su dimensión técnica, sino que debe abordar también sus implicaciones socioculturales, epistemológicas y éticas.

Ante este panorama, resulta necesario desarrollar modelos pedagógicos que estructuren y guíen el aprendizaje del diseño de prompts en contextos educativos. El modelo CRETA+R (Contexto, Rol, Ejemplos, Tarea, Ajustar, Refinar) surge como una propuesta para apoyar a los futuros docentes en la construcción progresiva y reflexiva de prompts de calidad, facilitando una interacción significativa con herramientas de IA generativa. Inspirado en principios de *scaffolding instruccional* (Reiser, 2004; Rosenshine, 2012), CRETA+R permite descomponer la tarea compleja de escribir prompts en pasos concretos y accesibles. Cada componente actúa como un recordatorio pedagógico: establecer el contexto de la situación de aprendizaje, definir el rol que debe asumir la IA, ofrecer ejemplos relevantes, describir la tarea esperada, ajustar el lenguaje según el destinatario y refinar el prompt a través de iteraciones. En esta misma línea, Federiakin et al. (2024) sostienen que el diseño de prompts debe abordarse como una competencia evaluable que combina estrategias lingüísticas, heurísticas y retóricas, lo que demanda marcos analíticos claros para su desarrollo formativo. Complementariamente, Debnath et al. (2025) plantean un marco sistemático para investigar y enseñar prompt engineering en educación, argumentando que los modelos didácticos deben guiar tanto la estructuración del prompt como su proceso iterativo de mejora. Estas perspectivas refuerzan la pertinencia de propuestas como CRETA+R, que buscan operacionalizar esta competencia emergente mediante pasos explícitos y pedagógicamente fundamentados.

Esta estructura favorece no solo la mejora técnica del prompt, sino también procesos de metacognición, reflexión ética y evaluación formativa. Estudios recientes (Bozkurt y Sharma, 2023; Oppenlaender et al., 2024) coinciden en señalar que los prompts bien diseñados no solo producen mejores resultados por parte de la IA, sino que también promueven aprendizajes más profundos al obligar al usuario a explicitar sus intenciones comunicativas y a revisar críticamente las respuestas generadas. La aplicación del modelo CRETA+R en la formación inicial docente permite además adaptar el diseño de prompts a las especificidades de cada disciplina, facilitando su integración curricular de forma contextualizada. Asimismo, ofrece un marco común para la evaluación de los prompts, permitiendo establecer rúbricas transparentes y procesos de mejora continua.

En los últimos dos años se ha producido un notable incremento en las investigaciones sobre la integración de la IA en los programas de formación inicial docente. Bond (2024), en una revisión sistemática de 138 estudios, identifica el diseño de materiales con IA, el uso de asistentes conversacionales y la evaluación automatizada como las aplicaciones más comunes. Sin embargo, también destaca la falta de propuestas pedagógicas concretas para desarrollar competencias críticas en relación con la IA. Moldavan y Nafziger (2024), por su parte, han trabajado con estudiantes de magisterio en el diseño de planes de clase asistidos

por IA generativa, mostrando que el uso guiado de prompts puede ayudar a los futuros docentes a cuestionar la autoridad de la máquina, desarrollar pensamiento crítico y reflexionar sobre la equidad y la personalización del aprendizaje. La investigación de Theophilou et al. (2023) representa otro ejemplo relevante. En un estudio piloto con estudiantes europeos, se exploró cómo el trabajo con prompts puede ser utilizado en el aula no solo para mejorar habilidades técnicas, sino también para discutir los límites de la IA, sus sesgos y sus implicaciones éticas. Todos estos estudios coinciden en señalar que la enseñanza de la IA no debe limitarse a la capacitación técnica, sino que debe estar acompañada de marcos pedagógicos que promuevan la comprensión crítica, el diseño ético y la interacción significativa con estas herramientas emergentes.

## 2. Metodología

El presente estudio adopta un enfoque descriptivo y exploratorio, con el objetivo de analizar la competencia emergente en el diseño de prompts por parte de estudiantes en formación inicial docente, a través de la aplicación del modelo CRETA+R. Esta aproximación metodológica resulta especialmente pertinente en investigaciones educativas orientadas a fenómenos poco explorados o que emergen en contextos de cambio tecnológico acelerado, como es el caso del uso de inteligencia artificial generativa en la formación del profesorado.

La investigación se inscribe dentro de un marco de análisis mixto, combinando el estudio cuantitativo de patrones generales y comparativas entre grupos, con un análisis cualitativo orientado a interpretar ejemplos representativos de los productos generados por el alumnado. Esta complementariedad de enfoques permite no solo describir el desempeño de los estudiantes, sino también comprender los matices discursivos, pedagógicos y comunicativos implicados en la escritura de prompts educativos.

### 2.1. Participantes

La muestra del estudio estuvo compuesta por 481 estudiantes del Máster Universitario en Formación del Profesorado de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanza de Idiomas, procedentes de diversas especialidades: Lengua y Literatura, Matemáticas, Inglés, Biología y Geología, Geografía e Historia, Educación Física, entre otras. Todos los participantes cursaban una asignatura centrada en la innovación y el uso de tecnologías digitales aplicadas a la enseñanza, en la que se introdujo el trabajo con inteligencia artificial generativa como parte de una experiencia formativa. El Máster pertenece a una universidad online.

La muestra presentó una distribución equilibrada en cuanto a género y edad (rango: 22–48 años), y todos los participantes contaban con formación universitaria previa en su disciplina, aunque con niveles de familiaridad muy diversos respecto al uso de herramientas de IA.

### 2.2. Instrumento

El instrumento principal de recogida de datos consistió en una tarea individual que implicaba el diseño de un prompt educativo para ser utilizado con un modelo de IA

generativa (ChatGPT u otro equivalente). La consigna pedía al alumnado que desarrollara un prompt alineado con una situación de aprendizaje realista de su especialidad, aplicando de forma explícita los componentes del modelo CRETA+R, una estructura didáctica compuesta por los siguientes elementos:

- Contexto: situación educativa clara y coherente.
- Rol: definición del papel que debe adoptar la IA (por ejemplo, tutor, evaluador, estudiante).
- Ejemplos: modelos o casos que orientan la respuesta.
- Tarea: descripción clara de la actividad esperada.
- Ajustar: adaptación de formato, tono o lenguaje.
- Refinar: iteración o mejora tras una primera interacción con la IA.

Cada uno de estos componentes fue evaluado por el equipo docente mediante una rúbrica analítica con cuatro niveles de logro: Excelente, Bueno, Adecuado, Insuficiente. La rúbrica fue consensuada por los profesores responsables de la asignatura y utilizada de forma sistemática con fines tanto formativos como investigativos.

Además de la calificación cualitativa por componente, se registraron otras variables relevantes del curso académico, tales como la nota final de la asignatura, la nota obtenida en el examen presencial final y la calificación específica de la actividad con IA generativa.

Para garantizar la fiabilidad del proceso evaluador, la rúbrica analítica fue aplicada por un equipo de cuatro docentes responsables de la asignatura, previamente coordinados mediante una sesión de calibración. En dicha sesión, los profesores revisaron de manera conjunta ejemplos reales de prompts y discutieron los criterios operativos de cada nivel de logro, con el fin de unificar criterios y reducir la variabilidad interevaluador. La rúbrica incluía descriptores específicos para cada componente del modelo CRETA+R, distribuidos en cuatro niveles (Excelente, Bueno, Adecuado e Insuficiente), que evaluaban la claridad del contexto, la pertinencia del rol asignado a la IA, la calidad de los ejemplos, la precisión de la tarea, el ajuste lingüístico y el refinamiento iterativo. Este procedimiento permitió maximizar la consistencia y transparencia en la evaluación, aspecto esencial dado que dichas calificaciones constituyen la base del análisis cuantitativo y cualitativo del estudio.

### 2.3. Variables analizadas

El conjunto de datos permitió trabajar con las siguientes variables principales:

- Especialidad del máster (variable categórica): clasificada en áreas disciplinarias normalizadas.
- Calidad del prompt (variables ordinales): nivel alcanzado en cada uno de los seis componentes del modelo CRETA+R.
- Nota de la actividad de diseño de prompts (variable continua): calificación específica obtenida por el estudiante en la tarea.
- Nota del curso (variable continua): calificación global obtenida en la asignatura.
- Nota del examen presencial (variable continua): evaluación escrita individual.

Estas variables fueron objeto de análisis tanto individual como relacional, para explorar posibles patrones de desempeño por especialidad, correlaciones entre la calidad del prompt y las calificaciones, y componentes que presentan mayor o menor desarrollo por parte del alumnado.

## 2.4. Procedimiento de análisis de datos

Los datos fueron organizados y tratados mediante un análisis mixto, que integró herramientas estadísticas y revisión cualitativa:

### 2.4.1. Análisis cuantitativo

- Estadísticas descriptivas generales (medias, frecuencias, desviaciones estándar).
- Análisis comparativos por especialidad (test de Kruskal-Wallis y gráficos de caja).
- Análisis de correlaciones entre notas y desempeño en los componentes del modelo (coeficientes de Spearman).

### 2.4.2. Análisis cualitativo

Revisión en profundidad de una muestra de prompts representativos, seleccionados por su nivel de desempeño y por su potencial explicativo. Este análisis permitió identificar patrones discursivos, estrategias recurrentes, y errores comunes en la aplicación de cada componente del modelo CRETA+R.

Todo el procesamiento y análisis estadístico se realizó con el software JASP versión 0.19.3 para Apple, una herramienta de análisis de código abierto que permite realizar procedimientos estadísticos robustos y visualizaciones interactivas. JASP fue elegido por su accesibilidad y transparencia, lo que lo convierte en un recurso especialmente adecuado para contextos educativos en los que se fomenta el pensamiento crítico y reproducible.

## 3. Resultados

El análisis cuantitativo realizado permitió obtener una visión detallada del desempeño del alumnado en el diseño de prompts aplicando el modelo CRETA+R. Las estadísticas descriptivas generales muestran que la nota media de la actividad fue elevada ( $M = 8.10$ ;  $DT \approx 0.49$ ), lo que indica un desempeño globalmente satisfactorio por parte de los estudiantes. No obstante, los valores atípicos observados en algunas especialidades (como Matemáticas o Biología y Geología) reflejan la existencia de diferencias individuales relevantes. El análisis comparativo por especialidad, realizado mediante el test no paramétrico de Kruskal-Wallis, no arrojó diferencias estadísticamente significativas ( $H= 5.13$ ;  $p=0.400$ ), lo cual sugiere que la nota obtenida en la actividad de diseño de prompt no depende de forma sustancial de la disciplina de procedencia del alumnado.

En cuanto a la relación entre el desempeño por componente del modelo CRETA+R y la nota final, se aplicó un análisis de correlación de Spearman, codificando ordinalmente los niveles de evaluación. Los coeficientes de correlación fueron bajos en todos los casos, destacando únicamente el componente “Ajustar”, que presentó una correlación débil pero

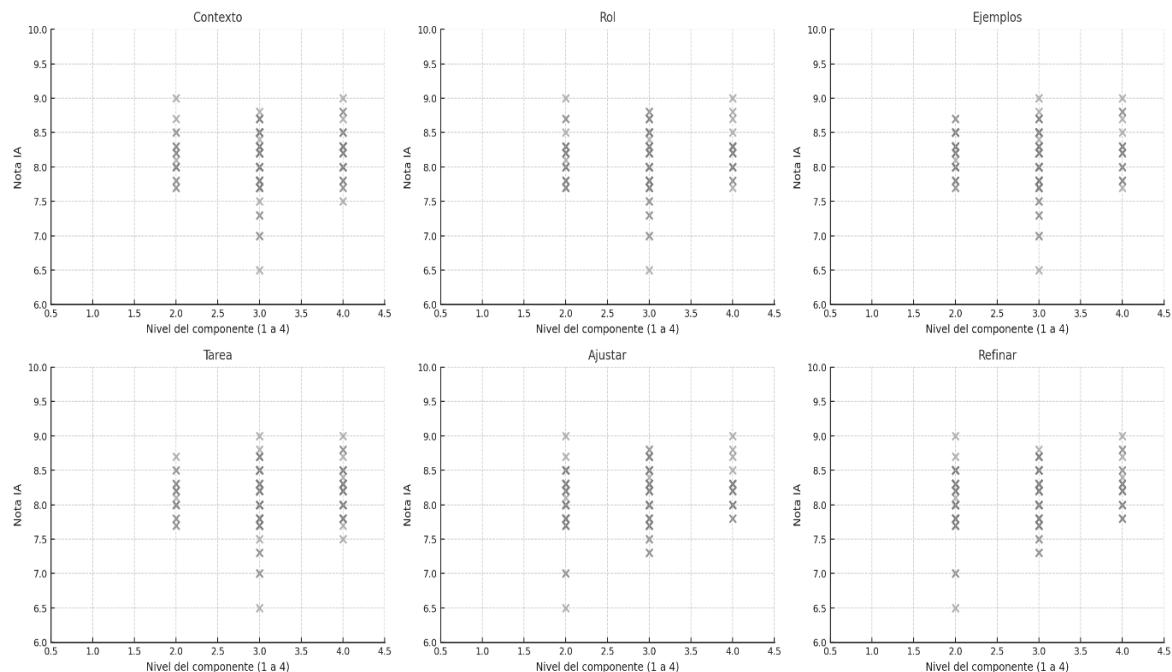
estadísticamente significativa ( $\rho = 0.111$ ;  $p = 0.031$ ), lo cual sugiere que una mayor precisión en la redacción y afinado del prompt podría estar ligeramente asociado a un mejor resultado global. El resto de componentes mostró correlaciones muy próximas a cero y no significativas, lo que refuerza la idea de que el rendimiento en la actividad no puede explicarse de forma directa por un solo factor, sino por una combinación compleja de elementos. Los gráficos de dispersión confirmarán esta interpretación (figura 1, siguiente página), mostrando distribuciones planas sin patrones claros, y apuntan a la necesidad de seguir explorando variables interviniéntes que puedan incidir en el éxito del diseño de prompts con IA.

### 3.1 Evaluación general por componente CRETA+R

La mayoría del alumnado obtuvo valoraciones en los niveles “Bueno” y “Excelente”, siendo los componentes de Contexto y Tarea los mejor resueltos. En contraste, los componentes Ajustar y Refinar presentaron una mayor concentración en los niveles “Adecuado”, lo que sugiere que los estudiantes encuentran mayores dificultades en los aspectos relacionados con el ajuste de tono, lenguaje y capacidad de iteración con la IA. A continuación, en la figura 2, se puede ver la distribución de niveles de evaluación por componente del modelo CRETA+R.

**Figura 1**

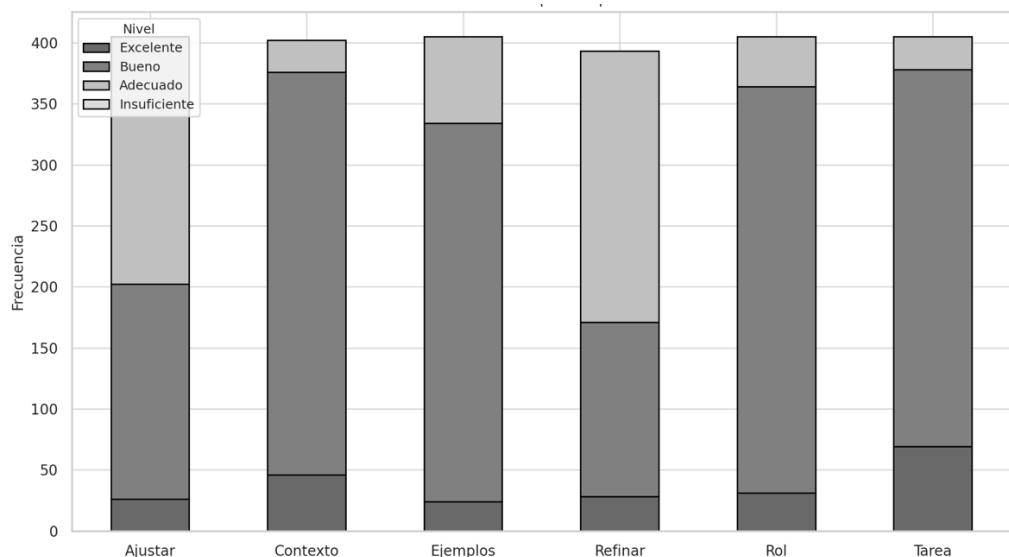
*Correlación entre desempeño en componentes CRETA+R y nota de la actividad*



Fuente: elaboración propia.

**Figura 2**

*Distribución de niveles de evaluación por componentes del modelo CRETA+R*



Fuente: elaboración propia.

Cada barra representa uno de los seis componentes del modelo CRETA+R. Los niveles de evaluación están representados en una escala de cuatro categorías: Excelente, Bueno, Adecuado e Insuficiente, cada uno con un tono de gris diferente. Los componentes con un mejor desempeño son Contexto, Rol, Tarea y Ejemplos: muestran un claro predominio del nivel "Bueno", con pocas apariciones de "Adecuado" y "Excelente". Este patrón sugiere que la mayoría del alumnado logra cumplir los criterios básicos de calidad en estos componentes, aunque sin destacar especialmente. Contexto es uno de los componentes con mayor número de valoraciones positivas (Excelente + Bueno), lo cual puede estar relacionado con la familiaridad del alumnado con la exposición de información contextual en tareas académicas.

Por el contrario, los componentes que muestran mayores dificultades son Ajustar y Refinar son los componentes que presentan un porcentaje considerablemente más alto en la categoría "Adecuado", lo que indica que estos aspectos del diseño de prompts fueron más complejos para los estudiantes. Esto puede reflejar la dificultad de revisar y mejorar iterativamente los prompts (refinar) o de redactarlos con claridad, precisión y sin errores (ajustar), lo que implica una mayor madurez lingüística, metacognitiva o técnica. Resulta pertinente señalar la ausencia del nivel "Insuficiente". El hecho de que ningún componente tenga una proporción significativa de evaluaciones como "Insuficiente" sugiere un desempeño mínimo aceptable en todos los casos, posiblemente vinculado a la calidad de la formación recibida o al uso de una rúbrica clara que orientó correctamente a los estudiantes.

En cuanto a las implicaciones didácticas de este análisis, se puede concluir que que los componentes más estructurales del diseño de un prompt (como establecer contexto, definir la tarea o indicar un rol) están más consolidados. En cambio, los aspectos más metacognitivos y de revisión (como ajustar y refinar) requieren un mayor apoyo formativo. Esto podría abordarse mediante actividades de andamiaje progresivo, ejercicios de retroalimentación entre pares o el uso de herramientas de IA para revisar iterativamente.

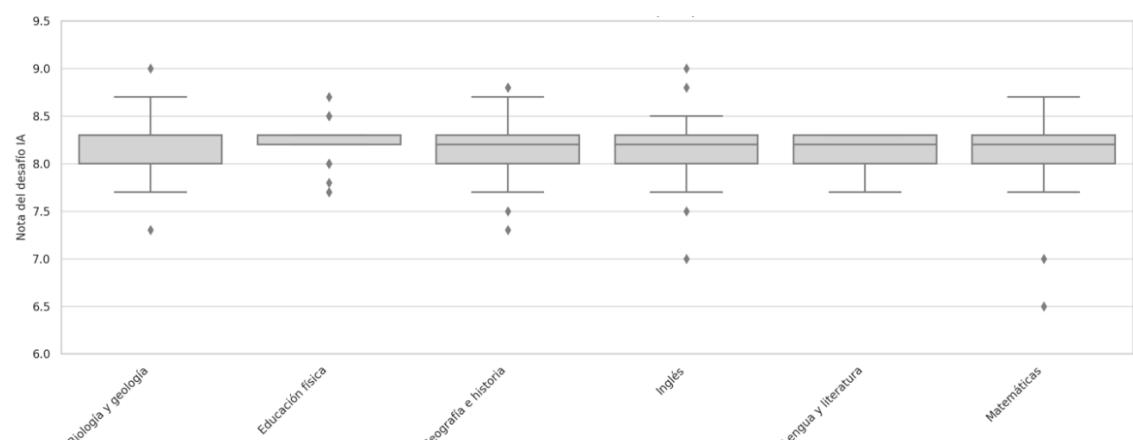
### 3.2 Análisis por especialidad

En cuanto a la calificación obtenida en la actividad de diseño de prompt, se observan diferencias en función de las especialidades de los alumnos en el Máster. La mayoría de las especialidades presentan notas medias relativamente altas, centradas en torno al 8.0 - 8.3, lo que sugiere un buen desempeño general en la actividad de diseño de prompts con IA. Los rangos intercuartílicos son bastante estrechos en varias especialidades, lo que indica poca variabilidad entre los estudiantes. Esto puede reflejar una rúbrica clara y un criterio de evaluación homogéneo. En cuanto a la comparación entre especialidades, Educación Física muestra una dispersión muy baja (casi sin caja visible), lo que sugiere que la mayoría de estudiantes obtuvieron la misma nota. Matemáticas presenta una distribución más baja con respecto al resto, con valores atípicos hacia el 6.5, lo que indica cierta dificultad del alumnado de esta especialidad para adaptarse a los requerimientos de la actividad. Esto puede relacionarse con una menor familiaridad con el lenguaje pedagógico o con la escritura reflexiva. Por su parte, Inglés, Lengua y Literatura, y Geografía e Historia muestran distribuciones similares, centradas en torno a 8.2, con una ligera asimetría negativa (algunos valores bajos aislados). Biología y Geología, y Matemáticas presentan outliers más bajos, lo que evidencia que algunos estudiantes tuvieron mayores dificultades con la tarea.

Las diferencias por especialidad podrían reflejar distintos niveles de alfabetización pedagógica o tecnológica, lo que sugiere la necesidad de adaptar la enseñanza del diseño de prompts según el perfil disciplinar del alumnado. Las especialidades que muestran menor rendimiento necesitarían estrategias de andamiaje más explícitas (por ejemplo, secuencias guiadas, ejemplos contextualizados o feedback iterativo). La ausencia de valores extremos altos indica que, aunque el desempeño fue bueno en general, hubo pocas respuestas sobresalientes, lo que podría sugerir una oportunidad para fomentar mayor creatividad o profundidad crítica en el uso de la IA. Asimismo, se observó que los estudiantes de Matemáticas, Inglés y Lengua y Literatura obtuvieron mejores promedios generales en la mayoría de los componentes. Por el contrario, las especialidades como Educación Física y Biología y Geología tendieron a concentrarse en niveles medios o adecuados.

**Figura 3**

*Distribución de la calificación de la actividad por especialidad (boxplots)*

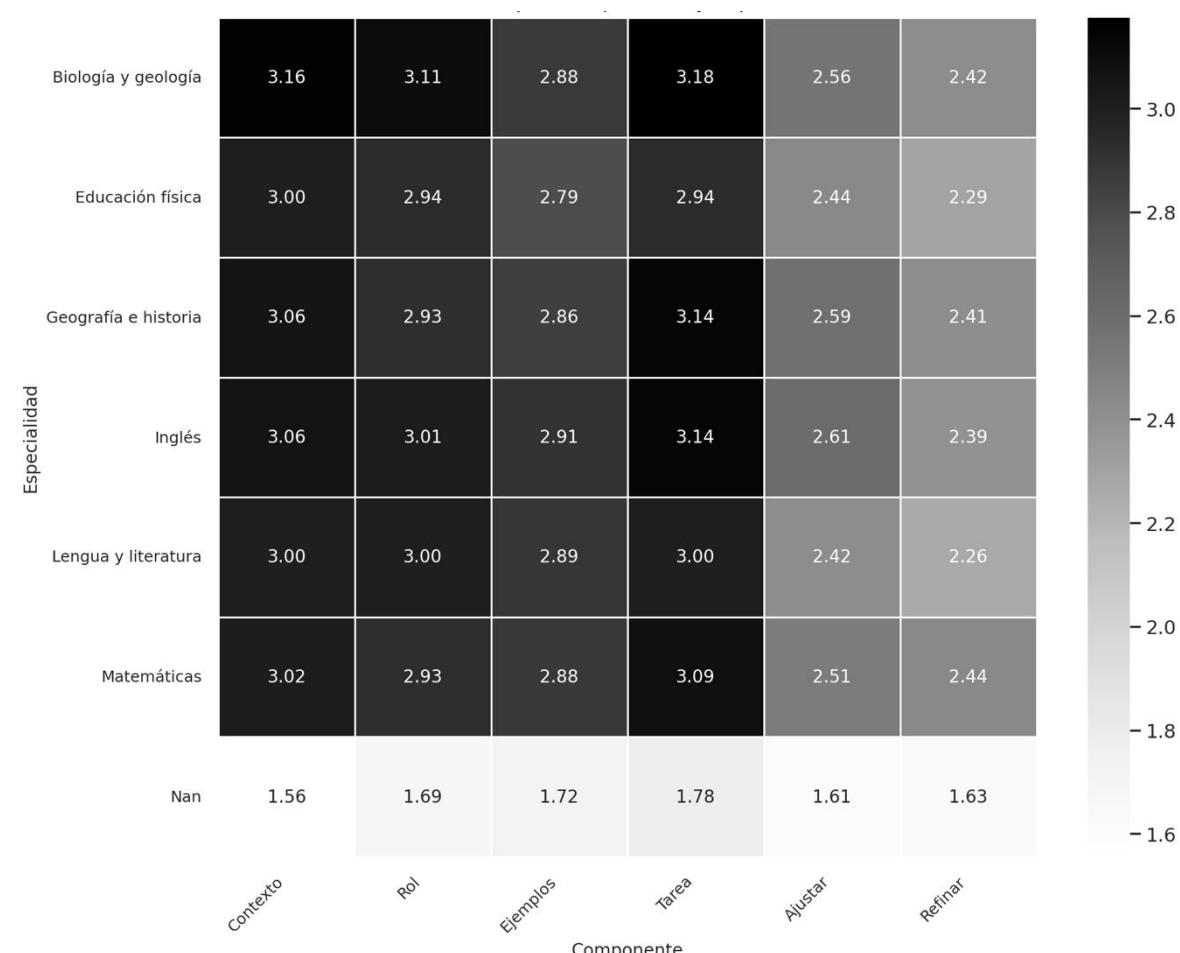


Fuente: elaboración propia.

Las especialidades con mayor cantidad de estudiantes muestran también una mayor distribución hacia los niveles altos de evaluación. Este patrón se repite, con matices, en los demás componentes del modelo CRETA+R. La visualización tipo *heatmap* (figura 4, siguiente página) muestra las puntuaciones medias por componente del modelo CRETA+R según la especialidad del máster, utilizando una escala de 1 (Insuficiente) a 4 (Excelente). En general, se observa un predominio de valoraciones cercanas al nivel “Bueno” (3) en la mayoría de los componentes y especialidades, lo que sugiere un desempeño sólido pero aún con margen de mejora. Las especialidades de Lengua y literatura, Geografía e historia y Orientación educativa presentan puntuaciones ligeramente superiores a la media en casi todos los componentes, destacando especialmente en “Contexto” y “Rol”. Por el contrario, especialidades como Educación física, Matemáticas y Filosofía muestran valores algo más bajos, en especial en los componentes más complejos como “Refinar” y “Ajustar”, lo cual podría vincularse a un menor entrenamiento en tareas discursivas o reflexivas propias del diseño de prompts educativos. Este resultado sugiere que, aunque el modelo CRETA+R es aplicable de forma transversal, algunas especialidades requieren de andamiajes pedagógicos específicos para mejorar en los componentes más vinculados con la revisión crítica y la mejora iterativa del prompt.

#### Figura 4

*Heatmap de puntuaciones medias por componente y especialidad*

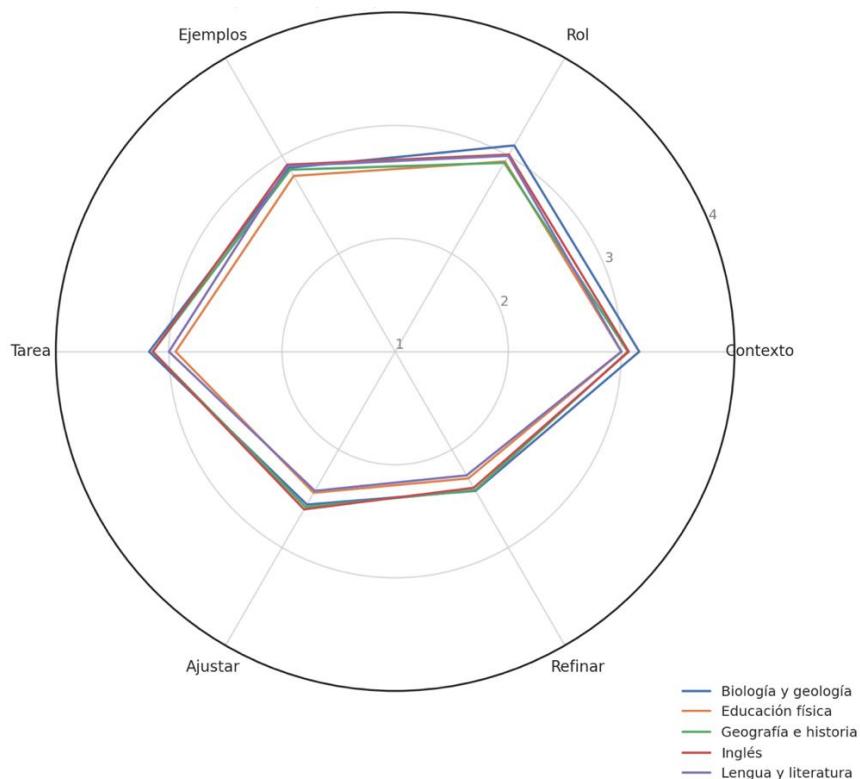


Fuente: elaboración propia.

El gráfico radar (figura 5, siguiente página) compara el perfil medio por especialidad en los componentes del modelo CRETA+R. Se observa un patrón general equilibrado, con puntuaciones cercanas al nivel “Bueno” (3), aunque con diferencias notables entre áreas. Las especialidades de Lengua y Literatura y Geografía e Historia presentan perfiles más amplios y consistentes, especialmente en los componentes “Contexto”, “Rol” y “Tarea”. En cambio, Matemáticas y Biología y Geología muestran rendimientos más bajos, sobre todo en “Refinar” y “Ajustar”, lo que sugiere dificultades en los procesos de revisión y mejora del prompt. Esta visualización evidencia la utilidad del modelo CRETA+R para detectar necesidades formativas específicas por área disciplinaria.

**Figura 5**

*Perfil comparativo por especialidad (gráfico radar)*



Fuente: elaboración propia.

### 3.3 Resultados del análisis cualitativo

El análisis cualitativo de una muestra representativa de prompts permitió identificar patrones discursivos que no emergen con claridad en el análisis cuantitativo. En los componentes estructurales (Contexto, Rol y Tarea), los estudiantes tendieron a ofrecer descripciones claras y coherentes, aunque en algunos casos el contexto se formuló de forma excesivamente general (“desarrolla un tema de mi asignatura”) sin especificar nivel educativo ni objetivo pedagógico. En cuanto al uso de ejemplos, se observaron diferencias por especialidad: los estudiantes de Lengua, Inglés y Geografía emplearon modelos

concretos y alineados con la tarea, mientras que en otras áreas, como Educación Física o Tecnología, los ejemplos eran escasos o poco informativos, lo que limitaba la capacidad de la IA para generar respuestas precisas.

Los componentes más problemáticos fueron Ajustar y Refinar. En Ajustar, varios estudiantes no lograron adaptar el tono, nivel de lenguaje o formato al destinatario previsto, produciendo instrucciones demasiado técnicas o, por el contrario, excesivamente coloquiales. En Refinar, la mayoría de los prompts no incluían ninguna indicación de iteración posterior, lo cual confirma una comprensión limitada del carácter cíclico del trabajo con IA generativa. Solo un pequeño número de estudiantes incorporó estrategias de revisión (“si la respuesta no cumple... vuelve a formularla de este modo”), demostrando mayor madurez metacognitiva. En conjunto, estos hallazgos cualitativos permiten profundizar en las dificultades observadas en el análisis cuantitativo y evidencian la necesidad de un mayor acompañamiento didáctico en las fases de ajuste y mejora del prompt.

#### 4. Discusión

La implementación del modelo CRETA+R ha permitido visibilizar patrones de desempeño y áreas de dificultad que coinciden con las tensiones actuales en torno a la alfabetización en inteligencia artificial en la educación superior. Diversos autores coinciden en señalar que el diseño de prompts representa una nueva forma de alfabetización digital, comparable a otras habilidades avanzadas de pensamiento crítico y comunicación (Lo, 2023). En este sentido, los estudiantes del máster que participaron en este estudio demostraron niveles aceptables en componentes estructurales como Contexto y Tarea, pero dificultades persistentes en elementos que exigen mayor conciencia comunicativa, como Ajustar o Refinar. Por tanto, enseñar a diseñar prompts va más allá de la técnica: implica pensar con la máquina, anticipar interpretaciones, modular instrucciones y aprender a iterar.

La variabilidad observada por especialidades sugiere que el perfil disciplinar influye notablemente en la forma en que se abordan los componentes del modelo. Mientras que los estudiantes de Lengua y Literatura, Inglés y Matemáticas muestran perfiles equilibrados y sólidos, otros como Educación Física y Biología y Geología presentan debilidades más marcadas, especialmente en el refinamiento y ajuste del lenguaje. Este patrón coincide con lo observado por Silva (2024) en el campo de la enseñanza de la química, donde los estudiantes presentan inicialmente una comprensión superficial del diseño de prompts y tienden a adoptar estrategias de copiado y pegado antes de desarrollar enfoques más sofisticados. Estas variaciones podrían explicarse en parte por la familiaridad previa con formas de expresión académica estructurada o por la cultura didáctica predominante en cada especialidad. Tal como argumentan Bozkurt y Sharma (2023), el arte de *susurrar al algoritmo* requiere habilidades que van desde la formulación clara hasta la creatividad y la empatía digital, habilidades que no siempre se desarrollan de manera homogénea entre disciplinas.

Desde un enfoque más cualitativo, el análisis de ejemplos representativos permitió observar que la componente “Refinar” fue la menos trabajada por la mayoría de los estudiantes. Esta debilidad se alinea con los hallazgos de Eager y Brunton (2023), quienes subrayan la importancia de enseñar estrategias iterativas en la relación con la IA generativa,

superando el uso superficial o unidireccional. La falta de revisión o ajuste del prompt tras obtener una primera respuesta de la IA sugiere la necesidad de fortalecer la dimensión metacognitiva de esta competencia, integrando mecanismos de autoevaluación y mejora progresiva. Asimismo, se observaron dificultades en la aplicación del componente “Ejemplos”, especialmente en disciplinas como educación física o tecnología, donde los estudiantes no siempre lograron proporcionar modelos comprensibles para la IA ni pertinentes al contenido. Como señalan Ranade et al. (2024), los prompts efectivos deben articular de forma clara el contexto, la audiencia y el tipo de respuesta esperada, lo que implica una alfabetización retórica que aún no está presente en todos los futuros docentes. Esta brecha indica que el desarrollo del diseño de prompts no puede abordarse únicamente desde una perspectiva funcional, sino que debe considerar elementos del diseño comunicativo, la teoría del discurso y la interacción semiótica con las tecnologías. Además, el hecho de que los estudiantes mejor evaluados mostraran mayores niveles de reflexión en las fases de ajuste y refinamiento conecta con lo que Sajja et al. (2024) denominan “personalización inteligente del aprendizaje”, una capacidad crítica en entornos asistidos por IA.

Los resultados ponen de manifiesto la necesidad de incluir explícitamente el diseño de prompts en los programas de formación docente como competencia pedagógica emergente, alineada con las directrices europeas sobre IA en educación (Comisión Europea, 2022) y con el nuevo Reglamento 2024/1689 (Unión Europea, 2024), que subraya la responsabilidad de educadores en el uso ético, transparente y seguro de estas tecnologías. Desde un enfoque crítico, Bearman et al. (2023) advierten que los discursos actuales sobre IA en educación suelen oscilar entre el entusiasmo tecnodeterminista y el rechazo alarmista. Frente a ello, este estudio aporta una evidencia concreta de cómo los futuros docentes pueden comenzar a relacionarse con la IA no solo como usuarios, sino como diseñadores reflexivos de experiencias de aprendizaje mediadas por tecnología. Tal como señalan Baidoo-Anu y Owusu Ansah (2023), el uso generalizado de herramientas como ChatGPT en la educación superior requiere de acompañamiento ético, formación crítica y políticas institucionales claras. El desarrollo de la competencia de diseño de prompts debe ir necesariamente acompañado de una reflexión sobre los límites y responsabilidades del uso de IA en el aula.

En este sentido, el modelo CRETA+R se revela como una propuesta valiosa no solo para estructurar la tarea de escribir prompts, sino también como mediador didáctico para enseñar a pensar con y sobre la IA. Su diseño se alinea con estrategias recomendadas en la literatura, como el desglose por tareas (Karakaya, 2025) y el refinamiento iterativo (Higginbotham & Matthews, 2024). El uso del modelo CRETA+R ha funcionado como una estrategia de scaffolding para estructurar el pensamiento del alumnado en torno a la IA generativa. Esta herramienta no solo facilita la evaluación formativa del trabajo con prompts, sino que, como sugieren Korzyński et al. (2023), puede convertirse en una base estructural para desarrollar competencias de ingeniería de prompts como parte del repertorio profesional docente. El hecho de que los componentes mejor valorados hayan sido “Tarea” y “Contexto” indica que el modelo ofrece un andamiaje eficaz para los aspectos más próximos a la planificación docente, mientras que aquellos más novedosos —como la iteración o el ajuste tonal— requieren mayor tiempo y práctica para consolidarse.

Finalmente, los hallazgos también ponen de relieve el valor del aprendizaje situado. Tal como se evidenció en el taller analizado por Graux et al. (2024), el dominio del prompt engineering no se adquiere únicamente por exposición a ejemplos, sino por ensayo, error,

retroalimentación y reconstrucción. Incorporar esta competencia en entornos colaborativos, donde los estudiantes puedan compartir, comentar y reelaborar prompts, puede aumentar tanto su eficacia técnica como su dimensión crítica y reflexiva.

## 5. Conclusiones

Este estudio ha permitido explorar, desde una perspectiva empírica y pedagógica, el desarrollo de la competencia para diseñar prompts educativos por parte de estudiantes del Máster de Profesorado de Secundaria. Los resultados obtenidos confirman que dicha competencia no solo es relevante en el contexto actual de transformación digital, sino que requiere de estrategias didácticas específicas para su fortalecimiento. Los datos muestran que los futuros docentes son capaces de generar instrucciones claras y coherentes (especialmente en los componentes Contexto y Tarea), pero enfrentan mayores dificultades en fases más sofisticadas del proceso, como el ajuste del lenguaje y el refinamiento iterativo. Estas limitaciones coinciden con las barreras identificadas por otros estudios sobre alfabetización en IA (Zamfirescu-Pereira et al., 2023; Knoth et al., 2024), y refuerzan la necesidad de incorporar enfoques sistemáticos como el modelo CRETA+R en la formación inicial. Además, la comparación entre especialidades revela que el enfoque disciplinar influye significativamente en el perfil de desempeño. Áreas como Lengua, Inglés o Matemáticas presentan mayor solidez general, mientras que otras, como Educación Física, evidencian la necesidad de mayor acompañamiento didáctico.

En este contexto, se proponen una serie de recomendaciones pedagógicas para favorecer la integración eficaz del diseño de prompts como competencia emergente en la formación docente:

**Tabla 1**

*Recomendaciones pedagógicas para el desarrollo de la competencia en diseño de prompts en la formación docente*

Área	Recomendación	Justificación
Integración curricular	Incluir el diseño de prompts como contenido explícito en asignaturas de didáctica, innovación educativa o competencia digital docente.	Responde a la necesidad de alfabetización en IA en formación inicial (Comisión Europea, 2022; Knoth et al., 2024).
Andamiaje metodológico	Utilizar modelos como CRETA+R para guiar y estructurar la redacción de prompts, con ejemplos progresivos y análisis colaborativo.	Mejora la calidad del diseño y promueve metacognición (Korzyński et al., 2023).
Iteración y refinamiento	Fomentar tareas con múltiples rondas de refinamiento del prompt tras interacción con la IA, incorporando reflexión crítica sobre los resultados.	Potencia habilidades adaptativas y metacognitivas (Bozkurt & Sharma, 2023; Lo, 2023).

Área	Recomendación	Justificación
Evaluación formativa	Diseñar rúbricas basadas en CRETA+R que incluyan criterios de claridad, adaptabilidad, ajuste del lenguaje y proceso de mejora del prompt.	Permite retroalimentación efectiva y seguimiento del progreso (González-Calatayud et al., 2021).
Perspectiva disciplinar	Adaptar los ejemplos y tareas de diseño de prompts a las necesidades de cada especialidad, promoviendo enfoques contextualizados.	Se ajusta a las diferencias observadas por área de conocimiento (Luckin et al., 2024; nuestros resultados).
Formación ética y crítica	Incluir espacios de debate sobre los riesgos, sesgos y límites de la IA generativa, especialmente en tareas de evaluación automatizada.	Alineado con el Reglamento (UE) 2024/1689 y propuestas de IA inclusiva (Roscoe, 2023; Bearman et al., 2023).

La integración de estas prácticas puede contribuir a formar docentes capaces de interactuar de forma crítica, creativa y ética con herramientas basadas en inteligencia artificial, avanzando así hacia una educación más inclusiva, reflexiva y tecnológicamente contextualizada.

## Referencias

- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence: Understanding the impact of ChatGPT on teaching and learning. *Education and Information Technologies*, 29, 739–758. <https://doi.org/10.1007/s10639-023-11948-4>
- Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, 86, 369–385. <https://doi.org/10.1007/s10734-022-00937-2>
- Bond, M. (2024). AI applications in Initial Teacher Education: A systematic mapping review. *Computers and Education: Artificial Intelligence*, 6, 100228. <https://doi.org/10.1016/j.caai.2024.100228>
- Bozkurt, A., & Sharma, R. C. (2023). Generative AI and prompt engineering: The art of whispering to let the genie out of the algorithmic world. *Asian Journal of Distance Education*, 18(2), i–vii. <https://tinyurl.com/mr2csd5u>
- Comisión Europea (2022). *Directrices éticas sobre el uso de la inteligencia artificial (IA) y los datos en la educación y formación para los educadores*. Oficina de Publicaciones de la Unión Europea. <https://tinyurl.com/yckdz2vk>

- Debnath, S., Dai, T., Smith, G., & Sridhar, S. (2025). Prompt engineering in education: A framework and research agenda. *Computers and Education: Artificial Intelligence*, 7, 100289. <https://doi.org/10.1016/j.caeai.2025.100289>
- Eager, B., & Brunton, R. (2023). Prompting Higher Education Towards AI-Augmented Teaching and Learning Practice. *Journal of University Teaching & Learning Practice*, 20(5). <https://doi.org/10.53761/1.20.5.02>
- Federiakin, M., Azaria, A., & Hershkovitz, A. (2024). Prompt engineering skills and strategies: Toward a framework for assessing student interaction with generative AI. *Computers and Education: Artificial Intelligence*, 6, 100267. <https://doi.org/10.1016/j.caeai.2024.100267>
- González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial Intelligence for Student Assessment: A Systematic Review. *Applied Sciences*, 11(12), 5467. <https://doi.org/10.3390/APP11125467>
- Graux, A., Brassier, C., & Guillemet, M. (2024). Prompt engineering as a learning activity in higher education: A case study of a design workshop. *Computers and Education: Artificial Intelligence*, 6, 100258. <https://doi.org/10.1016/j.caeai.2024.100258>
- Higginbotham, G.Z., & Matthews, N.S. (2024). Prompting and In-Context Learning: Optimizing Prompts for Mistral Large. *Research Square*. <https://doi.org/10.21203/rs.3.rs-4430993/v1>
- Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, 6, 100225. <https://doi.org/10.1016/j.caeai.2024.100225>
- Korzyński, P., Mazurek, G., Krzypkowska, P., & Kurasiński, 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>
- Lo, L. (2023). The Art and Science of Prompt Engineering: A New Literacy in the Information Age. *Internet Reference Services Quarterly*, 27, 203–210. <https://doi.org/10.1080/10875301.2023.2227621>
- Luckin, R., Rudolph, J., Grünert, M., & Tan, S. (2024). Exploring the future of learning and the relationship between human intelligence and AI. An interview with Professor Rose Luckin. *Journal of Applied Learning and Teaching*, 7(1), 346–363. <https://doi.org/10.37074/jalt.2024.7.1.27>
- Moldavan, C., & Nafziger, R. (2024). Scaffolding a Critical Lens of Generative AI for Lesson Planning. *Contemporary Issues in Technology and Teacher Education*, 24(1), 45–62. <https://tinyurl.com/mvaeu929>
- Oppenlaender, J., Linder, R., & Silvennoinen, J. (2024). Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering. *International Journal of Human–Computer Interaction*, 1-23. <https://doi.org/10.1080/10447318.2024.2431761>

- Ranade, D., Gillespie, T., & Lin, T. (2024). Rhetorical prompting: A framework for prompt design in educational contexts. *Learning, Media and Technology*, 49(1), 12–34. <https://doi.org/10.1080/17439884.2024.2278910>
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 13(3), 273–304. [https://doi.org/10.1207/s15327809jls1303\\_2](https://doi.org/10.1207/s15327809jls1303_2)
- Roscoe, R. D. (2023). Building inclusive and equitable artificial intelligence for education. *XRDS: Crossroads, The ACM Magazine for Students*, 29(3), 22–25. <https://doi.org/10.1145/3589637>
- Rosenshine, B. (2012). Principles of Instruction: Research-Based Strategies That All Teachers Should Know. *American Educator*, 36(1), 12–19. <https://eric.ed.gov/?id=EJ971753>
- Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. *Information*, 15(10), 596. <https://doi.org/10.3390/info15100596>
- Silva, B. (2024). Generative Artificial Intelligence in chemistry teaching: ChatGPT, Gemini, and Copilot's content responses. *Journal of Applied Learning and Teaching*, 7(2). <https://doi.org/10.37074/jalt.2024.7.2.13>
- Theophilou, E., Koyutürka, C., Yavari, M., Bursic, S., Donabauer, G., Telari, A., Testa, A., Boiano, R., Hernández-Leo, D., Ruskov, M., Taibi, D., Gabbiadini, A., & Ognibene, D. (2023). Learning to Prompt in the Classroom to Understand AI Limits: A Pilot Study. In *Proceedings of the 22nd International Conference of the Italian Association for Artificial Intelligence (AIXIA 2023)*, Rome, Italy. <https://doi.org/10.48550/arXiv.2307.01540>
- Unión Europea (2024). *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial*. Diario Oficial de la Unión Europea, L 144, 12 de julio de 2024. <https://tinyurl.com/9kn8ww8c>
- Zamfirescu-Pereira, J., Wong, R., Hartmann, B., & Yang, Q. (2023). Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3544548.3581388>

#### Material complementario

El conjunto de datos utilizados en este estudio están disponibles previa solicitud razonable al autor de correspondencia.

#### Aprobación ética

Este estudio no requirió aprobación de comité de ética, dado que se trató de cuestionarios autoaplicados anónimos sin intervención.

#### Conflictos de interés

Los autores declaran no tener conflicto de interés.

**Como citar:**

García-Beltrán, E. (2026). No es magia, es prompting: el diseño de prompts como competencia emergente en la formación docente. Un estudio desde el modelo CRETA+R [It's Not Magic, It's Prompting: Prompt Design as an Emerging Competence in Teacher Education. A Study Based on the CRETA+R Model]. *Pixel-Bit, Revista de Medios y Educación*, 75, Art. 6.  
<https://doi.org/10.12795/pixelbit.115487>