PIXEL-BIT
e-ISSN: 2171-7966
REVISTA DE MEDIOS Y EDUCACIÓN
MEDIA & EDUCATION JOURNAL

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.

Received: 2025-04-22; **Revised:** 2025-12-01; **Accepted:** 2025-12-02;**Published:** 2026-01-01

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.

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.

KEYWORDS · PALABRAS CLAVES
Teacher Education; Artificial Intelligence; Computer Assisted Instruction; Critical Thinking; Vocational Training.
Educación de Profesores; Inteligencia Artificial; Enseñanza Asistida por Ordenador; Pensamiento Crítico; Formación Profesional.

1. Introduction

In recent years, generative artificial intelligence (GenAI) has radically reshaped technological possibilities across multiple sectors, and education has been no exception. Unlike earlier forms of AI focused on predictive analytics or automation, generative models—such as large language models (LLMs) or systems for visual and multimedia generation—introduce capacities for dialogue, content creation and contextual adaptation that redefine traditional ways of teaching, learning and assessment.

As Bearman et al. (2023) point out, higher education is caught between two emerging discourses around AI: the discourse of imperative transformation—which assumes AI is inevitable and must be integrated urgently—and the discourse of altered authority, which questions how power relations in teaching shift with the incorporation of these technologies. From this perspective, GenAI is not merely another tool; it is a technology that profoundly alters cognitive, pedagogical and social dynamics in the classroom.

The development of GenAI has brought about the emergence of a new educational competence: the ability to design effective prompts. A prompt is far more than a textual instruction; it is a way of structuring knowledge, anticipating responses, contextualising intentions and modulating the behaviour of the AI system. Recent studies emphasise that prompt design requires a combination of linguistic, cognitive, technological and pedagogical skills (Bozkurt & Sharma, 2023; Korzynski et al., 2023). Writing a prompt requires the teacher to make decisions regarding tone, the role assigned to the AI, examples to be included, the type of response expected, and how the interaction will be refined based on the output received. For this reason, authors such as Lo (2023) and Zamfirescu-Pereira et al. (2023) argue that prompt writing constitutes an advanced form of digital literacy that should form part of teachers' professional repertoire. This competence is particularly relevant in contemporary educational contexts where AI does not merely provide technical support but becomes an active agent in the teaching–learning process. Mastering prompt writing enables teachers not only to better manage generative tools but also to design personalised, creative and student-centred learning experiences.

The effective integration of generative AI in educational settings demands a profound transformation in initial teacher education. Digital literacy for teachers can no longer focus solely on instrumental skills; it must incorporate critical understanding of algorithms, data ethics, human–machine interaction and, crucially, the design of interactions through language. In this sense, authors such as Knoth et al. (2024) propose the concept of "AI literacy" as an expanded form of digital literacy that encompasses the ability to interact with, evaluate and make pedagogical decisions about AI-based technologies. Critical digital literacy therefore requires future teachers to develop a reflective stance towards algorithms, the biases they may contain, the power structures they reproduce and the data they process. As Bearman et al. (2023) argue, educators must be equipped not merely as informed users of technology but as ethical mediators capable of making responsible decisions in AI-mediated educational contexts. Prompt design emerges here as a practical pathway to enact this literacy in authentic instructional design scenarios, requiring student teachers to understand how a language-model system "thinks," responds and learns.

Zamfirescu-Pereira et al. (2023) warn that even advanced users may fail to formulate effective prompts, highlighting the need for explicit and systematic instruction in this practice. Far from being a minor technical skill, prompt design entails decision-making about tone, role, format, examples and clarity of purpose. Recent literature also suggests that prompt design can serve as an entry point to critical reflection on AI in the classroom. For example, Bearman et al. (2023) emphasise that educational research on AI must not be reduced to its technical dimension but should also address its sociocultural, epistemological and ethical implications.

Given this panorama, there is a need for pedagogical models that structure and guide the learning of prompt design in educational settings. The CRETA+R model (Context, Role, Examples, Task, Adjust, Refine) is proposed as a framework to support future teachers in the progressive and reflective construction of high-quality prompts, fostering meaningful interactions with generative AI tools. Inspired by principles of instructional scaffolding (Reiser, 2004; Rosenshine, 2012), CRETA+R breaks down the complex task of prompt writing into concrete and manageable steps. Each component serves as a pedagogical cue: establishing the educational context, defining the role the AI should adopt, offering relevant examples, specifying the desired task, adjusting the language for the intended audience and refining the prompt iteratively. In this line, Federiakin et al. (2024) contend that prompt design should be approached as an assessable competence that combines linguistic, heuristic and rhetorical strategies, calling for clear analytical frameworks for educational development. Complementarily, Debnath et al. (2025) propose a systematic framework for studying and teaching prompt engineering in education, arguing that instructional models should guide both the structural composition of the prompt and its iterative improvement process. These perspectives reinforce the relevance of proposals such as CRETA+R, which aim to operationalise this emerging competence through explicit, pedagogically grounded steps.

This structure not only enhances the technical quality of the prompt but also supports metacognitive processes, ethical reflection and formative assessment. Recent studies (Bozkurt & Sharma, 2023; Oppenlaender et al., 2024) agree that well-designed prompts not only produce better AI outputs but also promote deeper learning by requiring users to articulate their communicative intentions and critically evaluate the responses generated. Applying the CRETA+R model in initial teacher education also makes it possible to adapt prompt design to discipline-specific needs, facilitating contextualised curricular integration. Furthermore, the model provides a common framework for evaluating prompts through clear rubrics and iterative improvement processes.

The past two years have seen a substantial increase in research on the integration of AI in initial teacher education programmes. In a systematic review of 138 studies, Bond (2024) identifies AI-supported material design, conversational agents and automated assessment as the most common applications. However, she also highlights the lack of concrete pedagogical proposals to develop critical competencies related to AI. Similarly, Moldavan and Nafziger (2024) worked with pre-service teachers on lesson plans assisted by generative AI, showing that guided prompt design can help student teachers question machine authority, develop critical thinking and reflect on equity and personalisation in learning. The pilot study by Theophilou et al. (2023) offers another relevant example. Conducted with European student teachers, the study explored how prompt-based work can be used in classrooms not only to improve technical skills but also to discuss the limits of AI, its biases and its ethical implications. Across these studies, there is a shared conclusion: teaching AI cannot be limited to technical training but must include pedagogical frameworks that foster critical understanding, ethical design and meaningful interaction with emerging technologies.

2. Methodology

This study adopts a descriptive and exploratory approach aimed at analysing the emerging competence of prompt design among pre-service teachers through the application of the CRETA+R model. This methodological choice is particularly appropriate for educational research focused on underexplored phenomena or those arising in contexts of rapid technological change, such as the integration of generative artificial intelligence in teacher education.

The research is situated within a mixed-methods framework, combining quantitative analysis of general patterns and group comparisons with qualitative analysis of representative examples of students' work. This combination allows not only for describing performance, but also for understanding the discursive, pedagogical and communicative nuances involved in writing educational prompts.

2.1. Participants

The sample consisted of 481 students enrolled in the Master's Degree in Teacher Training for Secondary Education, Upper Secondary Education (Bachillerato), Vocational Education and Training, and Language Teaching. Participants represented a range of subject specialisations—such as Spanish Language and Literature, Mathematics, English, Biology and Geology, Geography and History, and Physical Education. All students were enrolled in a course focused on innovation and digital technologies applied to teaching, within which work with generative AI tools was introduced as part of a structured learning experience. The master's programme is delivered fully online.

The sample showed a balanced distribution in terms of gender and age (range: 22–48 years). All participants held a prior university degree in their subject area, although their familiarity with AI tools varied considerably.

2.2. Instrument

The main data-collection instrument was an individual task requiring students to design an educational prompt to be used with a generative AI model (ChatGPT or equivalent). Students were instructed to create a prompt aligned with a realistic learning situation from their subject specialisation, explicitly applying the components of the CRETA+R model, which consists of:

- Context: a clear and coherent educational scenario.
- Role: the role the AI is expected to adopt (e.g., tutor, evaluator, student).
- Examples: models or illustrations guiding the expected response.
- Task: a precise description of the required output.
- Adjust: adaptation of tone, language or format.
- Refine: instructions for iterative improvement following the AI's initial response.

Each component was assessed by the teaching team using an analytic rubric with four performance levels: Excellent, Good, Adequate and Insufficient. The rubric was collaboratively developed by the instructors and applied consistently for both formative and research purposes.

In addition to component-level evaluations, the dataset included variables such as the student's final grade in the course, their mark in the final on-site examination, and the specific grade obtained on the generative-AI activity.

To ensure reliability in the assessment process, the analytic rubric was applied by a team of four instructors who completed a prior calibration session. During this session, instructors jointly reviewed real examples of prompts and discussed operational criteria for each performance level to minimise inter-rater variability. The rubric included detailed descriptors for each CRETA+R component across the four levels of achievement, covering clarity of context, appropriateness of role assignment, quality of examples, accuracy of the task description, linguistic adjustment and iterative refinement. This process ensured maximum consistency and transparency, essential given that the evaluations formed the basis for both the quantitative and qualitative analyses.

2.3. Variables analysed

The dataset enabled the analysis of the following variables:

- Master's specialisation (categorical): grouped into standardised disciplinary areas.
- Prompt quality (ordinal): performance level in each of the six CRETA+R components.
- Prompt activity grade (continuous): numerical mark for the task.
- Course grade (continuous): final mark in the module.
- Final on-site exam grade (continuous).

These variables were analysed both independently and relationally to explore patterns of performance by specialisation, correlations between prompt quality and academic results, and components with stronger or weaker development.

2.4. Data analysis procedure

Data were processed through a mixed-methods approach integrating statistical analysis and qualitative review.

2.4.1. Quantitative analysis

- Descriptive statistics (means, frequencies, standard deviations).
- Comparative analysis by specialisation (Kruskal–Wallis tests and boxplots).
- Correlation analysis between grades and CRETA+R performance (Spearman's rho coefficients).

2.4.2. Qualitative analysis

A focused review of a selection of representative prompts chosen for their performance level and explanatory potential. This review enabled the identification of discursive patterns, recurring strategies and common errors in the application of each CRETA+R component.

All quantitative processing and visualisation were carried out using JASP version 0.19.3 for macOS, an open-source statistical tool offering robust procedures and interactive graphical outputs. JASP was selected for its accessibility and transparency, making it particularly suitable for educational contexts that promote critical and reproducible analytical practices.

3. Results

The quantitative analysis provided a detailed picture of student performance in prompt design using the CRETA+R model. Descriptive statistics indicated a high average grade for the activity ($M = 8.10$; $SD \approx 0.49$), suggesting generally strong performance across the cohort. However, the presence of outliers in some specialisations (such as Mathematics or Biology and Geology) highlights notable individual variability. The comparative analysis by specialisation, conducted using the non-parametric Kruskal–Wallis test, yielded no statistically significant differences ($H = 5.13$; $p = 0.400$). This suggests that performance in the prompt-design task was not substantially dependent on students' disciplinary backgrounds.

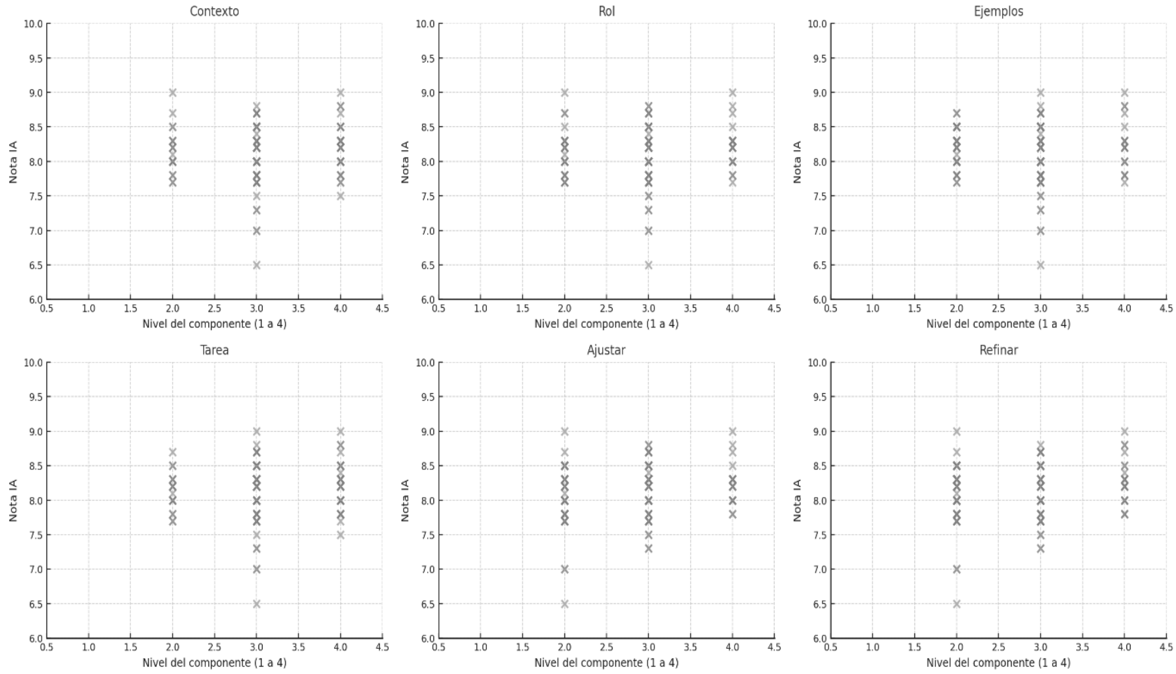
To examine relationships between performance in the CRETA+R components and final grades, Spearman correlations were calculated using ordinal encoding of rubric levels. The correlation coefficients were low for all components, with only "Adjust" showing a weak but statistically significant correlation ($\rho = 0.111$; $p = 0.031$). This result suggests that greater precision in fine-tuning the prompt may be slightly associated with higher overall performance. The remaining components showed correlations very close to zero and were not statistically significant, reinforcing the idea that success in the task is not driven by any single component but emerges from a more complex interplay of factors. The scatterplots (figure 1) support this interpretation, revealing flat distributions with no clear patterns and indicating the need for further investigation into variables that may influence successful prompt design.

3.1. Overall evaluation by CRETA+R component

Most students achieved ratings in the "Good" and "Excellent" categories, with Context and Task being the strongest components. In contrast, Adjust and Refine showed a higher concentration of ratings in the "Adequate" category, suggesting that students encountered more difficulty in aspects related to tone adaptation, language adjustment and iterative refinement. Figure 2 displays the distribution of performance levels across the six components of the CRETA+R model.

Figure 1

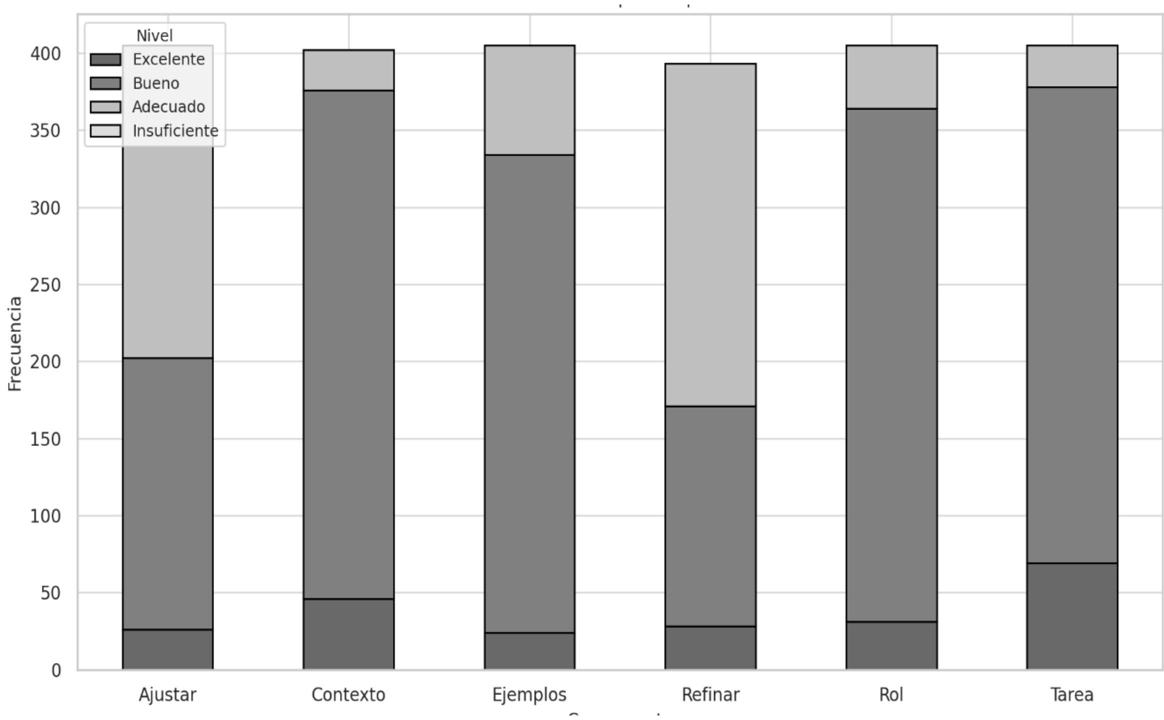
Correlation Between Performance in CRETA+R Components and Activity Grade



Source: own elaboration.

Figure 2

Distribution of Evaluation Levels Across CRETA+R Components



Source: own elaboration.

Each bar represents one of the six components. The evaluation scale ranges across four levels—Excellent, Good, Adequate and Insufficient—coded in varying shades of grey. The components with the strongest performance are Context, Role, Task and Examples, all showing a clear predominance of "Good," with relatively few "Adequate" or "Excellent" ratings. This pattern suggests that most students fulfilled the basic quality criteria in these components, although without consistently reaching the highest levels. Context stands out as one of the components with the highest proportion of positive evaluations (Excellent + Good), potentially reflecting students' familiarity with providing contextual information in academic tasks.

In contrast, the components showing the greatest difficulty were Adjust and Refine, both displaying a substantially higher proportion of ratings in the "Adequate" category. This indicates that these aspects of prompt design were more challenging for students, likely due to the linguistic, metacognitive or technical maturity required to adapt tone or revise prompts iteratively. It is noteworthy that the "Insufficient" level was virtually absent. The absence of significant proportions of "Insufficient" suggests a minimum acceptable level of performance across all components, possibly attributable to effective instructional guidance or the clarity of the rubric.

From a pedagogical perspective, these findings suggest that students have consolidated the more structural components of prompt design (setting context, defining the task, specifying the role), while the more metacognitive and revision-oriented components (adjusting and refining) require additional instructional support. Potential approaches include scaffolded activities, peer feedback exercises and guided iterative revision using AI tools.

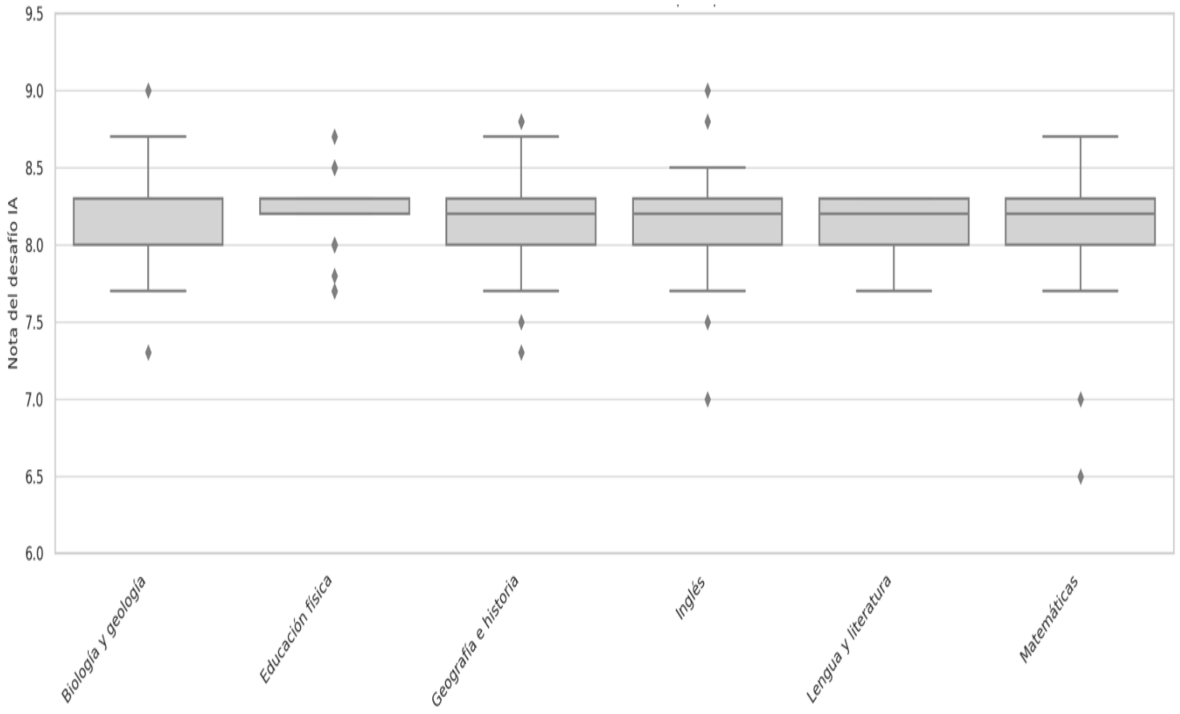
3.2. Analysis by specialisation

The grades obtained in the prompt-design activity varied across specialisations. Most specialisations exhibited relatively high mean scores, clustered around 8.0–8.3, indicating solid overall performance. Several specialisations displayed narrow interquartile ranges, suggesting low variability and a consistent application of the rubric. Physical Education showed minimal dispersion (almost no visible boxplot), indicating that most students received the same grade. By contrast, Mathematics presented a lower distribution with outliers around 6.5, suggesting some difficulty among students in adapting to the requirements of the task. This may be linked to less familiarity with pedagogical language or reflective writing. English, Spanish Language and Literature, and Geography and History showed similar distributions around 8.2, with slight negative asymmetry caused by isolated low-performing cases. Biology and Geology and Mathematics had more low outliers, evidencing greater challenges for some students.

Differences across specialisations may reflect varying levels of pedagogical or technological literacy, highlighting the need for discipline-sensitive instruction in prompt design. Specialisations with lower performance may benefit from more explicit scaffolding (e.g., guided sequences, contextualised examples, iterative feedback). The absence of very high outliers suggests that, although overall performance was good, very few submissions were truly exceptional—indicating scope for fostering greater creativity or critical depth in working with AI. Additionally, students in Mathematics, English, and Spanish Language and Literature tended to obtain higher mean scores across most components. Conversely, specialisations such as Physical Education and Biology and Geology showed more concentration in middle or adequate performance levels.

Figure 3

Distribution of Activity Grades by Specialisation (Boxplots)



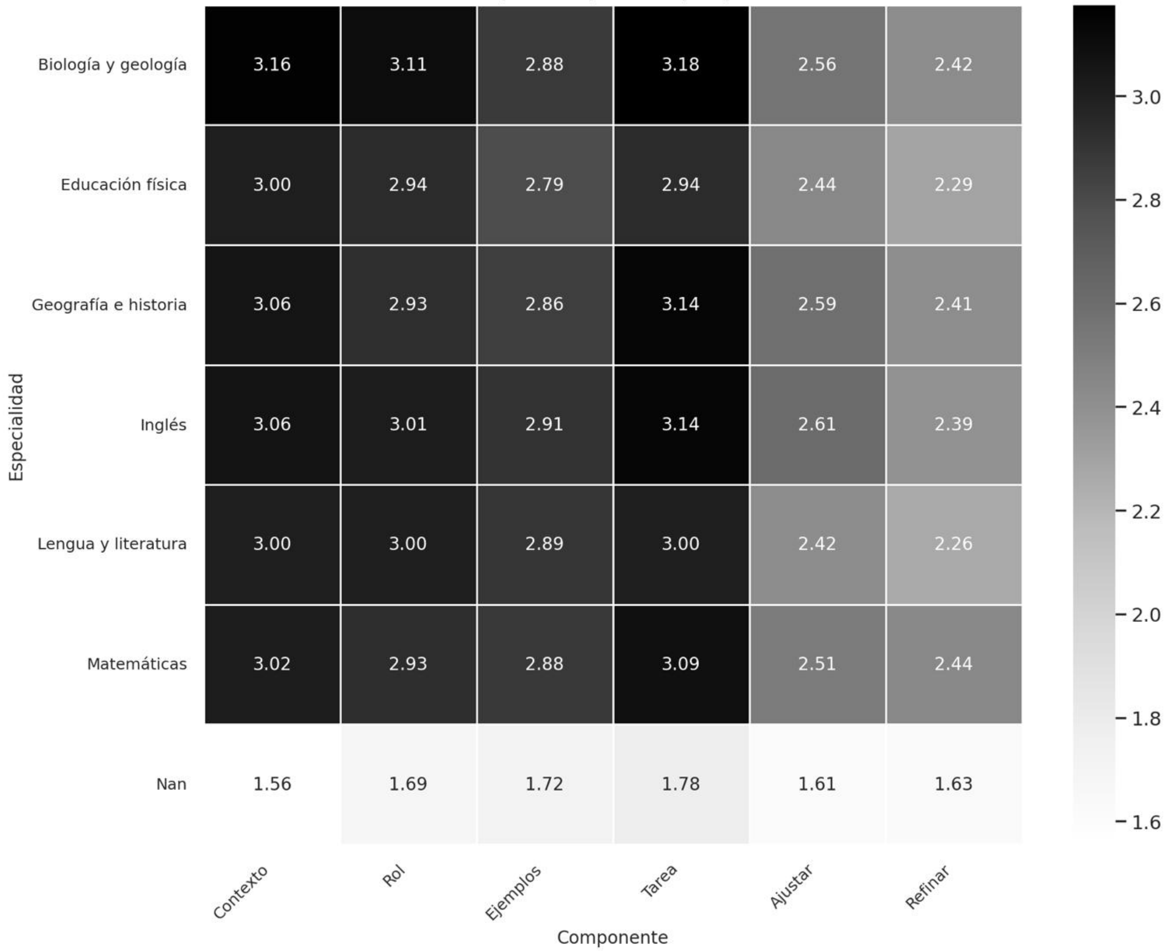
Source: own elaboration.

Specialisations with a larger number of students also show a wider distribution toward the higher evaluation levels. This pattern is generally repeated—with some nuances—across the remaining components of the CRETA+R model. The heatmap visualisation (Figure 4, next page) displays the mean scores for each CRETA+R component by master's specialisation, using a scale from 1 (Insufficient) to 4 (Excellent). Overall, ratings tend to cluster around the "Good" level (3) across most components and specialisations, indicating solid performance while still leaving room for improvement. Specialisations such as Spanish Language and Literature, Geography and History, and Educational Guidance show slightly above-average scores in nearly all components, particularly in **Context** and **Role**.

In contrast, specialisations such as Physical Education, Mathematics and Philosophy display somewhat lower values, especially in the more complex components **Refine** and **Adjust**, which may reflect less experience with the discursive or reflective tasks inherent to educational prompt design. This pattern suggests that, although the CRETA+R model is broadly applicable across disciplines, some specialisations require more targeted pedagogical scaffolding to improve performance in components related to critical revision and iterative refinement.

Figure 4

Heatmap of Mean Scores by CRETA+R Component and Specialisation

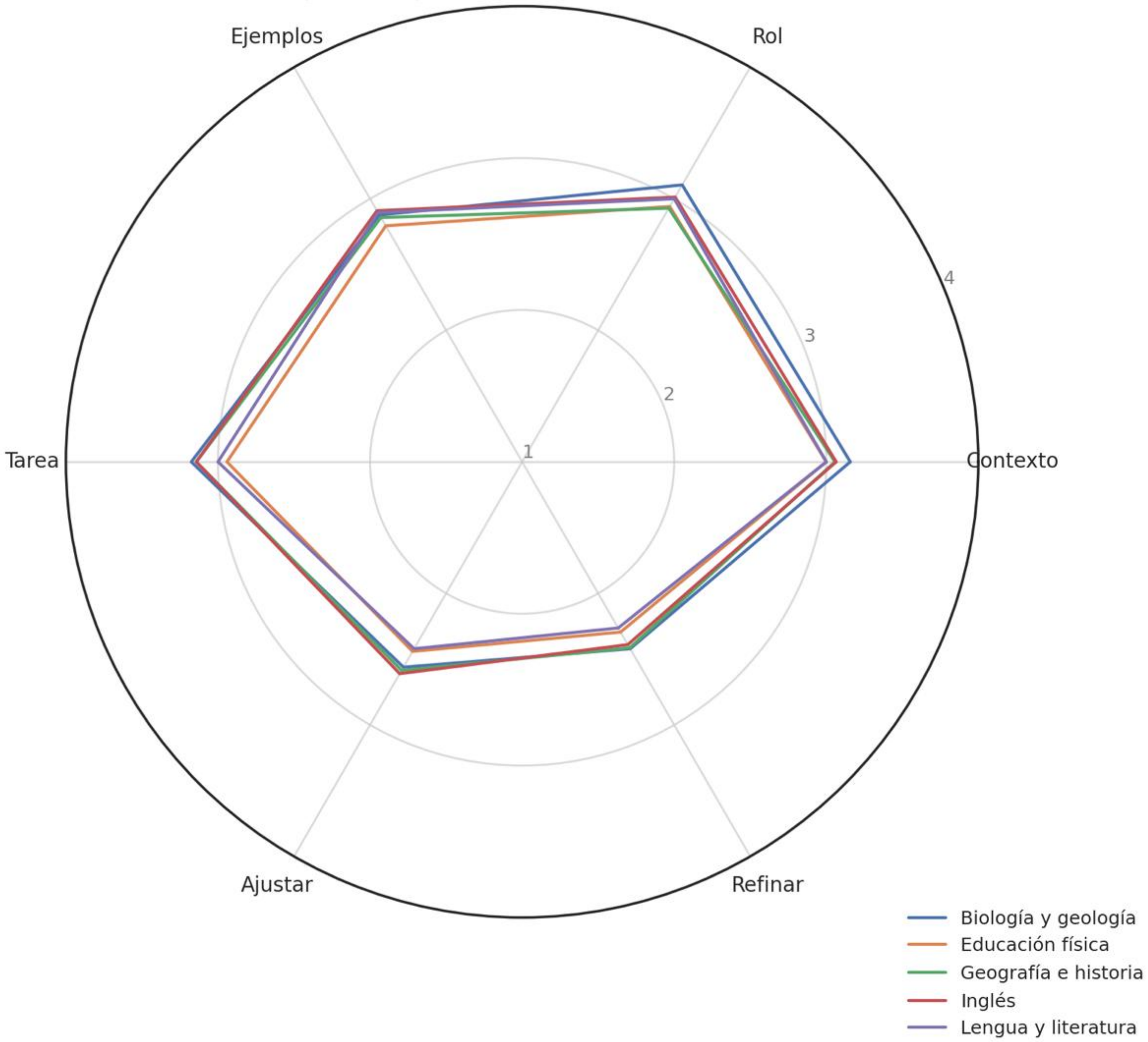


Source: own elaboration.

The radar chart (figure 5) compares the average profile by specialisation across CRETA+R components. A generally balanced pattern emerges, with scores close to “Good” (3), although notable differences appear among areas. Spanish Language and Literature and Geography and History show broader and more consistent profiles, particularly in Context, Role and Task. Mathematics and Biology and Geology demonstrate lower performance, especially in Refine and Adjust, indicating challenges in revision and iterative improvement. This visualisation further highlights the value of CRETA+R in identifying discipline-specific learning needs.

Figure 5

Comparative Profile by Specialisation (Radar Chart)



Source: own elaboration.

3.3. Qualitative analysis findings

The qualitative analysis of a representative sample of prompts revealed discursive patterns not visible in the quantitative results. In structural components (Context, Role, Task), students generally offered clear and coherent descriptions, although some contexts were excessively broad (e.g., "develop a topic from my subject") and lacked specificity regarding academic level or pedagogical goals. Differences also emerged by specialisation in the use of examples: students from Language, English and Humanities subjects tended to include detailed and relevant models, whereas other areas—such as Physical Education or Technology—often offered either minimal or uninformative examples, limiting the AI's ability to generate precise responses.

The components presenting the greatest difficulty were Adjust and Refine. In Adjust, several students did not adequately adapt tone, language level or format to the intended audience, producing instructions that were either overly technical or overly informal. In

Refine, most prompts did not include any indication of iterative revision, confirming a limited understanding of the cyclical nature of interactions with generative AI. Only a small subset of students incorporated revision strategies (e.g., "if the response does not meet the requirements, reformulate it as follows…"), demonstrating higher metacognitive maturity. Overall, these qualitative findings deepen the interpretation of the quantitative patterns and reinforce the need for greater instructional support in the adjustment and iterative stages of prompt design.

4. Discussion

The implementation of the CRETA+R model made it possible to identify performance patterns and areas of difficulty that align with current tensions surrounding AI literacy in higher education. Several authors concur that prompt design represents a new form of digital literacy, comparable to advanced skills in critical thinking and communication (Lo, 2023). In this regard, the master's students who took part in this study demonstrated solid performance in structural components such as *Context* and *Task*, while exhibiting persistent difficulties in aspects that demand greater communicative awareness, such as *Adjust* and *Refine*. Teaching prompt design therefore extends beyond technical proficiency: it involves thinking *with* the machine, anticipating interpretations, modulating instructions, and learning to iterate.

The variability observed across specialisations suggests that disciplinary background significantly influences how students engage with each component of the model. While students in Spanish Language and Literature, English, and Mathematics displayed more balanced and consistent profiles, others—such as Physical Education and Biology and Geology—showed more pronounced weaknesses, particularly in refining and adjusting language. This pattern echoes findings reported by Silva (2024) in the context of chemistry education, where students initially displayed a superficial understanding of prompt design and resorted to copy-and-paste strategies before developing more sophisticated approaches. These variations may stem partly from differences in prior experience with structured academic expression or from the didactic traditions prevalent in each discipline. As Bozkurt and Sharma (2023) argue, the "art of whispering to the algorithm" requires skills ranging from clarity of formulation to creativity and digital empathy—abilities not uniformly developed across subject areas.

From a qualitative standpoint, the analysis of representative examples revealed that *Refine* was the least developed component for most students. This finding aligns with the results of Eager and Brunton (2023), who highlight the importance of teaching iterative strategies when working with generative AI, moving beyond superficial or one-way use. The absence of revision or prompt adjustment after receiving an AI response points to the need to strengthen the metacognitive dimension of this competence, incorporating mechanisms for self-evaluation and progressive improvement. Difficulties also emerged in the use of examples, particularly in areas such as Physical Education or Technology, where students did not always provide clear or pedagogically relevant models for the AI. As noted by Ranade et al. (2024), effective prompts must clearly articulate context, audience and expected response type—an aspect that requires rhetorical literacy not yet well established among all future teachers. This gap suggests that prompt-design competency cannot be developed solely from a functional perspective; it must also address principles of communicative design, discourse theory, and the semiotic interaction between humans and technology. Additionally, the fact that the highest-performing students showed greater reflective capacity in the adjustment and refinement phases aligns with what Sajja et al. (2024) describe as “intelligent personalisation of learning,” a critical skill in AI-assisted environments.

The findings also highlight the need to explicitly include prompt design in teacher education programmes as an emergent pedagogical competence, aligned with European guidelines on AI in education (European Commission, 2022) and with Regulation (EU) 2024/1689, which emphasises educators’ responsibility in the ethical, transparent and safe use of AI technologies. From a critical standpoint, Bearman et al. (2023) argue that current discourses on AI in education often oscillate between technodeterminist enthusiasm and alarmist rejection. Against this backdrop, the present study provides concrete evidence of how future teachers can begin to relate to AI not only as users but as reflective designers of AI-mediated learning experiences. As Baidoo-Anu and Owusu Ansah (2023) observe, the widespread use of tools such as ChatGPT in higher education requires ethical guidance, critical training and clear institutional policies. Developing prompt-design competence must therefore be accompanied by reflection on the limits and responsibilities associated with AI use in the classroom.

In this regard, the CRETA+R model proves valuable not only as a structure for writing prompts, but also as a didactic mediator to support thinking *with* and *about* AI. Its design aligns with recommended strategies in the literature, such as task decomposition (Karakaya, 2025) and iterative refinement (Higginbotham & Matthews, 2024). The use of CRETA+R functioned as an effective scaffolding strategy, helping students organise their thinking around generative AI. The model not only supports formative assessment of prompt-design work but, as Korzyński et al. (2023) suggest, may also serve as a structural foundation for developing prompt-engineering competencies as part of teachers’ professional skillsets. The fact that *Task* and *Context* received the highest evaluations indicates that the model offers strong support for components closely related to instructional planning, whereas the more novel components—such as iteration or tonal adjustment—require more time and practice to consolidate.

Finally, the findings underscore the value of situated learning. As demonstrated in the workshop analysed by Graux et al. (2024), mastery of prompt engineering does not emerge solely from exposure to examples, but through trial, error, feedback and reconstruction. Embedding this competence in collaborative settings—where students can share, critique and iteratively refine prompts—can enhance both technical proficiency and critical–reflective engagement.

5. Conclusions

This study has explored, from both an empirical and pedagogical perspective, the development of prompt-design competence among students enrolled in a Master’s Degree in Secondary Teacher Education. The findings confirm that this competence is not only relevant within the current context of digital transformation, but also requires targeted instructional strategies to be effectively strengthened. The data indicate that future teachers are capable of producing clear and coherent instructions—particularly in the *Context* and *Task* components—yet face greater challenges in more sophisticated stages of the process, such as linguistic adjustment and iterative refinement. These limitations are consistent with barriers identified in other studies on AI literacy (Zamfirescu-Pereira et al., 2023; Knoth et al., 2024), reinforcing the need to integrate systematic approaches such as the CRETA+R model into initial teacher education.

Furthermore, the comparison across specialisations reveals that disciplinary background significantly shapes performance profiles. Areas such as Language, English and Mathematics demonstrated greater overall consistency, whereas others—such as Physical Education—showed a clearer need for enhanced instructional support. These findings highlight the importance of tailoring pedagogical strategies to disciplinary characteristics when developing AI-related competencies.

In light of the evidence gathered, several pedagogical recommendations are proposed to support the effective integration of prompt design as an emerging competence in teacher education:

Table 1

Pedagogical Recommendations for Developing Prompt-Design Competence in Teacher Education

Area	Recommendation	Rationale
Curricular integration	Include prompt design as an explicit topic in courses on didactics, educational innovation or digital competence.	Responds to the need for AI literacy in initial teacher education (European Commission, 2022; Knoth et al., 2024).
Methodological scaffolding	Use models such as CRETA+R to guide and structure prompt writing, incorporating progressive examples and collaborative analysis.	Enhances prompt quality and promotes metacognition (Korzyński et al., 2023).
Iteration and refinement	Design activities requiring multiple rounds of refinement following AI interaction, with explicit critical reflection.	Strengthens adaptive and metacognitive skills (Bozkurt & Sharma, 2023; Lo, 2023).
Formative assessment	Develop CRETA+R-based rubrics including criteria for clarity, adaptability, linguistic adjustment and iterative improvement.	Supports effective feedback and progress monitoring (González-Calatayud et al., 2021).
Disciplinary perspective	Adapt examples and prompt-design tasks to the needs of each specialisation, ensuring contextualised learning.	Addresses the differences observed across subject areas (Luckin et al., 2024; present results).
Ethical and critical focus	Incorporate opportunities to discuss risks, biases and limitations of generative AI, especially regarding automated assessment.	Aligns with Regulation (EU) 2024/1689 and proposals for inclusive AI (Roscoe, 2023; Bearman et al., 2023).

Integrating these practices can support the development of teachers capable of interacting critically, creatively and ethically with AI-based tools, contributing to more inclusive, reflective and contextually grounded educational environments.

References

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.caeai.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

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

Funding

This study did not receive financial support from any institution.

Suppelemntary material

The dataset used in this study is available upon reasonable request to the corresponding author.

Ethical Approval

This study did not require approval from an ethics committee, as it involved anonymous self-administered questionnaires without any form of intervention.

Conflict of interest

The authors declare no conflict of interest..

How to cite:

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