Prompting as a Professional Competence: Analysing Pre-service Teachers' Prompts and Critical Appraisal of AI-Generated Lesson Plans

Maria Rosaria Re[1], Stefania Capogna[2] and Alessandra Natalini[3]

[1]Link Campus University, Rome, Italy, m.re@unilink.it
[2]Link Campus University, Rome, Italy, s.capogna@unilink.it
[3]Link Campus University, Rome, Italy, a.natalini@unilink.it

Abstract

The widespread adoption of generative artificial intelligence is reshaping teachers' professional competences, requiring not only the ability to integrate digital technologies ef ectively into teaching, but also to critically evaluate AI-generated outputs and interact with AI systems through ef ective prompting. Within this context, the present study investigates the development of prompt literacy and critical reflection in initial teacher education. The research was conducted at Link Campus University within the Educational Technology Laboratory of the five-year Primary Teacher Education degree programme (LM-85bis). At the end of the laboratory, 133 pre-service teachers designed a technology-enhanced lesson plan, asked a generative AI system to produce a lesson on the same topic while documenting the prompt used, and finally wrote a critical comparison of the two lesson plans. The corpus was analysed through a mixed-methods content analysis adopting a quantitizing strategy. Prompts were coded according to indicators of quality and contextualisation, while students' reflections were analysed using established frameworks of critical reflection. The study explores the relationship between prompt quality and the depth of critical evaluation of AI-generated outputs, proposing prompt literacy as an emerging professional competence to be fostered within initial teacher education.

Keywords: *Artificial Intelligence in Education, Assessment of Learning, New Teaching Methods, Educating students for the future job market.*

ISSN 2991-9371

1. Theoretical framework

The rapid spread of generative artificial intelligence is reshaping expectations regarding teachers' professional competencies [1]-[3]. Beyond the well-established need to integrate technological, pedagogical, and content knowledge [4], teachers are now expected to critically evaluate AI-generated instructional materials and to communicate their pedagogical intentions effectively to AI systems. This emerging set of competencies is increasingly discussed within the broader concept of AI literacy [5] and, more specifically, in relation to the ethical and informed integration of AI tools into educational practice [6]. European policy frameworks reflect this evolution: the "Digital Competence Framework for Citizens" emphasises the critical evaluation of data, information, and digital content [7], while DigCompEdu defines the professional digital competencies required of educators, including the critical selection and creation of digital resources [8].

1.1 From digital competence to teachers' AI literacy

The digital transformation of education has reshaped teachers' professional competencies, extending beyond traditional digital literacy to include the critical integration of technologies into teaching and learning. Within this context, the European DigCompEdu framework provides a key reference for teachers' digital competencies. However, the rapid diffusion of Artificial Intelligence (AI) has highlighted the need to expand these models through the concept of Artificial Intelligence Literacy (AI Literacy), defined as the knowledge, skills, and dispositions required to understand, use, critically evaluate, and responsibly govern intelligent systems [9], [5]. AI Literacy extends digital competence by including the understanding of AI principles, the critical interpretation of AI-generated outputs, awareness of biases and limitations, and consideration of ethical, social, and pedagogical implications [5], [10]. Within Education 4.0, teachers act as learning facilitators, using AI to support students' cognitive, collaborative, and reflective processes without replacing the relational and pedagogical dimensions of teaching [11]. Developing AI Literacy requires the integration of technical, pedagogical, ethical, and transversal competencies, including critical thinking, problem solving, creativity, and algorithmic thinking [5]. Teacher preparation should therefore be supported by AI-specific training and continuous professional development, which can enhance digital competencies, confidence, and willingness to integrate AI into educational practice [11], [10]. Kasinidou et al. [10] identify a positive correlation between teachers' digital competence and their propensity to use and teach AI. Overall, AI Literacy represents an essential evolution of teachers' digital competence, integrating technological, pedagogical, ethical, and social dimensions necessary for the critical and responsible use of AI in education [5], [10], [11]. The studies highlight that prompt engineering is an important skill for interacting effectively with large language models (LLMs). Although prompting may appear intuitive and accessible even to non-experts in AI, formulating effective prompts and interacting with LLMs can be challenging. In particular, non-expert users tend to experiment with prompts in an opportunistic and unsystematic manner, developing their strategies primarily through trial and error. Among the main difficulties are expectations derived from human-to-human interactions and a tendency to overgeneralize based on the results obtained. Furthermore, prompt engineering can be regarded as a form of programming, as it enables users to define rules, customise interactions,

and guide the characteristics of generated outputs. To this end, a catalogue of prompt patterns, that is, structured and reusable strategies for addressing recurring problems in interactions with LLMs, can be adapted to different contexts and combined. Overall, the research underscores the importance of developing greater literacy regarding LLMs and prompting, highlighting that the ability to formulate effective instructions should not necessarily be regarded as intuitive. This aspect is particularly relevant in educational contexts, as the absence of specific training in prompting may influence the quality of the outputs generated by students [12], [13].

1.2 The Role of Artificial Intelligence in Teaching and Learning Processes

The adoption of Artificial Intelligence (AI) in education is transforming teaching design, delivery, and assessment. The literature emphasises that AI should support teachers' decision-making, personalise instruction, and enhance educational practice rather than replace teachers. Recent guidelines therefore promote a human-centred approach based on appropriate governance, staff training, risk assessment, and continuous monitoring [14]. AI can support personalised and inclusive learning by adapting educational content to students' needs, providing accessibility tools, and enabling learning analytics to inform targeted instructional interventions [11]. It can also automate administrative and repetitive tasks, allowing teachers to focus on instructional design, educational relationships, and active learning. Its effectiveness depends on the integration of disciplinary, pedagogical, and technological competencies, as highlighted by the TPACK framework, as well as on the development of algorithmic thinking. AI integration also requires attention to ethical, legal, and organisational issues. DigComp and DigCompEdu emphasise AI Literacy, responsible use, transparency, accountability, and digital well-being, while the European Artificial Intelligence Regulation (AI Act) [15] establishes specific requirements for AI systems in education, including human oversight, transparency, traceability, and risk management. Overall, the effective adoption of AI depends less on technological capabilities than on the ability of educational institutions to integrate AI into coherent, pedagogically informed instructional design, balancing innovation, inclusion, ethical responsibility, and continuous professional development.

1.3 Enablers, Challenges, and Future Directions for AI Literacy Development

Although the literature recognises the transformative potential of Artificial Intelligence (AI) in educational processes, there is broad consensus that its effective integration depends primarily on teachers' preparedness and the organisational context in which they operate. Empirical evidence indicates that the availability of advanced technologies does not automatically lead to educational innovation; rather, the successful adoption of AI is closely linked to the development of teachers' professional competencies, institutional support, and the presence of an organisational culture that fosters innovation and change [11]. Among the key enabling factors identified in the literature, continuous professional development emerges as the most influential. Teachers who participate in structured training programs develop higher levels of digital competence, greater familiarity with AI tools, and a stronger willingness to integrate them into their instructional practice. However, professional development should not be

limited to the acquisition of technical skills. Instead, it should promote critical reflection on the pedagogical, ethical, and social implications of intelligent technologies, fostering the professional competencies that enable teachers to evaluate both the opportunities and the limitations of educational innovation in an informed manner [9]. Alongside professional development, the school context plays a crucial role. Educational leadership, the availability of digital infrastructure, access to technological resources, professional learning communities, and organisational support systems are all essential conditions for fostering the development of AI Literacy. Indeed, the literature review indicates that most of the factors positively influencing technology adoption depend on conditions external to individual teachers, confirming that educational innovation is a collective process requiring a shared strategic vision and sustained institutional investment [11]. At the same time, numerous studies identify barriers that continue to hinder the widespread adoption of Artificial Intelligence in schools. Among the most frequently reported challenges are the limited availability of specialised training opportunities, inadequate technological infrastructure, insufficient time for pedagogical experimentation, and technophobia, defined as the insecurity or resistance that some teachers experience toward emerging technologies. These factors contribute to widening the gap between the increasing availability of AI-based tools and their actual implementation in educational settings. International research further demonstrates that teachers' digital competencies significantly influence their attitudes toward Artificial Intelligence. The study conducted by Kasinidou et al. [10] among Cypriot teachers found a positive correlation between teachers' level of digital competence and their willingness to adopt AI tools in their professional practice. Teachers with higher levels of digital competence exhibited greater confidence in emerging technologies, a better understanding of how AI systems function, and a stronger intention to use AI both as a pedagogical support tool and as a subject of instruction. Conversely, lower levels of digital literacy were associated with more cautious or neutral attitudes and with a more limited integration of AI into educational activities. Another important theme emerging from the literature concerns the role of ethics as a fundamental component of AI Literacy. Ethical competence should not be regarded as an additional dimension beyond digital competence but rather as an integral aspect of the responsible use of intelligent systems. Teachers are expected to recognise the risks associated with algorithmic bias, opaque decision-making processes, the misuse of personal data, and the potentially discriminatory effects of AI applications. At the same time, they should be capable of guiding students toward a critical, responsible, and informed use of digital technologies [5], [10]. Ethical reflection is closely connected to the issue of educational equity. The growing diffusion of Artificial Intelligence makes it increasingly necessary to address the digital divide not only in terms of access to devices and connectivity but also in terms of access to the competencies required to use these technologies effectively. Within this perspective, teachers assume a strategic role as mediators of innovation, responsible for ensuring equitable learning opportunities through the selection of accessible, inclusive, and learner-centred technologies that meet students' diverse needs. The ability to reduce digital inequalities is therefore regarded as a fundamental component of AI Literacy and, more broadly, of the social responsibility inherent in the teaching profession [9]. Taken together, the studies reviewed converge in highlighting that the diffusion of Artificial Intelligence across educational systems requires a systemic perspective integrating professional development, organisational innovation, and institutional governance. The effectiveness of AI depends not solely on technological advancement but on the capacity of educational institutions to create ecosystems that support professional learning by providing continuous professional development opportunities, adequate technological infrastructure, and educational policies oriented toward responsible innovation. In light of these findings, AI Literacy emerges as

a strategic competency for contemporary education, in which technological, pedagogical, ethical, and organisational dimensions are deeply interconnected. The development of these competencies constitutes a prerequisite for ensuring that Artificial Intelligence contributes meaningfully to improving the quality of teaching and learning while promoting educational innovation that is sustainable, inclusive, and human-centred.

2. Research and questions

This study investigates how pre-service primary teachers engage with generative AI as an instructional design partner, and whether the way they interrogate the AI relates to how critically they appraise its output. Building on the distinction between self-reported and demonstrated competence, the analysis focuses on two performance-based indicators drawn from an authentic laboratory task: the quality of students' prompts and the depth of their critical reflection. Three research questions guide the study.

RQ1: What features characterise the prompts that pre-service teachers formulate to have an AI system design a lesson, and to what extent do students transfer into their prompts the instructional design elements they master (context, objectives, methodologies, inclusion)?

RQ2: What level of reflective depth do students reach when critically comparing their own lesson design with the AI-generated one?

RQ3: Is there a relationship between prompt quality (prompt literacy) and the depth of critical reflection?

3. Methodology

The present study addresses this challenge within the context of initial teacher education in Italy. In the Italian education system, prospective preschool and primary school teachers obtain their teaching qualification by completing a five-year single-cycle Master's degree program in Primary Teacher Education (LM-85bis). This program integrates disciplinary coursework, compulsory laboratory activities, and school-based teaching placements, leading directly to teacher certification. As part of the degree program, all students are required to attend a face-to-face Educational Technology laboratory designed to develop operational competencies in the pedagogical use of digital technologies. The laboratory focuses on the search, selection, and critical use of digital educational resources; the design of technology-enhanced learning units; and the development of multimedia instructional materials for preschool and primary school pupils. During the final synchronous session of the laboratory, held at Link Campus University during the 2025–2026 academic year, students completed a structured final assignment. First, they designed a technology-enhanced lesson following an instructional planning template consisting of eleven elements: educational context and special educational needs, subject

areas, timing, learning objectives, content, prerequisites, instructional phases and activities, teaching methodologies, digital tools, learning environments, and assessment. Subsequently, they prompted a generative AI system to produce a lesson on the same topic, documenting both the exact prompt used and the AI-generated output. Finally, they prepared a critical comparison of the two instructional designs. It should be emphasised that the generative AI system used by the students was freely chosen by them, which may have influenced both the AI-generated output and the students' reflections. Furthermore, the students were expected to complete the activity independently at home, without a predetermined time limit and without receiving specific training on how to use prompts effectively beyond what was covered during the workshop. A total of 133 assignments were collected. Of the participating students, 131 were female, and 2 were male; 17 were in the second year of the Primary Teacher Education degree program, while 116 were in the third year. The authors conducted a mixed-methods content analysis [16] of the 133 assignments, adopting a quantitizing strategy in which qualitative codes were subsequently analysed quantitatively [17]. Each prompt was coded across six ordinal dimensions: contextual completeness (the number of the eleven instructional planning elements included in the prompt: 0–3 = 1; 4–7 = 2; 8–11 = 3), specificity of the request (generic = 1; partially constrained = 2; constrained = 3), use of professional vocabulary (absent = 1; present = 2), attention to inclusion (absent = 1; generic reference = 2; explicit request for adaptations = 3), structural organization (single block = 1; articulated structure = 2), and iterative prompting (single interaction = 1; successive refinements = 2). The scores for the six dimensions were summed to obtain a composite Prompt Literacy Index (range = 6–15), which was subsequently grouped into ordinal proficiency levels. In addition, contextual completeness was retained as a raw count (0–11) for descriptive analyses. Students' critical reflections were coded using the four-level framework developed by Kember, McKay, Sinclair, and Wong [18], based on Hatton and Smith's model of reflective thinking [19], comprising descriptive writing, understanding, reflection, and critical reflection. The coding framework was informed by the core dimensions of critical thinking identified in Facione's expert consensus [20]. To establish coding reliability, both authors independently double-coded a subsample of 30 assignments (approximately 27% of the dataset). Inter-rater reliability was assessed using Cohen's kappa [21] and Gwet's AC1 coefficient [22]. Coding discrepancies were discussed and resolved collaboratively, leading to refinement of the codebook. The remaining assignments were subsequently coded by one of the authors. Frequencies, contingency tables, and non-parametric statistical analyses (Spearman's rho and chi-square tests) were then employed to examine the relationship between Prompt Literacy and the depth of students' critical reflection.

4. Results

Regarding prompt evaluation, the results show that among the 133 participants, the most frequently obtained score was 10 (n = 31), followed by 9 (n = 24). Only three participants achieved a score of 14, while just one participant obtained the maximum score of 15 established by the prompt evaluation rubric. Three cases were excluded from the analysis because the prompts were not interpretable. As reported in Table 1, the most frequently identified category in the analysis of students' reflections on the use of AI for lesson planning was reflection, with a frequency of 100, followed by critical reflection, which occurred 23 times. These findings

indicate that most students focused their analyses on the advantages and limitations of using AI in instructional design, developing reflective considerations grounded in their experience during the Educational Technology laboratory.

Indicators for reflection on the use of AI for lesson planning	Category Frequency
Descriptive writing: the student simply reports what the AI and the student did	2
Understanding: the student shows conceptual elaboration but without reference to experience	5
Reflection: the student relates the experience to professional knowledge (e.g., "The AI didn't consider the real time of a class with 25 students because it doesn't know my context")	100
Critical reflection: this implies a shift in perspective on the use of AI in the teaching profession.	23

Table 1. Frequency Distribution of Reflection Levels in the Evaluation of Reflections on the Use of AI for Lesson Planning

To facilitate the interpretation of the final result, two illustrative examples of prompts and reflections are presented in the table below (Table 2).

Prompt example	Critical reflection example
Act as an inclusive instructional design expert for elementary school. Create a six-hour science lesson for a first-grade class (18 students, including one with mild autism and two with NAI) focused on exploring the five senses in nature. Use methodologies such as outdoor education and cooperative learning. Include the use of digital tools such as tablets and Book Creator. Organise the response according to the classic principles of phased instructional planning.	*Designing a lesson without the help of AI made me realise how much teaching is a living process, made of intuition, trial and error, second thoughts, and above all, attention to the real children I imagined before me. When I wrote my "human" version, I realised I was spontaneously thinking about their faces, their timing, their difficulties, and their strengths. Every choice stemmed from a concrete need, not an abstract structure. When I asked AI to design the same lesson, I immediately noticed a difference: its version was much more organised, complete, and linear. It seemed almost "perfect," as if every part was in the right place. This made me realise that AI is capable of seeing a very clear and systematic vision, which can be truly useful for beginners like me, since I don't yet teach elementary school, nor even preschool, unfortunately, and I'm afraid of forgetting something. In this sense, AI provides security: it shows you a solid structure, reminds you of elements you might not have considered, and allows you to see your idea from a different perspective. At the same time, however, I also perceived its limitations. The lesson generated by AI, while impeccable, isn't born from direct experience and therefore can't take into account children's small vulnerabilities, their rhythms, their emotions, and the group dynamics that only those who experience the classroom can understand. It's as if AI constructs an ideal lesson, while the teacher constructs a possible, concrete, tailor-made lesson. Another aspect I noticed is that AI tends to use very formal and academic language, while my version was more spontaneous, closer to my way of thinking and speaking. This made me reflect on the fact that instructional design isn't just a technical exercise; there's always a bit of the person who writes it within a lesson. In conclusion, I believe AI is a valuable tool, especially for those like me who are learning and have many doubts and insecurities. It can help structure, verify, and complete. But it can't replace the teacher's sensitivity, their ability to observe, listen, and adapt. The best design comes when these two worlds meet: the order of AI and the teacher's humanity. AI can shape, but we are the ones who give meaning! We can never be replaced.*

Table 2. Illustrative examples of prompts and student reflections

To assess coding reliability, an independent double-coding procedure was conducted on a subset of 30 prompts (27% of the total corpus) using a coding rubric comprising six indicators. Inter-rater agreement was evaluated using the percentage of observed agreement, Cohen's

kappa coefficient (κ), and Gwet's AC1 coefficient. The results indicated substantial agreement for the *Contextual Completeness dimension* (κ = 0.673; AC1 = 0.713) and moderate agreement for *Professional Vocabulary* (κ = 0.549; AC1 = 0.534), Attention to Inclusion (κ = 0.490; AC1 = 0.511), and *Prompt Structure* (κ = 0.467; AC1 = 0.502). The Request Specificity dimension showed lower agreement (κ = 0.193; AC1 = 0.229), suggesting that this category involved a greater degree of interpretative judgment during coding. For the Iteration variable, Cohen's kappa could not be computed because both raters assigned all prompts to the same category (1 = single exchange). The complete results are presented in Table 3. The findings for the Iteration dimension indicate that participants still tend to conceive interaction with Artificial Intelligence as a one-time stimulus–response exchange rather than as a dialogic and iterative process of progressively refining prompts. This finding suggests that AI interaction competence—understood as the ability to exploit successive cycles of feedback and prompt revision to improve the quality of AI-generated outputs—has not yet been fully developed. Further research is therefore needed to examine the extent to which this competency evolves through experience and targeted educational interventions, as well as to identify the factors that promote a more informed and effective use of iterative interaction with AI systems.

Prompt Evaluation Grid Indicator	Observed agreement (%)	Cohen's k	Gwet's AC1	Interpretation
Contextual completeness	**80,0**	**0,673**	**0,713**	Substantial agreement
Specificity of the request	**46,7**	**0,193**	**0,229**	Weak agreement (fair)
Professional lexicon	**76,7**	**0,549**	**0,534**	Moderate agreement
Attention to inclusion	**66,7**	**0,490**	**0,511**	Moderate agreement
Prompt structure	**73,3**	**0,467**	**0,502**	Moderate agreement

Table 3. Inter-Rater Reliability of the Coding Rubric (N = 30)

With regard to the reliability of the coding of the 30 participants' critical reflections (representing 27% of the total corpus), an independent double-coding procedure was conducted. The following tables present the frequencies assigned by the two raters for each of the four levels of Kember et al.'s coding framework [18] (Table 4), the 4 × 4 agreement matrix between the two raters (Table 5), and the percentage of agreement for each level (Table 6). In addition, the overall inter-rater reliability for the four-level ordinal scale was assessed using the weighted Cohen's kappa coefficient (quadratic weighting).

Level of reflection	Evaluator 1	Evaluator 2
1. Descriptive writing	0	0
2. Understanding	2	0
3. Reflection	22	25
4. Critical reflection	6	5
Total	**30**	**30**

Table 4. Frequencies attributed by the two evaluators to the reflections

Evaluator 1/ Evaluator 2	1	2	3	4	Total
1	0	0	0	0	0
2	0	0	2	0	2
3	0	0	21	1	22
4	0	0	2	4	6
Total	**0**	**0**	**25**	**5**	**30**

Table 5. Concordance matrix regarding reflections

Level	Agreements	Cases involved	% agreement
1. Descriptive writing	0	0	—
2. Understanding	0	2	0,0%
3. Reflection	21	27	77,8%
4. Critical reflection	4	7	57,1%

Table 6. Percentage of agreement by level

Inter-rater reliability was estimated using the weighted Cohen's kappa coefficient (quadratic weighting), which is appropriate for four-level ordinal variables, and Gwet's AC1 coefficient. The weighted Cohen's kappa value was $\kappa = 0.571$, indicating a moderate level of agreement between the two raters [23]. In contrast, Gwet's AC1 = 0.798 indicated substantial inter-rater agreement according to the classification proposed by Landis and Koch.

Prompt variable	Spearman ρ	p	Kendall τ-b	p
Contextual completeness	**0,176**	**0,045**	**0,164**	**0,045**
Specificity of the request	0,157	0,074	0,149	0,075
Professional lexicon	**0,235**	**0,007**	**0,227**	**0,008**
Attention to inclusion	0,141	0,111	0,129	0,114
Prompt structure	0,085	0,334	0,084	0,332
Total prompt score	**0,264**	**0,002**	**0,225**	**0,003**

Table 7. Nonparametric correlations between prompt quality and level of critical reflection

Prompt variable	Spearman ρ	p	Holm-adjusted p	Kendall τ-b	Significant after Holm correction
Contextual completeness	0.176	0.045	0.180	0.164	No
Specificity of the request	0.157	0.074	0.222	0.149	No
Professional lexicon	0.235	0.007	0.035	0.227	Yes
Attention to inclusion	0.141	0.111	0.222	0.129	No
Prompt structure	0.085	0.334	0.334	0.084	No
Total prompt score	**0.264**	**0.002**	n.a.	0.225	Not included in the correction

Table 8. Nonparametric correlations between prompt quality and level of critical reflection, with Holm-adjusted p-values for the individual indicators

The analysis revealed a positive and statistically significant correlation between the overall prompt score and the level of critical reflection (Spearman's ρ = 0.264, p = 0.002; Kendall's τ-b = 0.225, p = 0.003). After the Holm–Bonferroni correction, Professional Lexicon was the only individual indicator significantly associated with reflection (adjusted p = 0.035). Contextual Completeness did not remain statistically significant (adjusted p = 0.180) and is therefore considered an exploratory indicator. The Kruskal–Wallis test further confirmed significant differences in the overall prompt score across the four levels of reflection ($H(3) = 9.52$, $p = 0.023$; $\varepsilon^2 = 0.052$) (Table 8). The analysis revealed a positive and statistically significant correlation

between the overall prompt score and the level of critical reflection (Spearman's ρ = 0.264, p = 0.002; Kendall's τ-b = 0.225, p = 0.003). Although the strength of the association was modest, students who produced higher-quality prompts tended to demonstrate deeper levels of critical reflection. Among the individual prompt indicators, Contextual Completeness showed a weak but statistically significant positive correlation with the level of critical reflection (ρ = 0.176, p = 0.045), while Professional Vocabulary was the indicator most strongly associated with reflection depth (ρ = 0.235, p = 0.007). In contrast, Request Specificity, Attention to Inclusion, and Prompt Structure did not reach statistical significance. The findings for the Request Specificity indicator suggest that its limited predictive value is consistent with the relatively low level of coding reliability observed for this dimension. The Kruskal–Wallis test further confirmed that the overall prompt score differed significantly across the four levels of critical reflection (H(3) = 9.52, p = 0.023). However, the effect size was small (ε^2 = 0.052), indicating that the level of critical reflection accounted for a limited, albeit non-negligible, proportion of the variance in prompt quality. Overall, these findings are consistent with the hypothesis that greater prompt design competence—particularly demonstrated through the use of professional terminology and more comprehensive contextualization—is associated with deeper critical reflection on the use of Artificial Intelligence in instructional design (Table 9).

Statistics	Value
H	9,518
gl	3
p	0,023
ε^2	0,052

Table 9. Kruskal–Wallis Test - Comparison of the total prompt score across the four levels of critical reflection

5. Conclusion and discussion

The present study examined the quality of prompts produced by students enrolled in the Primary Teacher Education degree program and the level of critical reflection they developed after using Artificial Intelligence (AI) to design a lesson. The aim was to explore the relationship between prompt engineering competencies and reflective capacities in educational contexts. The findings indicate, first, that most participants were able to construct prompts of intermediate quality, whereas only a limited number achieved the highest scores according to the evaluation rubric. This distribution suggests that, although students had acquired the fundamental skills required to interact with generative AI systems, the ability to formulate comprehensive, specific, and methodologically well-structured prompts remains a competency that is still developing. With regard to critical reflection, the reflection level, according to Kember et al.'s [18] classification, was the most frequently observed. This finding indicates that participants were generally able

to relate their laboratory experience to their emerging professional knowledge, recognising both the strengths and the limitations of Artificial Intelligence in instructional design. The critical reflection level was less frequent but represented a substantial proportion of the sample. This higher level of reflection involves a deeper reconsideration of one's understanding of the role of AI within the teaching profession. These findings suggest that the proposed learning activity went beyond promoting a purely instrumental use of technology, instead fostering professional reflection and encouraging the development of a critical stance toward Artificial Intelligence. The reliability analyses showed that the coding rubrics used to evaluate prompts and reflective writings achieved generally satisfactory levels of inter-rater agreement. In particular, the assessment of Contextual Completeness demonstrated substantial agreement, whereas the remaining dimensions achieved moderate levels of agreement. The lower reliability observed for Request Specificity suggests that this construct involves a greater degree of interpretative judgment and may benefit from more detailed coding criteria in future studies. Similarly, the coding of critical reflections yielded moderate inter-rater agreement, a result consistent with the interpretative complexity involved in classifying reflective texts. The most significant finding of the study concerns the relationship between prompt quality and the depth of critical reflection. The non-parametric analyses revealed a positive and statistically significant association between the overall prompt score and the level of critical reflection, although the magnitude of this relationship was modest. These findings suggest that students who are capable of designing more comprehensive and better-structured prompts also tend to produce more sophisticated reflections on the educational use of AI. In particular, the use of appropriate professional vocabulary and the ability to contextualise requests were the prompt characteristics most strongly associated with higher levels of critical reflection. These results were further supported by the Kruskal–Wallis test, which revealed significant differences in prompt quality across the four levels of critical reflection. Overall, the findings are consistent with the hypothesis that prompt engineering should not be regarded merely as a technical skill, but rather as an integral component of teachers' professional competence, closely connected to processes of pedagogical analysis, instructional design, and reflective practice. In this regard, future studies adopting an experimental or longitudinal design would be valuable—for example, by providing training in prompting skills and subsequently measuring its effects on reflective practice—to further strengthen the evidence base in this area. Constructing effective prompts requires the ability to articulate instructional objectives, contextualise requests, select appropriate teaching methodologies, and employ professional educational terminology—competencies that reflect cognitive processes analogous to those involved in educational planning and curriculum design. This study nevertheless has several limitations. First, the sample consisted of students enrolled in a single teacher education program who participated in a single laboratory experience, thereby limiting the generalizability of the findings. Second, although statistically significant, the observed correlations were relatively modest, indicating that prompt quality represents only one of several factors influencing the ability to develop critical reflections on the educational use of Artificial Intelligence. Finally, prompt quality was assessed using an ad hoc evaluation rubric, the validity of which should be further established through application to larger samples and diverse educational contexts. Another potential limitation concerns the use of non-uniform rating scales across the study. Specifically, four indicators were rated on a three-point scale (1–3), whereas the remaining indicators were rated on a two-point scale (1–2). This asymmetry may have resulted in an implicit weighting of certain elements over others, potentially influencing the overall interpretation of the findings. Furthermore, it should be noted that the "reflection" category is highly concentrated (with approximately 77% of the sample

falling within this category). This limited variability may have reduced the variance available for correlation analysis, making it more difficult for a genuine relationship to emerge as statistically significant. In light of these findings, future research should investigate the development of prompt engineering competencies through longitudinal studies, evaluate the effectiveness of educational interventions specifically designed to foster critical reflection on the use of AI, and examine the role of individual variables—such as prior digital competence, technological self-efficacy, and pedagogical beliefs—in mediating the relationship between prompt quality and reflective processes. From this perspective, integrating Artificial Intelligence into initial teacher education should be viewed not only as an opportunity to enhance instructional design but also as a valuable context for fostering the reflective competencies required for the critical, informed, and pedagogically grounded use of emerging technologies.

ISSN 2991-9371

References

[1] N. Selwyn, "Should robots replace teachers? AI and the future of education", Cambridge (UK), Polity Press, 2019.

[2] W. Holmes, M. Bialik and C. Fadel, "Artificial intelligence in education: Promises and implications for teaching and learning", Boston (MA), Center for Curriculum Redesign, 2019.

[3] O. Zawacki-Richter, V.I. Marín, M. Bond and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?", *International Journal of Educational Technology in Higher Education*, Volume: 16 Issue: 39, 2019.

[4] P. Mishra and M.J. Koehler, "Technological pedagogical content knowledge: A framework for teacher knowledge", *Teachers College Record*, Volume: 108 Issue: 6, pg.1017–1054, 2006.

[5] D.T.K. Ng, J.K.L. Leung, S.K.W. Chu and M.S. Qiao, "Conceptualizing AI literacy: An exploratory review", *Computers and Education: Artificial Intelligence*, Volume: 2, 100041, 2021.

[6] I. Celik, "Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education", *Computers in Human Behavior*, Volume: 138, 107468, 2023.

[7] R. Vuorikari, S. Kluzer and Y. Punie, "DigComp 2.2: The Digital Competence Framework for Citizens", Luxembourg, Publications Office of the European Union, 2022.

[8] C. Redecker, "European framework for the digital competence of educators: DigCompEdu", Luxembourg, Publications Office of the European Union, 2017.

[9] I. Tenberga and L. Daniela, "Artificial intelligence literacy competencies for teachers through self-assessment tools", *Sustainability*, Volume: 16 Issue: 23, 10386, 2024.

[10] M. Kasinidou, S. Kleanthoys and J. Otterbacher, "Cypriot teachers' digital skills and attitudes towards AI", *Discover Education*, Volume: 4 Issue: 1, 1, 2025.

[11] W. Flores and A. Chiappe, "Integrating AI into Education: Preparation Factors and Teachers' Digital Competencies", *Revista colombiana de educación*, Volume: 97, e20825-e20825, 2025.

[12] J. D. Zamfirescu-Pereira, R.Y. Wong, B. Hartmann and Q. Yang, "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 (CHI '23)*, pg. 1–21, 2023.

[13] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, D.C. Schmidt, "A prompt pattern catalog to enhance prompt engineering with ChatGPT", *arXiv*, 2023.

[14] Ministero dell'Istruzione e del Merito, "Le Linee Guida per l'adozione dell'Intelligenza Artificiale nella scuola", 2025,

[15] https://www.mim.gov.it/documents/20182/0/MIM_Linee+guida+IA+nella+Scuola_09_08_2025-signed.pdf/b70fdc 45-4b75-1f7e-73bf-eab12989b928?t=1756468797694.

[16] Unione Europea, "REGOLAMENTO (UE) 2024/1689 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)", 2024, ELI: http://data.europa.eu/eli/reg/2024/1689/oj.

[17] K. Krippendorff, "Content analysis: An introduction to its methodology (4th ed.)", Thousand Oaks (CA), Sage, 2018.

[18] M. Sandelowski, C.I. Voils and G. Knafl, "On quantitizing", *Journal of Mixed Methods Research*, Volume: 3 Issue: 3, pg. 208–222, 2009.

[19] D. Kember, J. McKay, K. Sinclair and F.K.Y. Wong, "A four-category scheme for coding and assessing the level of reflection in written work", *Assessment & Evaluation in Higher Education*, Volume: 33 Issue: 4, pg. 369–379, 2008.

[20] N. Hatton and D. Smith, "Reflection in teacher education: Towards definition and implementation", *Teaching and Teacher Education,* Volume: 11 Issue: 1, pg. 33–49, 1995.

[21] P.A. Facione, "Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction", Millbrae (CA), California Academic Press, 1990.

[22] J. Cohen "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, Volume: 20 Issue: 1, pg. 37–46, 1960.

[23] K. Gwet "Handbook of Inter-Rater Reliability: How to Estimate the Level of Agreement Between Two or Multiple Raters". Gaithersburg (MD), STATAXIS Publishing Company, 2001.

[24] J. R. Landis and G.G Koch "The measurement of observer agreement for categorical data", *Biometrics*, Volume: 33 Issue: 1, pg. 159–174, 1977.

THE USE OF AI TOOLS

The authors used ChatGPT (OpenAI) and Claude (Anthropic) for language editing and translation of selected parts of the manuscript. All content was reviewed and verified by the authors.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Maria Rosaria Re wrote § 1.2, 2 and 3; Stefania Capogna wrote § 1 and 1.3; Alessandra Natalini wrote § 1.1, 4 and 5. All authors read and approved the final manuscript.