

Multidisciplinary Journal of Educational Research
Volume 16, Issue 2, 15th June 2026, Pages 112-128
Online First Published: 27th May 2026
© The Author(s) 2026

<http://dx.doi.org/10.17583/remie.19875>

Academic Performance, Originality Indicators, and Student Reflections in an AI-Mediated Service-Learning Context in Initial Teacher Education

José M. Rodríguez-Ferrer¹ 
<https://orcid.org/0000-0003-1811-0040>

Eduardo Miguel Blasco-Delgado¹ 
<https://orcid.org/0009-0005-7408-5582>

Ana Manzano-León² 
<https://orcid.org/0000-0001-6966-0355>

1) University of Burgos, Spain

2) University of Almería, Spain

Abstract

This study examines an AI-integrated service-learning experience in initial teacher education in which 161 undergraduate students designed educational games for real school cases. Artificial intelligence was explicitly integrated into the design process. A mixed-methods approach was adopted. Quantitative data were drawn from final project grades and Turnitin text-similarity reports; qualitative data comprised documentary analysis of students' written reflections on their AI use throughout the project. Results show high overall academic performance. Turnitin indicators revealed that most projects contained no AI-generated text, while a minority exceeded the threshold for project suspension under the assessment criteria. Thematic analysis of written reflections revealed that most students described AI primarily as a creative support tool and consistently framed their own judgment as central to pedagogical decision-making. A substantial proportion of reports connected AI use to the broader demands of inclusive design. However, a notable minority of groups showed discrepancies between their declared AI use and Turnitin indicators, and a small number of reports engaged with AI only superficially or not at all, despite explicit task requirements. These results contribute to the growing empirical literature on AI integration in teacher education and point to the need for sustained instructional scaffolding to develop critical AI literacy.

Keywords

AI literacy, service-learning, teacher education, academic performance, academic integrity

To cite this article: Rodríguez-Ferrer, J.M., Blasco-Delgado, E.M. & Manzano-León, A. (2026). Academic Performance, Originality Indicators, and Student Reflections in an AI-Mediated Service-Learning Context in Initial Teacher Education. *Multidisciplinary Journal of Educational Research*, 15(2), pp. 112-128. <http://dx.doi.org/10.17583/remie.19875>

Corresponding author: Ana Manzano León

Contact address: aml570@ual.es

Multidisciplinary Journal of Educational Research

Volumen 16, Número 2, 15 de junio de 2026, Páginas 112-128

Publicado en Online First: 27 de mayo de 2025

© Autor(s) 2026

<http://dx.doi.org/10.17583/remie.19875>

Rendimiento Académico, Indicadores de Originalidad y Reflexiones del Alumnado en un Contexto de Aprendizaje-Servicio Mediado por IA en la Formación Inicial Docente

José M. Rodríguez-Ferrer¹ 

<https://orcid.org/0000-0003-1811-0040>

Eduardo Miguel Blasco-Delgado¹ 

<https://orcid.org/0009-0005-7408-5582>

Ana Manzano-León² 

<https://orcid.org/0000-0001-6966-0355>

1) *University of Burgos, Spain*

2) *University of Almería, Spain*

Resumen

Este estudio examina una experiencia de aprendizaje-servicio integrada con inteligencia artificial en la formación inicial del profesorado, donde 161 estudiantes diseñaron juegos educativos para casos escolares reales. La IA se integró explícitamente en el proceso de diseño. Se adoptó un enfoque de métodos mixtos: los datos cuantitativos procedieron de las calificaciones finales y de los informes de Turnitin, mientras que los cualitativos consistieron en el análisis de las reflexiones escritas de los estudiantes. Los resultados muestran un alto rendimiento académico. La mayoría de los proyectos no contenían texto generado por IA, aunque una minoría superó el umbral de suspensión establecido. El análisis temático reveló que la mayoría de los estudiantes describieron la IA como herramienta de apoyo creativo y situaron su propio juicio como elemento central en la toma de decisiones, vinculando su uso con las demandas del diseño inclusivo. No obstante, una minoría mostró discrepancias entre el uso declarado de la IA y los indicadores de Turnitin, y algunos informes la abordaron de forma superficial. Estos resultados contribuyen a la literatura empírica sobre IA en la formación docente y señalan la necesidad de un andamiaje instruccional para desarrollar la alfabetización crítica en IA.

Palabras clave

Alfabetización en inteligencia artificial, aprendizaje-servicio, formación docente, rendimiento académico, integridad académica

Cómo citar este artículo: Rodríguez-Ferrer, J.M., Blasco-Delgado, E.M. & Manzano-León, A. (2026). Rendimiento Académico, Indicadores de Originalidad y Reflexiones del Alumnado en un Contexto de Aprendizaje-Servicio Mediado por IA en la Formación Inicial Docente. *Multidisciplinary Journal of Educational Research*, 15(2), pp. 112-128. <http://dx.doi.org/10.17583/remie.19875>

Correspondencia Autores(s): Ana Manzano León

Dirección de contacto: aml570@ual.es

The emergence of artificial intelligence (AI) is rapidly redefining teaching and learning processes in higher education. In just a few years, these technologies have gone from rarely used tools to accessible, widely used resources that play a role in key tasks such as content generation and lesson planning (Giannakos et al., 2025). In this process, educational innovation at the university needs profound reconfiguration of pedagogical models, teaching roles, and the critical use of these technologies (Alam & Mohanty, 2023).

In this context, one noteworthy aspect is the rapid adoption of AI by university students in their daily academic lives. Several recent studies show that the use of generative AI tools has become widespread. For example, Baig and Yadegaridehkordi (2024) find that most students use tools such as ChatGPT for academic tasks, including generating ideas, writing texts, and understanding complex concepts. Similarly, empirical studies indicate that students use AI as an active part of their academic work, integrating it into the processes of studying, writing assignments, and preparing for assessments (Baldrich et al., 2024; Strzelecki, 2023). This intensive use has direct implications for the nature of university learning. On the one hand, interaction with AI systems has been shown to facilitate more autonomous learning by providing immediate, level-tailored feedback. Along these lines, some studies indicate that the use of generative AI can improve conceptual understanding and knowledge organization, especially when used to clarify content or structure information (Chiu, 2024; Tlili et al., 2025).

However, the widespread use of AI in academic tasks raises fundamental questions about authorship, academic integrity, and the authenticity of learning (Bozkurt, 2024; Eke, 2023). In particular, the fact that generative systems can produce coherent texts without guaranteeing their veracity introduces the risk of confusing the formal quality of discourse with actual comprehension of content (Kasneci et al., 2023). Consequently, there is a risk of reduced cognitive agency, understood as the learner's capacity to direct, regulate, and take responsibility for their own thinking processes, and a less in-depth understanding of content, especially when AI is used as a substitute for thinking rather than as a support for it. In such cases, students may produce high-quality outputs mediated by AI without this reflecting genuine comprehension, critical reflection, or knowledge construction (Böttcher et al., 2025; Zhai et al., 2024).

These issues take on special relevance in the initial teacher training, where the use of AI affects both student learning and their future professional practice. In this context, AI should be understood as a learning object that requires the development of critical literacy. This literacy involves understanding how these technologies work, assessing their limitations, identifying biases, and making informed decisions about their use in educational contexts (Ayanwale et al., 2024). Moreover, critical AI literacy in teacher education entails understanding the technical dimensions of these tools and developing the capacity to reflect on their ethical, pedagogical, and social implications. These competencies are necessary for future educators who will face AI-integrated classrooms (Mienye & Swart, 2025).

At the same time, the integration of AI is linked to one of the structural challenges of contemporary education: Educational inclusion and attention to diversity. In increasingly heterogeneous educational contexts, the need to design accessible learning experiences has shifted the focus from individual adaptation to universal design for learning (Evmenova, 2018; Rodríguez-Ferrer et al., 2025). To promote educational inclusion, various teaching methodologies should be taught to future teachers at the university level, with technologies

integrated to support greater personalization and multimodality. For example, service-learning integrates academic learning with addressing real-world needs in community contexts (Manzano-León, 2025). The literature shows that this type of methodology promotes the development of complex competencies, such as decision-making, critical reflection, and professional responsibility, especially when combined with the design of educational resources geared toward real-world contexts.

The combination of service-learning and AI-mediated design in teacher education represents an emerging and underexplored intersection. While research has separately documented the benefits of service-learning for professional competence development (Rodríguez-Ferrer et al., 2023) and the potential of generative AI to support personalized and adaptive educational design (Zawacki-Richter et al., 2019), empirical evidence on their joint implementation in initial teacher training contexts remains scarce. This study aims to contribute to this gap. This study aims to analyze an innovative teaching experience in initial teacher training that integrates artificial intelligence into service-learning projects focused on designing inclusive educational resources. Specifically, the following research questions are posed:

RQ1. What level of academic performance do teacher education students achieve in AI-mediated service-learning projects, as assessed through their final submitted products?

RQ2. What patterns of artificial intelligence use are identified in student projects through the combined analysis of Turnitin text similarity reports and instructor review based on linguistic criteria?

RQ3. How do students assess the use of artificial intelligence in the process of designing inclusive educational resources within a service-learning project?

Methodology

This study adopts a descriptive mixed-methods approach, combining quantitative and qualitative methods to provide a comprehensive understanding of the educational experience under analysis (Creswell & Plano Clark, 2018).

Participants

The sample consisted of 161 second-year undergraduate students enrolled in the Primary Education degree program at a Spanish university. Participants were selected through non-probabilistic convenience sampling. The students participated in the Teaching Innovation Project "Aprender creando con IA: desarrollo de contenidos educativos por el alumnado de Ciencias de la Educación (25_26_1_24C)" at the University of Almería.

It is important to note that, while the unit of analysis for grades was 161 individual students, the text-similarity detection instruments and documentary analysis were applied to the 35 group projects produced by those students, organized in teams of 3 to 7 members. All members of each group received the same grade.

Instruments

Three types of instruments were used to collect information, selected to analyze both the students' academic performance and the quality, originality, and reflective depth of the products developed within the framework of the teaching innovation project.

- a) Final grades for students' completed projects are an indicator of academic performance. This indicator provided a quantitative measure of the degree to which learning outcomes were achieved.
- b) Documentary analysis of the sections of “Discussion and conclusions” included in the projects developed by the students. This instrument provided qualitative evidence of student learning and perceptions, complementing information obtained from grades.
- c) Anti-plagiarism reports of the projects. These reports were generated using a text similarity detection tool (Turnitin, LLC, 2024, current online platform version). They were used as indirect indicators of artificial intelligence use. However, it is important to clarify that Turnitin does not detect AI-generated content per se; rather, it relies on probabilistic language models that assign a score based on the likelihood that certain textual sequences were produced by a generative system (Cotton et al., 2024; Rudolph et al., 2023). Accordingly, the reported percentages should be interpreted as indicative signals and not as objective measures. To compensate for this limitation, the instructor conducted a secondary review when the Turnitin report exceeded 20%. This review was guided by linguistic criteria documented in scientific literature as characteristic of generative writing (Herbold et al., 2023; Kobak et al., 2024). Based on this combined review, an interpretive judgment was made regarding whether AI tools had likely been used in drafting the document.

Procedure

The project was developed during an academic semester as an educational intervention integrated into the Special Education subject of the Degree in Primary Education. The practice was structured around the design of educational resources, mediated by artificial intelligence, and delivered through a service-learning approach.

The students organized themselves into groups of between 3 and 7 people. The students themselves freely chose the composition of the teams. The role of the teaching staff was primarily to serve as a guide, providing continuous support and explaining various games and digital tools tailored to the specific needs of the cases under review.

The intervention was structured into several phases that progressively guided the learning process and project development.

First, the students received a series of real cases described by active teachers. These previously anonymized cases included relevant information about the students' characteristics, learning difficulties, strengths, educational context, and specific support needs. This material served as a starting point for the groups to identify learning barriers and guide the design of their educational proposals from an inclusive perspective.

Next, various board games were explained to support content related to therapeutic pedagogy, and artificial intelligence was explicitly introduced as a teaching tool, addressing

both its technical use and its pedagogical and ethical implications. After testing different games, the groups developed their educational projects autonomously and collaboratively. During this process, they designed resources adapted to their assigned cases, integrating theoretical knowledge about addressing diversity, principles of Universal Design for Learning (UDL), and the use of digital tools and AI. The teaching staff supported this process by offering feedback and specific guidance according to the needs of each group.

In the presentation phase, each group presented its project to the rest of the class, explaining the pedagogical decisions made, the design process, and the use of artificial intelligence. Students were intentionally asked to explain how they had used AI, at which points in the process, and for what purpose, thereby encouraging reflection on its educational and ethical use. Subsequently, the designed resources were given to the educational centers from which the cases came, reinforcing the service-learning dimension of the project and adding applied and social value to the experience.

Finally, each group submitted a written report, which included the theoretical foundation, a description of the educational resources developed, and a critical reflection on the process followed. Project grading followed the rubric below, which was developed based on Manzano-León (2025).

Table 1

Project Learning Rubric

Criterion	Excellent (4)	Good (3)	Pass (2)	Fail (1)
Theoretical grounding	Solid, up-to-date, and well-integrated theoretical framework. Rigorous use of academic sources.	Adequate framework, with some limitations in depth or integration.	Basic framework, with limited depth or use of sources.	Absence or significant weakness of the theoretical framework.
Educational game design	Highly coherent, innovative design adapted to the case. Adequately integrates objectives, mechanics, and educational needs.	Adequate and functional design with some pedagogical coherence.	Underdeveloped design or with limitations in its suitability	Incoherent, incomplete, or unadapted design.
Alignment with the case (SEN)	Excellent alignment with the specific needs of the case. Well-grounded inclusive proposal.	Correct alignment, though with limitations.	Superficial or poorly justified alignment.	Does not align with the case or contains conceptual errors.

Criterion	Excellent (4)	Good (3)	Pass (2)	Fail (1)
Critical reflection	Deep, critical, and well-argued reflection. Integrates theory and practice.	Adequate reflection, though with less critical depth.	Descriptive or limited reflection.	Absence of reflection or superficial character.
Originality and own elaboration	High degree of original elaboration with original, well-developed ideas.	Own elaboration predominates, with some external dependencies.	Limited elaboration or excessively dependent on external sources.	Minimal own elaboration or plagiarism.
Writing quality and structure	Clear, coherent, and academic writing. Excellent structure.	Adequate writing with minor errors.	Writing that could be improved and unclear structure.	Poor or disorganized writing.
Use of GenAI for game design	Appropriate, critical, and justified use of generative AI tools. How they were used is explicitly stated, evidencing decision-making, adaptation, and original elaboration. AI is used as support and does not replace the creative process.	Adequate use of AI, though with a lower level of justification or critical reflection. Use is described in general terms.	Limited or unclear use of AI. Little justification or integration into the process.	Inappropriate or unjustified use of AI. Excessive dependency or absence of explanation of its use.

Data Analysis

The data analysis was conducted using a mixed-methods approach, combining quantitative and qualitative techniques to provide a comprehensive understanding of the phenomenon under study.

First, a descriptive quantitative analysis of the final grades and the anti-plagiarism reports of the projects. Basic statistics (mean, standard deviation, minimum, and maximum) were calculated to characterize students' academic performance and the percentages of textual similarity, and to explore variability in the quality of the submitted work

Second, a qualitative documentary analysis (Bowen, 2009) of the written reports was conducted. This analysis followed a thematic approach structured in four phases: (a) exhaustive and repeated reading of the 35 written reports to ensure familiarization with the corpus; (b) initial inductive coding of relevant units of meaning, defined as text segments—sentences or brief paragraphs—that expressed a student perception, an assessment, or a reflective judgment about AI use; (c) grouping of codes into thematic categories based on their semantic similarity

and conceptual coherence; and (d) interpretive review and theoretical integration of the resulting themes.

The coding process was carried out independently by two members of the research team. Each researcher coded the full corpus separately, without prior coordination of labels, to minimize individual interpretive bias. Once the independent coding phases were complete, both researchers met to compare their categorization schemes. Disagreements were resolved through deliberative consensus: when a unit of meaning had been assigned to different categories, both researchers discussed the interpretive logic until they reached agreement. No unit was assigned to a category without consensus. Codes with equivalent meaning were merged. The final categorical structure stabilized when no new codes emerged after re-reading the corpus (theoretical saturation). Five categories were identified in the final structure: “AI as a creative support tool”; “complementary use subordinated to personal judgment”; “development of pedagogical competencies and inclusion”; “tensions between declared discourse and actual AI use”; and “absence or superficiality of reflection on AI”.

Ethical considerations

The study was conducted in accordance with the ethical principles applicable to research with human participants. At the start of the intervention, students were informed that their participation in the teaching innovation project could be subject to research analysis, and informed consent was obtained for the use of data derived from their projects, grades, and text similarity reports. Participation in the educational experience was integral to the course, whereas participation in the research component had no bearing on academic assessment, either positively or negatively. Cases provided by in-service teachers were previously anonymized before being handed to students, so that students could not identify either the name or the school associated with the case. The educational resources designed were transferred to the participating schools with the knowledge and authorization of their directors. The data used in the analysis were treated confidentially and do not allow individual identification of participants. The research was validated through the Teaching Innovation Project "Aprender creando con IA: desarrollo de contenidos educativos por el alumnado de Ciencias de la Educación (25_26_1_24C)" at the University of Almería.

Results

In relation to the first research question, a descriptive analysis of the final project ratings was conducted. The results show a high overall academic performance (8.36 out of 10), with a standard deviation of 1.92. In terms of dispersion, a high concentration of scores is observed at higher ranks, especially around 8.5 and 10 points. The highest grade is 10, while the lowest recorded value is 0 (not delivered), which corresponds to very few isolated cases. There is a small group rated 4 (suspended). This could indicate difficulties in the project's development or a lower level of involvement.

Regarding the second research question, the AI percentages from the Turnitin reports were recorded. The results show a high predominance of jobs without evidence of complete

generation through artificial intelligence. Specifically, 24 of the 35 projects (68.6%) had 0% of the writing attributed entirely to AI. This suggests a generally supervised, partial, or nonexistent use of these tools in drafting documents. On the other hand, 8 projects (22.9%) had estimated AI usage between 20% and 30%. This could be interpreted as a moderate use of these tools, probably linked to support tasks such as idea generation, initial drafting, or content reformulation. Finally, 3 projects (8.6%) presented values greater than 40%. This was considered an inappropriate or unjustified use of artificial intelligence, constituting grounds for suspension of the project under the established evaluation criteria.

Regarding the third research question, five thematic categories were identified through documentary analysis. Table 2 provides an overview of each category, including its prevalence in the corpus, the type of evidence on which it is based, and its relationship to the other categories.

Table 2

Qualitative Frequency Table (N = 35 Student Project Reports)

Category	N (%)	Representative excerpts from student reports	Descriptive pattern	Analytical inference
C1. AI as a creative support tool	31/35 88.6%	<i>"We mainly used AI during the brainstorming session, helping us filter and select an original idea."</i> <i>"Generating new ideas and different approaches to game design was where it helped most."</i> <i>"We used it to overcome the initial block and propose several creative directions before deciding."</i>	The most prevalent pattern in the corpus. Students describe AI use primarily in the early stages of the task: idea generation, initial structuring, and overcoming cognitive blocks. Language is process-oriented and treats AI as a facilitator of creativity within a broader workflow rather than as a final product generator.	This pattern is consistent with a scaffolding function in open-ended tasks. Whether the concentration of AI use at moments of uncertainty reflects productive support or reduced engagement with the generative challenge cannot be determined from reflective texts alone; process-level data (working drafts, interaction logs) would be required to address this question.
C2 Complementary use subordinated to personal judgment	28/35 80.0%	<i>"AI has been used as a complementary resource, never replacing human work or decision-making."</i> <i>"The proposals generated were</i>	Students consistently present AI as auxiliary and bounded by human agency. In this context, the risks associated with excessive	Whether this discourse faithfully describes actual practice or also reflects awareness of institutional expectations cannot be established. The

Category	N (%)	Representative excerpts from student reports	Descriptive pattern	Analytical inference
		<i>analyzed and selected by each group member."</i> <i>"Although it suggested ideas, all were reviewed, adapted, or discarded according to our criteria."</i>	dependency were known in advance: academic dishonesty, loss of authorship, and grade penalties. This constitutes a recurring discursive pattern: 28 of 35 reports contain at least one coded unit of this type. What the texts directly demonstrate is a shared declarative commitment to ethically conscious AI use.	cross-evidence in C4 warrants interpretive caution but does not permit a generalized reinterpretation of C2. The two categories should be read together.
C3 Development of pedagogical skills and inclusion	25/35 71.4%	<i>"Designing inclusive activities requires reflection, creativity, and sensitivity to student diversity."</i> <i>"AI helped us think about different formats, but it was us who had to understand the child's needs and decide what was appropriate."</i> <i>"Working on a real case made us realize how complex inclusive design is and how much professional knowledge it requires."</i>	Across 25 reports, students foreground the project's pedagogical challenge: understanding specific student profiles, applying UDL principles, and designing appropriate resources. AI appears as one instrument within a broader professional reasoning process. The service-learning context, grounded in real cases, is frequently cited as the element that gave the task its complexity and professional meaning.	This pattern is consistent with the idea that the pedagogical value of AI tools may depend substantially on the methodological framework in which they are embedded. Whether the service-learning context was the explanatory factor, or whether the same pattern would emerge in other task designs, cannot be determined without comparative data. The finding suggests that, in this context, students positioned AI as one tool within their pedagogical decision-making.

Category	N (%)	Representative excerpts from student reports	Descriptive pattern	Analytical inference
C4 Tensions between declared discourse and actual AI use	7/35 20.0%	<i>"We used AI only for inspiration in the creative process. The writing was done by the team members."</i> [Turnitin: +40% attributed to AI] <i>"AI has been used as a minor support tool."</i> [Turnitin: 20-30% attributed to AI]	In 7 reports, the self-reported AI use does not match the Turnitin data. Students describe limited or supplementary use of AI, while the anti-plagiarism reports attribute between 30% and 50% of the text to AI-generated content.	Two plausible, non-mutually exclusive interpretations may account for this pattern. First, students may have underestimated their use of AI in their projects. Another possible explanation is bias in Turnitin's AI detection tool. The available data do not allow these explanations to be disentangled.
C5 Absence or superficiality of reflection on AI	4/35 11.4%	<i>"This project has allowed me to learn much more about Down syndrome and how to design meaningful activities for these students."</i> <i>"The most valuable part was understanding how to adapt games to real educational needs."</i> <i>"We focused on making the resource accessible and motivating for the student in our case."</i>	In 4 reports, AI is entirely absent from the reflective content, despite being an explicit task requirement. Reflection focuses on academic content, the assigned case, or the service-learning experience. The absence of AI-related discourse is itself an observable and codeable feature of these texts.	The texts do not allow distinguishing between competing explanations: students may have used AI minimally and have little to report; they may have prioritized more substantive pedagogical content; or they may have interpreted the task in a way that did not foreground AI as a required object of reflection. Regardless of the explanation, the finding demonstrates that explicit task requirements are not sufficient to generate reflective engagement with AI as a pedagogical tool.

Note. N = number of reports containing at least one coded unit of meaning for that category. Categories are not mutually exclusive; a single report may contribute to more than one.

C1. AI as a Creative Support Tool

The most prevalent pattern in the corpus centers on the use of AI to facilitate generative and creative processes, particularly in the early stages of project development. Students consistently describe using AI when confronted with open-ended design challenges: generating initial ideas, exploring alternative approaches, structuring a proposal, or overcoming what several groups described as a "cognitive block" at the start of the task.

The results directly show that students position AI as a starting point rather than an endpoint, and describe subsequent stages (selection, design, adaptation, contextualization to the assigned case, and production) as student-directed. This approach is consistent with constructivist accounts of digitally mediated learning (Jonassen, 1999; Kim et al., 2018), in which external resources support engagement with complex and ill-defined problems.

C2. Complementary use Subordinated to Personal Judgment (n = 28; 80.0%)

The second most prevalent category reveals a recurring discursive positioning of AI as secondary and bounded by human agency. Across 28 reports, students construct a consistent narrative in which AI suggestions are described as subject to review, selection, and modification, and the locus of decision-making is explicitly situated within the student group.

What is directly observable is a shared declarative commitment to ethically conscious use (Chan, 2023; Perkins et al., 2024). What cannot be established from reflective texts alone is whether this account accurately describes the actual process or also reflects strategic awareness of how AI use should be presented in an assessed context. The cross-evidence in C4 introduces grounds for interpretive caution, but does not allow a generalized reinterpretation of C2.

C3. Development of Pedagogical Skills and Inclusion

A third category refers to the development of inclusive pedagogical competency. In this category, students foreground the complexity of the task: understanding the specific characteristics and needs of the students described in their assigned case, applying Universal Design for Learning principles, and designing resources that are both educationally sound and practically usable. AI appears in this context as one of several tools and sources of knowledge mobilized in the service of a broader professional goal.

What the texts show is that AI is not the object of reflection in the case studies, but rather an instrument within them: students articulate professional reasoning processes (why they chose certain materials, how they adapted them, what pedagogical principles guided design decisions). This is consistent with the idea that service-learning methodology may foster a more instrumental relationship with AI. However, whether the service-learning context was causally responsible for this pattern cannot be determined without comparative data from other contexts.

C4. Tensions Between Declared Discourse and Actual AI Use

In 7 of the 35 reports, the self-reported account of AI use does not match the corresponding Turnitin data. Students in these groups describe their AI use as limited, supplementary, or carefully controlled, while the anti-plagiarism reports attributed between 30% and 50% of the text to AI-generated content.

Two plausible interpretations may account for this pattern. First, students may have underestimated or underreported their use of artificial intelligence in developing their projects. This could be due to a lack of clear criteria regarding what constitutes AI use, particularly in cases where generated content is subsequently edited, adapted, or integrated into original work. In this sense, students' perceptions of their own practices may not fully align with the actual extent of AI involvement. A second possible explanation relates to the limitations of AI detection tools such as Turnitin. Current detection systems are known to have reliability issues, including false positives and difficulty identifying AI-generated content that has been modified or reworked. As a result, the similarity indicators may not accurately reflect the real degree of AI use. The available data do not allow these explanations to be disentangled, as no direct measures of students' interaction with AI tools were collected. This highlights the need for future research to incorporate process-based data that enables a more precise understanding of AI use in educational contexts.

C5. Absence or Superficiality of Reflection on AI

In 4 reports, AI is entirely absent from the reflective content or appears only as a brief mention, despite the task explicitly requiring that its use be addressed. Reflection in these reports focuses instead on curricular learning, the characteristics of the assigned case, or the value of the service-learning experience. This does not constitute a category of negative educational outcomes, but it does represent a structurally distinct response to the task.

The texts themselves do not reveal why AI was omitted. Possible explanations include minimal or nonexistent AI use by these groups, prioritization of what students considered more substantive pedagogical content, or an interpretation of the task that did not foreground AI as a required object of reflection. Regardless of the explanation, the finding demonstrates that explicit curricular requirements were insufficient in these groups to foster reflective, substantive engagement with AI as a pedagogical tool.

Discussion and Conclusions

This study describes the implementation of an AI-integrated service-learning project with 161 undergraduate students enrolled in a Primary Education degree program and examines three dimensions: final project grades, text-similarity indicators as proxies for AI use patterns, and students' written reflections on their experience with AI.

In relation to the first research question, the data show a high mean grade (8.36 out of 10) with a concentration of scores in the upper ranges, in this subject and assessment context. This outcome is descriptively consistent with the broader literature suggesting that AI-supported

tasks can be associated with high-quality final products (Kasneci et al., 2023; Dwivedi et al., 2023), and with evidence indicating that service-learning methodologies are associated with positive academic outcomes (Celio et al., 2011). However, the results cannot establish whether the observed performance reflects the contribution of AI tools, the service-learning methodology, prior student ability, group dynamics, the specific assessment instrument used, or their interaction. Furthermore, as the literature cautions, high grades on final products do not preclude the possibility that some students produced outputs of apparent quality without equivalent depth of understanding (Böttcher et al., 2025; Wu et al., 2025).

Regarding the second research question, the Turnitin reports show that many groups (68.6%) had 0% of text attributed to AI generation, while 22.9% fell in the 20–30% range and 8.6% exceeded 40%. As noted in the methodology section, Turnitin's AI detection relies on probabilistic language models and is subject to significant reliability limitations, including false positives and sensitivity to edited or reworked text (Cotton et al., 2024; Rudolph et al., 2023). The predominance of low similarity scores is descriptively notable and consistent with the finding in C2 that students framed their AI use as complementary rather than substitutive.

Regarding the third research question, the documentary analysis of students' written reflections yielded five thematic categories. The most prevalent pattern (C1, 88.6% of reports) describes AI as a support tool used primarily in early generative stages such as idea generation and overcoming initial cognitive blocks. The second most prevalent category (C2, 80.0%) reveals a discursive positioning of AI as subordinate to human judgment. Both patterns are consistent with research indicating that students tend to frame AI as a supplementary cognitive resource (Liu et al., 2024) and with expectations that, when made explicit in the assessment context, norms of ethical AI use are reflected in student discourse.

The category related to pedagogical competence development and inclusion (C3, 71.4%) shows that students positioned AI as one instrument within a broader professional reasoning process, with the service-learning context. This finding is consistent with research indicating that service-learning promotes the integration of academic knowledge with professional practice (Rodríguez-Ferrer et al., 2023) and that AI may support personalized educational design when embedded in applied contexts (Zawacki-Richter et al., 2019). However, the study cannot determine whether the service-learning context was the explanatory factor for this pattern. In this context, most students described AI as subordinate to their pedagogical decision-making rather than as the primary driver of their design process.

The data document how students described their AI use in these written texts, not whether those descriptions accurately represent their actual practices. The discrepancy category (C4, 20.0% of reports) is important in this regard: in seven groups, self-reported AI use did not align with Turnitin indicators, warranting interpretive caution across the corpus. The variability in depth of critical reflection across reports, ranging from substantive engagement to near-absence of AI-related content in the study case (C5, 11.4%), indicates that explicit task requirements were not sufficient in all cases to generate reflective engagement with AI as a pedagogical tool, a finding consistent with the argument that AI literacy requires deliberate instructional scaffolding beyond task design alone (Achab-Moukram et al., 2025; Chiu et al., 2024).

Limitations and Future Research

While this study provides relevant evidence on the integration of artificial intelligence into initial teacher education, several limitations must be acknowledged to adequately contextualize the findings. These can be grouped into two broad categories: limitations affecting the validity of the results, and limitations concerning the measurement instruments employed.

Regarding validity, the non-experimental design prevents establishing causal relationships between AI use and the observed outcomes. The association between AI-mediated projects and academic performance may reflect the influence of the service-learning methodology, students' prior characteristics, or their interaction, rather than AI integration per se. Compounding this, the sample was drawn from a single course at a single university through convenience sampling, which restricts the generalizability of the findings to other degree programs, institutions, or cultural settings.

Regarding measurement, the qualitative analysis relies exclusively on students' written reflections, a single textual genre collected at one point in the learning process. These texts may be shaped by social desirability bias, given that AI use was explicitly regulated and included in the assessment criteria, a concern reinforced by discrepancies between declared use and Turnitin-derived indicators. Turnitin itself constitutes an additional limitation: its scores reflect the statistical probability that certain sequences were generated by AI, not actual authorship or cognitive engagement, and should therefore be treated as heuristic signals rather than validated measures. Finally, final project grades were assessed by a single evaluator who was also the researcher, without moderation, and captured product quality rather than underlying learning processes, conceptual understanding, or capacity for transfer.

Taken together, these limitations point to a common agenda for future research: quasi-experimental or longitudinal designs with comparison groups; replication across diverse institutional contexts; richer data sources such as interviews and process logs; more transparent and validated protocols for detecting AI use; and independent assessment criteria with inter-rater reliability measures.

Conclusion

This study describes an AI-integrated service-learning experience in initial teacher education and examines project grades, text-similarity indicators, and students' written reflections across 35 groups. In this specific context, most students achieved high final grades; most projects showed low text-similarity scores attributed to AI; and most students described their use of AI as complementary and subordinate to their own judgment. The descriptive contribution of the study lies in the detailed documentation of how AI use manifested across multiple dimensions (performance outcomes, similarity indicators, and reflective discourse) within a service-learning methodology. The data show that AI use in this context was not uniform: it ranged from minimal or absent to extensive and potentially substitutive, and students' critical engagement with AI varied considerably across groups. This heterogeneity itself is a pedagogically meaningful finding, suggesting that explicit task requirements and institutional guidelines, while necessary, are insufficient to ensure reflective and ethical use of AI among

all students. These observations are consistent with arguments in the literature that AI literacy requires explicit and sustained pedagogical attention.

From a practical standpoint, this experience illustrates one way in which AI tools can be incorporated into service-learning projects in teacher education, with explicit use guidelines, assessment criteria that address both product quality and reflective process, and real-world cases that situate AI within professional decision-making. Whether this design produces deeper learning, stronger AI literacy, or better ethical practice than alternative designs remains an open question for future research. The present study offers a documented case that can serve as a reference point for such future work and contributes to the growing body of empirical descriptions of AI integration in teacher education contexts.

Funding

This research is funded by and linked to the Teaching Innovation Project "Aprender creando con IA: desarrollo de contenidos educativos por el alumnado de Ciencias de la Educación (25_26_1_24C)" of the University of Almería.

Author Contributions

AML: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. JMR: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. EMBD: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. All authors have read and agreed to the published version of the manuscript

References

- Achab-Moukram, Y., Rodríguez-Rivera, P., Aguilar-Parra, J. M., Rodríguez-Ferrer, J. M., Rodríguez-Moreno, J., & Manzano-León, A. (2025). Evaluating a UDL-Based Digital Escape Room for Bullying Awareness in Secondary Education. *Multimodal Technologies and Interaction*, 9(6), 55. <https://doi.org/10.3390/mti9060055>
- Alam, A., & Mohanty, A. (2023). Educational technology: Exploring the convergence of technology and pedagogy through mobility, interactivity, AI, and learning tools. *Cogent Engineering*, 10(2), 2283282. <https://doi.org/10.1080/23311916.2023.2283282>
- Ayanwale, M. A., Adelana, O. P., Molefi, R. R., Adeeko, O., & Ishola, A. M. (2024). Examining artificial intelligence literacy among pre-service teachers for future classrooms. *Computers and Education Open*, 6, 100179. <https://doi.org/10.1016/j.caeo.2024.100179>
- Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in the higher education: A systematic literature review and research challenges. *International Journal of Educational Research*, 127, 102411. <https://doi.org/10.1016/j.ijer.2024.102411>
- Baldrich, K., Dominguez-Oller, J. C., & García-Rocca, A. (2024). La Inteligencia Artificial y su impacto en la alfabetización académica: una revisión sistemática 1. *Educatio Siglo XXI*, 42(3), 53-74. <https://doi.org/10.6018/educatio.609591>
- Böttcher, A., Thurner, V., & Zönnchen, B. (2025, 22-25 April 2025). Concepts for Teaching Software Development in the Age of AI-Tools. 2025 IEEE Global Engineering Education Conference (EDUCON), <https://doi.org/10.1109/EDUCON62633.2025.11016505>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, 9(2), 27-40. <https://doi.org/10.3316/QRJ0902027>
- Bozkurt, A. (2024). GenAI et al.: Cocreation, Authorship, Ownership, Academic Ethics and Integrity in a Time of Generative AI. *Open Praxis*. <https://doi.org/10.55982/openpraxis.16.1.654>
- Celio, C. I., Durlak, J., & Dymnicki, A. (2011). A Meta-Analysis of the Impact of Service-Learning on Students. *Journal of Experiential Education*, 34(2), 164-181. <https://doi.org/10.1177/105382591103400205>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20, 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Chiu, T. K. F. (2024). The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. *Interactive Learning Environments*, 32(10), 6187-6203. <https://doi.org/10.1080/10494820.2023.2253861>
- Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, 6, 100171. <https://doi.org/10.1016/j.caeo.2024.100171>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

- Crompton, H., Burke, D., & Gregory, K. H. (2017). The use of mobile learning in PK-12 education: A systematic review. *Computers & Education*, 110, 51-63. <https://doi.org/10.1016/j.compedu.2017.03.013>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D.,... Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology*, 13, 100060. <https://doi.org/10.1016/j.jrt.2023.100060>
- Evmenova, A. (2018). Preparing teachers to use universal design for learning to support diverse learners. *Journal of Online Learning Research*, 4(2), 147-171. <https://doi.org/10.70725/806763ugjaeu>
- Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2025). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, 44(11), 2518-2544. <https://doi.org/10.1080/0144929X.2024.2394886>
- Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, 13, 18617. <https://doi.org/10.1038/s41598-023-45644-9>
- Jonassen, D. H. (1999). Designing Constructivist Learning Environments. In C. M. Reigeluth (Ed.), *Instructional Design Theories and Models: A New Paradigm of Instructional Theory* (Vol. II, pp. 215-239). Lawrence Erlbaum Associates.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., & Hüllermeier, E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kim, N. J., Belland, B. R., & Walker, A. E. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education: Bayesian meta-analysis. *Educational Psychology Review*, 30(2), 397-429. <https://doi.org/10.1007/s10648-017-9419-1>
- Kobak, D., González-Márquez, R., Horvát, E. Á., & Lause, J. (2024). Delving into ChatGPT usage in academic writing through excess vocabulary. *Science Advances*. <https://doi.org/10.1126/sciadv.adt3813>
- Liu, Y., Zhang, H., Jiang, M., Chen, J., & Wang, M. (2024). A systematic review of research on emotional artificial intelligence in English language education. *System*, 126, 103478. <https://doi.org/10.1016/j.system.2024.103478>
- Manzano-León, A. (2025). *Aprendizaje-Servicio lúdico: elaboración de juegos para la transformación social* (Vol. 213). Universidad de Almería. Recuperado de: https://editorial.ual.es/libro/aprendizaje-servicio-ludico-elaboracion-de-juegos-para-la-transformacion-social_166197/

- Mienye, I. D., & Swart, T. G. (2025). ChatGPT in education: A review of ethical challenges and approaches to enhancing transparency and privacy. *Procedia Computer Science*, 254, 181-190. <https://doi.org/10.1016/j.procs.2025.02.077>
- Perkins, M., Roe, J., Postma, D., McGaughan, J., & Olivier, J. (2024). Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. *Assessment & Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2024.2415654>
- Rodríguez-Ferrer, J. M., Manzano-León, A., & Aguilar-Parra, J. M. (2023). Game-Based Learning and Service-Learning to Teach Inclusive Education in Higher Education. *International Journal of Environmental Research and Public Health*, 20(4), 3285. <https://www.mdpi.com/1660-4601/20/4/3285> <https://doi.org/10.3390/ijerph20043285>
- Rodríguez-Ferrer, J. M., Manzano-León, A., García-Roca, A., & Soler-Delgado, R. (2025). Embark on the adventure! Advancing inclusive secondary education through narrative game learning for students with intellectual disabilities. *Journal of Intellectual Disabilities*, 29(4), 1033-1048. <https://doi.org/10.1177/17446295241302400>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of applied learning and teaching*, 6(1), 342-363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Strzelecki, A. (2023). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*, 32(9), 5142–5155. <https://doi.org/10.1080/10494820.2023.2209881>
- Tlili, A., Saqer, K., Salha, S., & Huang, R. (2025). Investigating the effect of artificial intelligence in education (AIEd) on learning achievement: A meta-analysis and research synthesis. *Information Development*, 41(3), 825-842. <https://doi.org/10.1177/026666669241304407>
- Wang, D., Liu, Y., Jing, X., Liu, Q., & Lu, Q. (2025). Catalyst for future education: An empirical study on the Impact of artificial intelligence generated content on college students' innovation ability and autonomous learning. *Education and Information Technologies*, 30(8), 9949-9968. <https://doi.org/10.1007/s10639-024-13209-6>
- Wu, H., Ni, H., Luo, W., & Wu, T. (2025). Development and validation of the AI dependence scale for Chinese undergraduates and a preliminary exploration. *Frontiers in Psychology*, 16, 1725393. <https://doi.org/10.3389/fpsyg.2025.1725393>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11(1), 28. <https://doi.org/10.1186/s40561-024-00316-7>