



OPEN ACCESS

EDITED BY

Antonino Ferraro,
Pegaso University, Italy

REVIEWED BY

Ingrid Kirschning,
University of the Americas Puebla, Mexico
Lucas Jasper Jacobsen,
University of Hamburg, Germany

*CORRESPONDENCE

Pilar Ester Mariño,so
✉ pester@ucjc.edu

RECEIVED 01 April 2025

ACCEPTED 02 September 2025

PUBLISHED 24 September 2025

CORRECTED 03 October 2025

CITATION

Ester Mariño,so P, Caballero García PA,
Morales Jareño I and Cañadas Rodríguez E
(2025) Evaluating the use of AI in the design of
learning situations by university students of
early childhood education.
Front. Psychol. 16:1604414.
doi: 10.3389/fpsyg.2025.1604414

COPYRIGHT

© 2025 Ester Mariño,so, Caballero García,
Morales Jareño and Cañadas Rodríguez. This
is an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Evaluating the use of AI in the design of learning situations by university students of early childhood education

Pilar Ester Mariño,so*, Presentación Angeles Caballero García,
Isabel Morales Jareño and Emilio Cañadas Rodríguez

Facultad de Educación, Universidad Camilo José Cela, Spain

The use of Artificial Intelligence (AI) in Higher Education (HE) is an expanding reality, exerting an increasingly significant influence on teaching and learning processes. Its integration into the university environment is reshaping how instructors design and deliver their courses while providing students with new opportunities to personalize their learning experiences. Not only does AI enable task automation and resource optimization, but it also presents methodological, technological, and ethical challenges that warrant thorough investigation. In this context, we assessed the academic performance of two hundred and thirty eight students enrolled in an undergraduate degree program for early years education teachers, defining performance as their ability to design learning situations aimed at promoting mathematical thinking in young children. Our analysis distinguished between those who used AI to complete the task and those who relied on traditional teaching resources. To this end, we adopted a sequential explanatory mixed-methods design. In the quantitative phase, we employed a quasi-experimental post-test only design with experimental and control groups. In the qualitative phase, we conducted an in-depth analysis of the learning materials produced by the students in the experimental group with the support of AI. The findings indicate that the application of AI brought about significant changes in the teaching-learning process. The experimental group obtained better academic results than the control group. These results underscore, on one hand, the transformative potential of AI to improve pedagogical practices, and on the other, the need for research to examine its long-term effects on student learning, teacher engagement and ethical use, and the development of HE curricula that include AI as a teaching resource.

KEYWORDS

artificial intelligence (AI), higher education (HE), mathematics education, academic performance, pedagogical innovation

1 Introduction

Artificial Intelligence (AI) is reshaping and redefining content creation by leveraging initial datasets to generate, optimize, and personalize educational materials more efficiently (Jiao, 2024; Navas-Franco et al., 2024). Not only does AI facilitate the automation of repetitive tasks and streamlining processes, but it also permits users to have more free time to focus on strategic and creative aspects of content development (Katiyar, 2024). However, AI also presents challenges that must be addressed from social, technological, scientific, and educational perspectives (Bit et al., 2024). In this respect, empirical evidence

underscores AI's significant impact on education, enhancing teaching-learning practices, and improving student academic performance.

2 Theoretical framework

It has been 70 years since John McCarthy coined the term *Artificial Intelligence* (Cristianini, 2016), with a huge step forward over the past 7 years (Chu et al., 2022); and it is not until 2017 that academic discourse on the use of AI in HE has gained traction. Nowadays, AI's scope, meaning, and application have expanded significantly across all educational levels, and its use helps to personalize learning, guide and support individual students considering their own preferences or personal characteristics (Hwang, 2014; Hwang et al., 2020).

As noted by Chan (2023), AI's accessibility and integration have permeated all educational domains and have also underscored the urgency for universities to develop AI-specific educational policies to equip students with the necessary skills to work with and understand this technology. Among those skills the mixture of ethical, technical, and human skills is necessary to prepare students to confidently face with the working future (Pervez et al., 2024; Qudrat-Ullah, 2025), and be aware of how to be responsible, effective and curious to inquire into innovative technologies that add value to any context. Moreover, universities bear the responsibility of preparing graduates to actively engage with AI-driven developments and emerging societal issues linked to their proliferation.

From a foundational perspective, AI is defined as the machines' capacity to imitate human intelligence (Turing, 1950), and over time, AI has evolved through different paradigms depending on its capabilities. Three primary AI waves can be identified: the first so-called symbolic AI operates based on predefined logic and can diagnose diseases; the second wave, known as data-driven AI, utilizes machine learning and data mining techniques; and the third one is referred to as context-aware AI, which is capable of understanding the real world and offering innovative solutions to novel or previously unknown problems (Holmes et al., 2019; Franganillo, 2023; Zhang et al., 2023). In any case, the current iteration of AI fosters uncertainty, particularly in education, where it aims to address complex objectives.

2.1 Evolution of AI in education

The trajectory of AI in education spans from the development of ELIZA in 1966 to the launch of ChatGPT in 2022 (Martins Ramos, 2023). While chatbots such as ALICE (1990's) improved response coherence without achieving contextual understanding, a significant transformation occurred with OpenAI's ChatGPT, a tool that takes advantage of deep neural networks and advanced machine learning models (Serrano and Moreno-García, 2024). The emergence of ChatGPT marks a turning point, and its potential long-term impact on education is projected to be substantial.

The present study focuses on generative AI, a subset of AI that has sparked considerable debate due to its increasing integration into educational environments (Franganillo, 2023). The

integration of generative AI into educational systems has gained global relevance, particularly following the COVID-19 pandemic (Nedungadi et al., 2024). The limitations of traditional education models during the pandemic highlighted the need for AI-driven applications that could support students at risk of falling behind in fundamental skills such as those of literacy and arithmetic capabilities (Goessmann et al., 2023; Pantelimon et al., 2021; Slimi, 2023).

In this respect, UNESCO (2019) established three primary intersections between AI and education: learning with AI (using AI tools in classrooms), learning about AI (understanding AI technologies and methodologies), and preparing for AI (enabling all citizens to comprehend AI's potential impact on human life). These dimensions underscored the importance of equipping educational communities with the necessary competencies to critically and productively engage with AI technologies.

AI and education have converged in the realm of values, judgements and politics (Selwyn, 2019), and due to the significant role technology plays now more than in previous eras, the current generation is technologically literate (Limna et al., 2022). It is also relevant to note that contemporary education faces the challenge of developing innovative teaching and learning practices that uphold the principles of quality, inclusion, and equity, aligning with the Sustainable Development Goal 4 (SDG4) of the 2030 Agenda (García-Martínez et al., 2023; Jiao, 2024).

2.2 Artificial intelligence (AI) in Higher Education (HE): opportunities and challenges

The impact of AI in education was unknown or even disappointing, due to the limited progress achieved (Bates et al., 2020; Serrano and Moreno-García, 2024), and as Zawacki-Richter et al. (2019) found out in their review, the low participation of education professionals shows it is necessary to focus on the pedagogical and ethic aspects in the application of AI in educational contexts. Accordingly, in the university sphere, AI presents both opportunities and concerns. A major issue is students' potential over-reliance on AI tools, which may undermine critical thinking and original writing skills (Chan, 2023; Civil, 2023; Warschauer et al., 2023). In the United States, a 2023 survey cited by Chan revealed that nearly one-third of students (out of 1,000 respondents) used AI chatbots like ChatGPT for written assignments, with 60% relying on AI for more than half of their tasks. While students were aware of the ethical implications, nearly 30% of faculty members were unaware of AI's prevalence in academic work.

This raises a dilemma for universities: they must either prohibit and penalize AI usage—potentially revising plagiarism policies (Alqahtani and Wafula, 2024; Chan, 2023; Hughes and Roblyer, 2023; Wood, 2023; Yau and Chan, 2023)—or integrate AI into curricula while promoting alternative learning strategies and faculty training. As Gallent-Torres et al. (2023) suggest, students frequently exposed to AI-related fraud and malpractice discussions may develop generalized behaviors that lead to unethical academic practices. Hence, it is crucial that teachers receive training in ethical

and privacy aspects related to the use of AI, ensuring responsible and respectful use of students' rights (Kasneci et al., 2023).

Nevertheless, AI's potential in HE extends beyond concerns over academic dishonesty. AI can enhance learning by personalizing assessments, identifying specific areas for improvement, and adapting content and activities to students' needs (González-González, 2023). AI-driven personalization aligns with the broader objective of providing high-quality, inclusive, and equitable education (Monika Singh et al., 2025).

In favor of the use of AI, a new step in the “normalization” of the AI process is being developed with the publication of the AI Assessment Scale (AIAS) (Perkins et al., 2024). This scale provides a new perspective to think of AI as a tool not to be hidden but rather used in the interest of transparency in the educational field (Caldeiro and Odetti, 2024). The use and philosophy of this tool clearly affects the role of the teacher and the students, and that transparency takes us necessarily to the evaluation of the AI itself as HE institutions around the world.

The emergence of AI tools in education presents a significant challenge for educational institutions, particularly universities: whether to restrict their use or to promote informed and ethical integration into teaching and learning processes. In line with authors such as Alqahtani and Wafula (2024), Chan (2023), and Hughes and Roblyer (2023), we argue that banning such technologies may hinder rather than support educational development. Instead, integrating AI into the classroom, supported by appropriate training for both students and educators, offers a more constructive approach. This also requires a re-evaluation of academic integrity policies, including those concerning plagiarism and originality. As highlighted by Kasneci et al. (2023), it is essential that educators receive specific training on ethical considerations and data privacy issues related to AI use, ensuring practices that are both responsible and respectful of students' rights.

Navarrete and Manzanilla (2023) argue that adaptive learning platforms use AI algorithms to assess student performance and adjust pedagogical resources promoting more effective and personalized learning. It is true that there is a paradigm of uncertainty regarding the use of AI, and the integration of these tools in education faces several barriers to their full acceptance in educational institutions (Macías et al., 2024). Likewise, Tomalá et al. (2023) state that AI has shown significant advances that allow its implementation in educational systems to improve teaching and learning. AI is also a potentially transformative tool capable of creating interactive learning environments that stimulate the active participation of students. Likewise, Cabrera (2024) highlights the potential these tools provide in transforming the context of assessment in education, due to the detailed analysis provided on students' performance. Using traditional methods, the identification of such patterns and trends could obviously go unnoticed, however using AI tools can aid to perform more accurate assessment and tailored to the competencies or skills of each student.

Building upon this broader potential of AI in education, lesson planning has emerged as a particularly promising area where such technologies can directly assist teachers. Lesson planning remains a cornerstone of effective teaching, functioning as a structured guide for achieving specific learning outcomes. A sound lesson

plan typically aligns instructional objectives, materials, classroom activities, and assessment strategies within broader curricular frameworks. Yet, despite its centrality, many educators—both in local and international contexts—face persistent challenges in producing coherent and pedagogically sound plans. These difficulties often include unclear objectives, unbalanced assessment loads, and ineffective instructional language. Such struggles are compounded by the inherent complexity of addressing diverse student needs, integrating culturally responsive pedagogy, and navigating high-stakes testing environments (Loh and Liew, 2016).

In this scenario, artificial intelligence emerges as a promising support tool. AI applications have shown potential to assist teachers in generating adaptable lesson content, fostering instructional creativity, and personalizing materials for varied learner profiles (Yeh, 2024; Octavio et al., 2024). However, the pedagogical effectiveness of these tools depends not merely on their availability but on the teacher's capacity to guide them through well-crafted prompts. As Karpouzis (2024) notes, prompt engineering has become a core skill—essential to translating AI's computational power into meaningful educational practices. Consequently, integrating AI into lesson planning is not a replacement for teacher expertise but rather an invitation to reconceptualize instructional design through a lens of human-AI collaboration, supported by targeted professional development.

2.3 Mathematics and AI in HE

AI has been used in various disciplines (Crompton and Burke, 2023) such as linguistics (Liang et al., 2021), engineering (Shukla et al., 2019), mathematics (Hwang and Tu, 2021), and medicine (Winkler-Schwartz et al., 2019), primarily for access to inclusive learning study materials, self-assessment, and personalized feedback (Hwang and Tu, 2021). However, practical research, like the one we present in this paper, is needed to delve into the changes that AI generates in the teaching-learning process and student performance in specific contexts.

In mathematics, several AI-driven platforms have been developed, such as Khan Academy's personalized learning platform, MathSpring's problem-solving tutorials, and ALEKS' adaptive assessment and learning tools. Of particular interest in this study are Matthew, an AI tool developed by Adaptical, and MagicSchool which, although not exclusively focused on mathematics, helps educators and educational institutions reduce administrative burdens, enabling more personalized instruction, and fostering individual student growth. These tools have been employed for the activity evaluated in our research, as they are applicable to all educational levels and subjects and allow for the generation of practical content and learning situations aligned with the new Spanish educational law (LOMLOE) (Vilar, 2024).

Personalization of learning is one of the most promising opportunities for AI in education, as through adaptive systems, the content and difficulty of tasks this tool can be adjusted according to the progress of the student, improving learning efficiency and increasing motivation (Bolaño-García and Duarte-Acosta, 2024).

Regarding the impact of AI on academic performance, many studies have identified success factors and challenges

in implementing AI in educational settings, particularly online environments. However, there is a lack of critical analysis concerning the direct influence of AI on academic performance (Adewale et al., 2024; Tiwari, 2023; Hashim et al., 2022; Seo et al., 2021; Zhou et al., 2021), and research further underscores AI's growing role in mathematics education (Hwang and Tu, 2021).

García-Martínez et al. (2023) note that assessment measures not only students' acquired skills but also internal and external factors influencing learning, such as aptitude and motivation. AI's ability to analyze large volumes of student performance data enables educators to identify trends and make data-driven pedagogical decisions (Mera Castillo, 2023; Serrano and Moreno-García, 2024). Furthermore, AI tools can provide immediate feedback, accelerating the learning process and allowing educators to focus on qualitative teaching aspects (Zhai, 2022; Monge-Vera et al., 2024).

Concerning instructional designs, Caldeiro and Odetti (2024) point out the importance of transparency in the use of generative AI. They further say that some educators find the AI as a great potential to improve the efficiency and quality of their educational proposals, whereas others have doubts about the authenticity of the content generated by AI. It is, therefore, essential to underscore the ethical and pedagogical value. In this respect, the AIAS Scale of Perkins et al. (2024) provides 5 levels for the use of AI in evaluation instances and the degree of intervention of AI in the educational design and evaluation of the process (Caldeiro and Odetti, 2024). According to this scale designed primarily for use in HE settings copes with the ethical challenges brought about using Gen AI to provide a practical solution in its adaptation to a Gen AI assessment approach. In this way, not only is students' engagement enhanced, but it also ensures its ethical usage and promotes skill development. In scale level 4, students are suggested to use Gen AI to create content as a collaborative and creative tool and compare students' results against AI's. This way learners develop critical thinking and evaluative skills by carrying out their own analysis of the use of Gen AI tools.

AI in education offers transformative potential as far as its implementation is guided by ethical considerations, responsible data usage, and a commitment to enhancing student learning experiences. Research is needed to explore how AI can be effectively integrated into HE while preserving academic integrity and critical thinking skills.

The main aim of this study was to integrate AI as an innovative pedagogical tool in university classrooms, encouraging students to utilize it for generating content related to didactic programming in early childhood mathematics education and evaluating its impact on academic performance compared to traditional methodologies.

3 Method

To achieve this objective, we used a sequential explanatory mixed-methods design (Creswell and Creswell, 2018), combining both quantitative and qualitative strategies to gain a comprehensive understanding of the effects of AI use in the planning of learning situations. In the quantitative phase, we employed a quasi-experimental post-test only design with experimental and control group. In the qualitative phase, an in-depth analysis was

TABLE 1 Distribution of participants by group and gender.

Experimental	<i>n</i>	%	Control	<i>n</i>	%
Woman	74	89	Woman	135	87
Men	9	11	Men	20	13
Total	83	100	Total	155	100

conducted on participants' responses regarding their experiences and outcomes when using AI for the design of didactic proposals. This analysis provided insights into their perceptions, evaluations, and potential improvements in the implementation of this technology within an educational context. Furthermore, it helped us to complete and explain the quantitative results in greater detail.

3.1 Participants

The study included two hundred and thirty eight university students enrolled in the undergraduate program in early childhood education. Out of the total number, 2two hundred and nine were female (87.82%) and twenty nine were male (12.18%). The distribution of participants into study groups was non-random. The sample selection was based on convenience, respecting the composition of intact classrooms to ensure that participants remained within their natural class groups. This strategy allowed for the analysis of the effects of implementing AI in the design of learning situations compared to a conventional method. Table 1 shows the distribution by group and gender.

The experimental group comprised 83 students (34.87%) who used AI tools for the design and adaptation of learning situations in a specific educational context. In contrast, the control group, consisting of one hundred and fifty five students (65.13%), carried out the same process using a traditional methodology without AI support. The primary difference between the two groups lies in the use of this technology as a resource for didactic planning.

3.2 Instruments

The instruments used to measure the study variables were as follows:

This *ad hoc* template was designed to assess students' initial competence in designing mathematics learning activities aimed at fostering mathematical thinking in early childhood education. While employed as an initial evaluation task, its structure and requirements were comparable to those of the post-test assignment, as both involved the development of an instructional plan focused on a specific numeracy concept. However, unlike the post-test—which required students to integrate all the content covered throughout the course, including both pre-numeracy and numeracy concepts—this initial activity focused exclusively on an introductory topic presented at the beginning of the program. The task was completed without the use of artificial intelligence tools and served as a comparative baseline for evaluating differences in product quality between the control and experimental groups. Both

the initial and post-test tasks were assessed using the same rubric to ensure consistency and comparability.

This template was also designed *ad hoc* to collect information regarding the title, educational stage, grade level, subject area, context, objectives, competencies, content, methodology, and assessment. It was provided to students to guide them in the creation of their mathematics teaching units, incorporating usage guidelines for both AI-based tools and traditional methodological resources. The development and validation of this template involved a panel of three experts in mathematics education and instructional design, who reviewed the instrument for clarity, relevance, and curricular alignment. Based on their feedback, a Content Validity Index (CVI) of 0.89 was obtained, indicating acceptable content validity. A pilot trial was also conducted with a small group of pre-service teachers ($n =$) not included in the main study sample, allowing refinement of language, structure, and clarity of the template before implementation.

Assessment rubric for instructors evaluating students' curricular design. This rubric was likewise developed *ad hoc* for the purposes of the study. The design of mathematical learning situations for early years pupils was assessed using a five-point Likert scale (1 = not appropriate, 5 = highly appropriate) across each component: formulation of objectives, competences, content, methodology, and assessment. It also recorded information on the resources used in the task. For students in the experimental group, the evaluation distinguished the level of alignment in the adaptation of AI-generated content according to four predefined categories (no adaptation required, adaptation required but not applied, poor adaptation, and adequate adaptation), based on subject-specific assessment criteria and current educational legislation. A codebook containing detailed operational definitions for each category was used during the evaluation process (see [Annexure 3](#)).

This coding scheme was developed deductively based on a general assessment rubric and iteratively refined through expert review to ensure conceptual clarity and relevance. Two independent raters underwent a five-hour training session using exemplar cases and a codebook to calibrate their interpretations of the four-level adaptation scale. They independently coded a subset of twenty five student responses across six rubric dimensions: Objectives, Competences, Content, Context, Resources, and Assessment. Inter-rater agreement was assessed using Cohen's kappa. The overall kappa coefficient was $\kappa = 0.831$, reflecting substantial reliability between raters ([Landis and Koch, 1977](#)). Item-specific kappa values also indicated strong consistency: Objectives ($\kappa = 0.728$), Competences ($\kappa = 0.713$), Content ($\kappa = 1.000$), Context ($\kappa = 0.839$), Resources ($\kappa = 0.729$), and Assessment ($\kappa = 0.946$). Discrepancies were resolved through a structured consensus process involving rubric review and collaborative interpretation of disagreements. This procedure led to minor refinements in the codebook, enhancing its clarity and applicability across the full dataset. The methodological approach adopted here contributed to the transparency, rigor, and reliability of the qualitative coding process.

Grade records. Provided data on students' academic performance, based on final marks, measured on a Likert-type scale from 0 to 10. These marks reflected all the competences

and learning outcomes that students were required to achieve in order to pass the module. These grades were based on a comprehensive rubric aligned with the official learning outcomes and competences defined in the curriculum and were assessed by qualified instructors using standardized criteria across both groups. This approach is supported by prior research indicating that course grades, when derived from structured and criterion-referenced assessments, can serve as valid proxies for academic achievement (e.g., [Adarkwah, 2021](#); [Allen, 2005](#); [Andersson, 2025](#); [York et al., 2015](#)). In our case, the grading system incorporated multiple components, including instructional design tasks, theoretical understanding, and practical application, thus reflecting a holistic evaluation of students' performance in relation to the module's objectives. Furthermore, the use of grades allowed for a consistent and ecologically valid comparison between the experimental and control groups, as it captured students' actual performance within the authentic learning environment. The integration of detailed rubrics, expert validation procedures, inter-rater reliability statistics, and triangulation with qualitative data (e.g., analysis of submitted tasks and adaptation strategies) helped mitigate potential bias and strengthened the methodological transparency and validity of our findings.

3.3 Procedure

In accordance with the ethical standards of the Declaration of Helsinki, students were informed of the objectives of the study, invited to participate voluntarily, assured of the anonymity and confidentiality of their responses, and their informed consent was obtained and recorded. Following this, participants were assigned to two groups: an experimental group and a control group. The experimental group used AI for the design of their didactic units, while the control group followed a traditional methodology without the use of AI.

In both groups, and at each stage (validation task/posttest), the task ([Annexure 1](#)) was conceived as a learning situation applicable to any course within early childhood education, which students were required to specify in their proposal. It was designed to be developed collaboratively in class, in pre-assigned groups, with the explicit purpose of preventing the control group from utilizing AI tools during the process. The activity aimed to integrate and apply the contents covered in the subject "Development of Mathematical Thinking," including early number concepts, verbal problem-solving, classification, and seriation activities. The task was implemented over several class sessions during the second semester of the previous academic year (2023-2024), within the regular schedule of the aforementioned course. Its primary objective was to assess students' general and specific competencies in designing learning situations that foster the development of mathematical thinking in early childhood education.

Students in the experimental group were allowed to work with one of three open-access and free AI tools (ChatGPT, Mathew, or MagicSchool) according to their preferences and needs. They were required to document the prompts used, to include the

AI-generated responses, and highlight in a contrasting color any modifications made, accompanied by a theoretical justification explaining the pedagogical relevance of the adapted content and the improvement of their final proposal. In contrast, students in the control group completed the same task without the integration of AI tools, working exclusively during scheduled class hours and under teacher supervision to ensure that AI was not used in completing the task, thereby preserving the integrity of the group comparison intended for the purposes of our research.

It is important to note that the overall pedagogical approach used throughout the course was consistent across both groups. The theoretical components were delivered in a shared format for all students, focusing on key principles of early mathematical thinking and instructional design. Practical learning was organized through collaborative projects and case studies that encouraged the application of theory to authentic classroom scenarios. The main difference between the groups was the integration of AI: while the control group was restricted to conventional tools and resources, the experimental group received explicit instructions on how to formulate and refine prompts to maximize the relevance and quality of AI-generated outputs.

This emphasis on prompt engineering allowed students to critically evaluate the pedagogical potential of different types of AI input, aligning generated content with learning objectives, and adapting it for classroom use. As such, the experimental design not only measured outcomes, but also aimed to develop students' reflective and metacognitive skills in working with emerging technologies.

Once the intervention was completed, data collection for both groups was conducted through the Blackboard Ultra platform over a period of 2 months.

The qualitative analysis of the planning content was carried out in several phases: first, an exploratory reading of the tasks submitted by the students was conducted; subsequently, emergent categories were generated from the content; and finally, the data were systematically coded.

The qualitative coding of content adaptations focused on evaluating how students used AI to generate content suitable for diverse classroom needs. Coding was conducted deductively, based on a four-level rubric ([Annexure 2](#)): Does not require adaptation, requires adaptation but not applied, Poor adaptation, and adequate adaptation. The coding logic was as follows:

- No adaptation required. It was assigned when the content generated by AI reached at least level 4 (Good) in all criteria of the general rubric without further modification.
- Adaptation required but not applied. Used when AI-generated content did not reach level 4 in one or more key criteria, and the student made no adaptations.
- Poor adaptation: Applied when the student attempted to adapt the AI-generated content but the resulting quality remained below level 4. Adaptations were minor, superficial, or lacked didactic relevance.
- Adequate adaptation: assigned when the student effectively modified the AI-generated content to reach at least level 4 in the rubric, ensuring quality, inclusivity, and curricular alignment.

This coding system enabled the linkage between pedagogical quality levels and the effective use (or lack thereof) of AI in instructional design, maintaining consistency between qualitative analysis and the predefined evaluative framework.

As previously explained, the qualitative analysis was conducted by two independent researchers to ensure the validity of the study and the consistency of data interpretation. The coding process was developed deductively from a general assessment rubric and refined through iterative expert review. Both coders received specific training using practical examples and a codebook, which allowed them to align their criteria before beginning the analysis. A subset of student submissions was coded independently, and discrepancies were resolved through a structured consensus process involving the review of rubric criteria and collaborative discussion until agreement was reached.

This methodology helped clarify and strengthen the coding system, ensuring its coherence and applicability across the entire dataset. It also contributed to a deeper understanding of students' performance and the pedagogical quality of the learning situations they designed. The final assessment of academic performance was based on the tasks submitted by students from both groups, allowing for a comparative analysis of the impact of AI use versus non-use on academic outcomes.

In order to illustrate the type of coding carried out, an excerpt of the students' work has been included in [Annexure 4](#), where the analysis can be observed in context. The bolded text in the excerpt highlights the specific adaptations made by the students.

3.4 Data analysis

We began the statistical treatment of the data with descriptive analyses that included frequencies, percentages, means, standard deviations, medians, and extreme values related to sociodemographic variables, AI tool usage, content adaptations, and academic performance.

To assess potential associations between AI use and the adaptation of curricular content, we conducted chi-square tests of independence that were conducted, along with residual analysis to delve deeper into the interpretation of the results and identify where significant differences were located. Cramer's *V* calculations helped to determine the strength of the association ([Cook and Campbell, 1979, 1986](#)).

Subsequently, we conducted inferential analyses using independent samples t-tests. In cases where statistically significant differences were observed, effect sizes were calculated using Cohen's *d* ([Cohen, 1988](#)) to evaluate the magnitude of those differences.

The statistical data processing was conducted using Jamovi software, version 1.6.23, with a significance level of 5% and a confidence interval of 95%.

4 Results

4.1 Initial diagnostic assessment: baseline comparison between groups

To begin the analysis, an initial assessment was conducted to determine whether the control and experimental groups were

TABLE 2 Shapiro–Wilk test of normality for initial competencies and final performance by group.

Group	Variable	<i>W</i>	<i>p</i>
Experimental	Initial competencies	0.978	0.174
Experimental	Final performance	0.977	0.152
Control	Initial competencies	0.99	0.424
Control	Final performance	0.986	0.129

TABLE 3 Comparison of initial competencies between control/experimental groups.

Group	<i>M</i>	<i>SD</i>	<i>t</i> (<i>gl.</i>)	<i>p</i>
Experimental	6.57	1.27		
Control	6.67	1.15	−0.58(150)	0.56

TABLE 4 Frequency distribution of AI types.

Levels	<i>fr</i>	% of Total	Cumulative %
Chatgpt	49	59.0 %	59.0 %
Mathew	29	34.9 %	94.0 %
Magic school	5	6.0 %	100.0 %

comparable in terms of their baseline competencies. Prior to conducting inferential analyses, the assumption of normality was tested using the Shapiro–Wilk test, confirming that the distribution of scores met the requirements for parametric testing in both groups (see Table 2).

At the beginning of the course, students were tasked with designing a classification activity as an initial diagnostic assessment, explicitly without the use of artificial intelligence. This activity was intended to evaluate their baseline competencies in developing early childhood education activities aimed at fostering mathematical thinking. As shown in Table 3, the evaluation of these tasks revealed no statistically significant differences between the control and experimental groups, $t(150) = -0.58$, $p = 0.56$, indicating that both groups started the course with a comparable level of competence in this area.

4.2 Use of artificial intelligence tools

Table 4 presents the distribution of frequencies and percentages of the different types of AI tools used by participants.

The results indicate that, out of all the AI tools used by the students for their designs of mathematics learning situations for elementary education, ChatGPT was the most frequently used AI tool, accounting for 59.0% of the total responses. This suggests that ChatGPT is the preferred AI tool among participants, likely due to its versatility and widespread adoption in educational and professional settings. Mathew was the second most frequently used AI tool, representing 34.9% of the total responses. Finally, Magic School was the least frequently used AI tool, with only 5.0% of participants selecting it.

4.3 Adaptation of content for each curricular component

Table 5 (below) shows the percentage distribution of adaptation levels across various educational components, including objectives, competences, context, knowledge, methodology, resources, and assessment. The adaptation levels are categorized as “No needed”, “Need but not do it”, “Bad adaptation”, “Good adaptation”, providing insight into the extent of modification and implementation across different areas.

The results indicate varying degrees of adaptation across different educational components. While objectives (45.8%), competences (45.8%), resources (45.8%), knowledge (44.6%), and methodology (44.6%) exhibit the highest levels of successful adaptation, context (63.4%), and assessment (60.2%) show the highest percentages of poor adaptation, highlighting significant challenges in these areas. A notable proportion of participants recognized the need for adaptation but did not implement it, particularly in knowledge (25.3%), methodology (25.3%), and assessment (20.5%), suggesting barriers to execution. Additionally, objectives (34.9%) and competences (33.7%) were most frequently considered as not requiring adaptation. The complete absence of responses in the “No needed” category for assessment (0.00%) suggests a universal recognition of its necessity. These findings emphasize the need for improved strategies and support, particularly in contextual adaptation and assessment, to ensure a more inclusive and effective learning environment.

Table 6 presents the chi-square (χ^2) test results, analyzing whether differences in adaptation levels across educational components are influenced by the type of AI tool used.

The results indicate that differences in adaptation levels are significantly influenced by the type of AI tool used for Objectives ($\chi^2 = 45.678$, $p < 0.001$, $\varepsilon^2 = 0.55704$) and Competences ($\chi^2 = 41.065$, $p < 0.001$, $\varepsilon^2 = 0.50079$), with large effect sizes. There is a significant and strong association between the use of AI and the adaptation of content for the curricular elements’ “objectives” and “competencies”, in favor of those who use AI. This suggests that the AI tool employed has a notable impact on how educators modify learning objectives and competences. However, for Context ($p = 0.896$), Knowledge ($p = 0.361$), Methodology ($p = 0.820$), Resources ($p = 0.637$), and Assessment ($p = 0.406$), the results are not statistically significant, implying that the choice of AI tool does not lead to meaningful differences in adaptation within these components. These findings highlight that while AI use plays a crucial role in shaping learning objectives and competences, other areas may require additional guidance or structured frameworks to ensure effective adaptation. Table 7 displays the standardized results corresponding to the curricular component ‘Objectives’.

Chi-square test revealed a significant association between the type of AI tool used and the level of objectives, $\chi^2(6, N = 83) = [\text{valor}]$, $p < 0.001$. The analysis of standardized residuals showed that Mathew was significantly overrepresented in the category “no adaptation needed” (residual = +6.70), while ChatGPT was significantly underrepresented in the same category (residual = −6.14). Furthermore, ChatGPT was strongly associated with the “good adaptation” category (residual = +5.18), whereas Mathew was significantly underrepresented in that same category (residual

TABLE 5 Adaptation levels in educational components: percentage distribution.

Levels	Objectives	Competences	Context	Knowledge	Methodology	Resources	Assessment
	% of total	% of total	% of total	% of total	% of total	% of total	% of total
No needed	34.9	33.7	1.2	1.2	1.2	10.8	0.00
Need but not do it		10.8	22.9	25.3	25.3	19.3	20.5
Bad adaptation	6.0	9.6	63.4	28.9	28.6	24.1	60.2
Good adaptation	45.8	45.8	12.0	44.6	44.6	45.8	19.3

TABLE 6 Chi-square test results: differences in adaptation levels based on AI tool used.

Curriculars items	χ^2	df	p	ε^2
Objectives	45.678	2	< 0.001*	0.55704
Competences	41.065	2	< 0.001*	0.50079
Context	0.220	2	0.896	0.00268
Knowledge	2.035	2	0.361	0.02482
Methodology	0.396	2	0.820	0.00483
Resources	0.901	2	0.637	0.01098
Assessment	1.801	2	0.406	0.02196

TABLE 7 Standardized residuals for the association between type of AI and adaptation of objectives.

Objectives adaptation level	ChatGPT (Residual)	Mathew (Residual)	Magic school (Residual)
No adaptation needed	−6.14	+6.70	−0.72
Needs but not do it	−0.32	+0.11	+0.46
Bad adaptation	+1.92	−1.69	−0.58
Good adaptation	+5.18	−5.67	+0.66

TABLE 8 Standardized residuals for the association between type of AI and adaptation of competencies.

Competency adaptation level	ChatGPT (Residual)	Mathew (Residual)	Magic School (Residual)
No adaptation needed	−5.44	+6.44	−1.65
Needs but not do it	−0.94	+0.64	+0.68
Bad adaptation	+1.72	−2.18	+0.81
Good adaptation	+4.74	−5.21	+0.66

= −5.67). These residuals indicate that ChatGPT and Mathew have different content adaptation patterns, while the School of Magic does not show significant differences in its adaptation levels. The use of ChatGPT was more frequently linked to effective adaptation of curricular content, while Mathew tended to be associated with lower levels of adaptation or lack of perceived need for adaptation. In Table 8, we can see the standardized results calculated for the curricular component ‘competencies’.

Chi-square test revealed a significant association between the type of AI tool used and the level of curricular competencies adaptation, $\chi^2(6, N = 83) = [\text{valor}]$, $p < 0.001$. The analysis of standardized residuals showed that Mathew was significantly overrepresented in the category “no adaptation needed” (residual = +6.44), while ChatGPT was significantly underrepresented in this same category (residual = −5.44). Furthermore, ChatGPT showed a strong positive association with the “good adaptation” category (residual = +4.74), whereas Mathew was notably underrepresented in that same category (residual = −5.21). These standardized residuals indicate, as in the previous case, that ChatGPT and Mathew have different competency adaptation patterns, while the School of Magic does not show significant differences in its adaptation levels. These results indicate that ChatGPT use was more consistently associated with a higher level of competency adaptation, while Mathew tended to cluster in categories requiring no adaptation or showing lower quality adaptation.

4.4 Academic performance and use of AI

Finally, Table 9 presents a comparison of performance scores between the experimental and control groups.

The results indicate a statistically significant difference in performance between the experimental and control groups ($t = 2.11$, $p = 0.020$). The experimental group achieved a higher mean score ($M = 8.10$, $SD = 1.07$) compared to the control group ($M = 7.75$, $SD = 0.75$) suggesting that the intervention applied in the experimental group positively influenced performance.

However, Cohen’s d (0.327) indicates that the effect of these differences is between small and moderate implying that, while the difference is statistically significant, its practical impact may be limited. These findings suggest that the implemented intervention had a positive, but modest, effect on performance, highlighting the need for further exploration of factors that could enhance its effectiveness.

5 Discussion and conclusions

The objective of this research was to examine how future teachers of early childhood education utilize AI to create learning situations for their students within the didactics of mathematics course. Specifically, we sought to identify which AI tools they employ, the levels of content adaptation they carry out, and which elements of early years mathematics planning involve the use of AI. Subsequently, an analysis was carried out to determine whether

TABLE 9 Performance differences by Group.

Group	Experimental		Control		<i>t</i>	<i>p</i>	<i>Cohen's d</i>
Performance	8.03	1.07	7.75	0.75	2.12	0.020	0.320

there were differences in the content adaptations made by the students depending on the type of AI tool used, as well as in the outcomes achieved, comparing those who employed artificial intelligence with those who relied on traditional teaching resources.

5.1 Uses of AI

Based on the results obtained, it was observed that students prefer to use tools such as ChatGPT rather than specific applications that generate lesson plans or learning situations in accordance with the national educational legislation. This trend may be linked to the popularity and growing use of ChatGPT in educational settings, where its adoption has increased exponentially among both students and educators (Hu, 2023). Factors such as ease of use and the adaptability of technological tools are crucial for their acceptance (Selwyn, 2019). This situation raises the need to reflect on the impact that the predominant use of general AI tools, such as ChatGPT, might have on the training of future teachers. It is particularly relevant in terms of their ability to design learning situations that meet legal and pedagogical standards. It also highlights the need to implement specific training programs to enable students to fully harness the potential of AI as a means to address this issue (Luckin et al., 2016).

To create learning situations, students must develop a series of components that structure and give meaning to them, such as objectives, competencies (based on legislation), content, methodology, resources, and evaluation. Additionally, they must adapt the material to the context provided, which includes student profiles requiring one or more adaptations. As observed, different AI tools offer varying levels of accuracy in providing objectives, competencies, and resources. However, they show less precision in technical aspects such as methodology and evaluation. Moreover, none of the submitted works included proper personalization of student learning for those with difficulties by adapting to the provided context.

This discrepancy may arise because the AI-generated tasks require technical and human knowledge derived from direct experience in implementing learning situations, which is difficult to replicate due to the specificity of the task. Another possibility is that students used overly generic prompts, limiting the AI's ability to develop the detailed specifications expected in such activities.

Nonetheless, students successfully adapted to AI-generated objectives, competencies, content, and resources. However, other components, such as methodology, evaluation, and context adaptation, were not redesigned according to the minimum requirements necessary for classroom implementation. As Martines Rizo (2013) points out, teachers' knowledge of evaluation tends to be limited, making professional development essential for bringing about change in their practices. This aspect has often been neglected, contributing to the persistence of ineffective assessment

practices. Proper training in evaluation techniques and approaches is crucial for improving educational quality and ensuring that assessments fulfill their formative and diagnostic purposes.

In the field of mathematics didactics, assessment is often based on the summative evaluation focused on the results obtained, sometimes overlooking the process or the strategies used by students (Boaler, 2016). This was evident in how students adapted AI-generated proposals, as many of the assessment tools, such as rubrics, were measured dichotomously, depending on whether the student achieved the goal or not.

Regarding context adaptation, an essential action to take for personalizing learning for students with diverse difficulties, it was carried out in a generic and imprecise manner, with simple modifications that did not fully address the needs of the proposed classroom. Angenscheidt Bidegain and Navarrete Antola (2017) argue that as education moves toward a more inclusive approach, less experienced teachers or those still in training face greater challenges in adapting their practices to the current context. These challenges can hinder the effective implementation of inclusive strategies, as they require in-depth knowledge and specific preparation that many pre-service teachers have not yet fully acquired. Consequently, these educators are more likely to encounter obstacles when attempting to integrate inclusive practices.

5.2 Levels of content adaptation by type of AI used

Our data shows significant differences in the way students adapted learning objectives and competencies based on the specific AI tool they employed. Notably, substantial variations emerged between ChatGPT, Mathew, and Magic School, highlighting the distinctive features of each tool in terms of design and functionality.

It is highly interesting to note that, when focusing on the adaptations made by students based on the AI tool used, significant differences emerged regarding the alignment of the objectives and competencies of the learning situations with the specificity of each tool. In particular, substantial variations were identified between ChatGPT, Mathew, and MagicSchool. Mathew, for example, excels in generating learning situations that are closely aligned with the curriculum and current educational legislation. Its design allows for the automatic creation of personalized content that precisely matches educational objectives and competencies. As a result, students who used Mathew found that the generated objectives and competencies were sufficiently aligned with curricular expectations, significantly reducing the need for additional adaptation. Conversely, students working with more general tools like ChatGPT or MagicSchool often needed to make manual adjustments to ensure compliance with curriculum standards, highlighting the differences in the level of alignment and the effort required by the user.

Mathew, for instance, is particularly noteworthy for his ability to generate learning scenarios that are strictly aligned with national curriculum standards and educational legislation. Its design allows it to produce customized content that precisely adheres to the objectives and competencies established in official regulations. Consequently, students using Mathew often find that the generated objectives and competencies already meet curricular expectations, significantly reducing the need for additional modifications. In other words, Mathew enables students to focus on other aspects of the instructional process, as the fundamental components related to objectives and competencies are pre-defined and aligned with official standards.

Another factor contributing to the reduced need for adaptation among students utilizing Mathew is the tool's high level of relevance in selecting objectives and content. Mathew appears to provide exceptionally accurate correspondence between proposed learning situations and educational objectives, leading students to perceive minimal necessity for further adjustments. This precise alignment between generated content and curricular needs stands as a key factor in explaining the observed differences in adaptation efforts compared to tools like ChatGPT and Magic School, where students often found it necessary to make adjustments to ensure alignment with educational objectives and competencies.

In light of the above, we can affirm that while AI tools may offer a valuable foundation for designing learning situations, their effectiveness not only depends on the precision of the tool itself but also on the user's ability to critically interpret and refine the generated results. Tools like Mathew demonstrate that highly specialized AI can reduce the workload associated with curricular alignment, allowing users to focus on methodological and contextual adaptations. However, more general tools require greater professional intervention. In this way, this study highlights the need for robust teacher training that will equip future educators with the analytical skills and pedagogical judgement necessary so that they can effectively use AI tools, compensate for their limitations, and ensure educational quality and compliance with regulations.

5.3 Adaptation to context, content, methodology, resources, and assessment

No significant differences were identified among the various AI tools regarding the adaptation of contextual elements, content, instructional methodology, resources, and assessment strategies. Regardless of the AI tool used, students applied similar adaptations in these aspects. However, it is essential to highlight that a lower degree of effectiveness was observed in the contextual adaptation and assessment of AI-generated content. This suggests that while AI tools provide useful proposals, students must apply their technical and pedagogical expertise to refine and adapt them effectively for a real-world implementation.

This process of customization requires specific competencies that students are still in the process of developing, limiting their ability to make fully effective adaptations. As a result, despite AI tools are valuable resources, their use demands a level of proficiency that students have not yet fully acquired. It, therefore, underscores the need for further refinements to make sure that

content is appropriately tailored to the educational context and assessment criteria.

Nonetheless, most students successfully adapted instructional methodologies and resource selection for the implementation of the activity. This success may be attributed to the comprehensive teacher training programs that integrate methodological aspects across various didactic approaches, fostering constructivist teaching practices. Such an approach is particularly relevant in early mathematics education. The collected data reinforces the notion that transforming teaching conceptions and practices is a gradual process fraught with challenges. Teacher education does not seek to replace students' initial ideas with those deemed superior by educational research but rather adopts progressive, constructivist approaches that encourage the evolutive development of teaching practices.

This philosophy aligns with the student-centered training model, which prioritizes the gradual and reflective development of competencies and knowledge. Consequently, teacher education is geared toward a transformation process that, although slow, promotes more sustainable and effective changes in educational practices, aligning with the demands of a more constructivist and adaptive teaching paradigm (Rivero et al., 2011).

5.4 Critical reflection and investigative learning approach in the use of AI

It is essential to emphasize that all student adaptations involved a rigorous process of critical reflection, regardless of the final outcome. This process required students to critically evaluate which content, resources, and methodologies were more suitable for the designed learning situation while identifying areas that required modifications for successful implementation. Consequently, the task was taken on a research nature, demanding continuous assessment and refinement of educational components, irrespective of the AI tool utilized.

This investigative approach fosters critical thinking, enabling students to question and select the most relevant and appropriate information for their specific contexts (Betancourth Zambrano, 2015). Furthermore, this process encourages an in-depth and deliberate analysis of pedagogical decisions, strengthening the critical thinking skills essential for effective teaching. It is crucial for training future educators to develop these competencies through reflection and inquiry, as it helps them adapt their practices to diverse learning needs while cultivating a more conscious and evidence-based approach to their professional responsibilities.

5.5 Impact of AI on academic performance

The findings of this study reveal statistically significant differences between students who used AI in the development of learning situations and those who did not. Students who incorporated AI tools into their analytical and adaptation processes demonstrated slightly higher academic performance compared to their peers who did not utilize such technologies, although the effect size (Cohen's $d \approx 0.33$) suggests a small to moderate impact, limiting practical implications.

The results align with existing literature suggesting that AI tool implementation can be a beneficial strategy to enhance academic outcomes. Essel et al. (2022), for instance, examined the impact of virtual assistants in higher education in Ghana, finding that students interacting with chatbots outperformed those in traditional instruction. However, given this study's posttest-only design, lack of random group assignment, and use of unvalidated rubrics, these results should be interpreted cautiously.

These findings point to AI's promise as an educational aid, reinforcing the need to explicitly link empirical outcomes with established theories. Students must be equipped with strong analytical skills to critically assess AI-generated content. Nevertheless, claims about AI's effectiveness should be tempered due to methodological constraints, and the evidence does not support broad generalizations about its overall impact on learning.

AI may assist with repetitive tasks and shift focus toward deeper learning processes. Yet, despite statistical significance, the effect's limited size and absence of a pretest reduce the strength of conclusions. Future research should adopt more rigorous designs and clearly connect results with theoretical frameworks to clarify AI's educational value.

The data show no major performance differences based on the specific AI tool used. All tools require users to apply interpretive and evaluative skills, indicating that these competencies—not tool type—are central to academic success. Still, results should not be overextended beyond what the evidence allows.

This study explored how pre-service early childhood and elementary teachers use AI to design learning scenarios in the undergraduate course of didactics of mathematics. Students favored general tools like ChatGPT over specialized applications aligned with national curricula. This suggests accessibility influences preferences, though it raises questions about pedagogical adequacy and legal compliance.

While AI supports drafting objectives and competencies, it lacks precision in methods, assessment, and adaptation for students with special needs. This contrasts with literature on AI's personalization potential, revealing that complex educational tasks still require expert human judgment. Limitations should be more explicitly addressed, ideally in a dedicated section.

AI can offer personalized learning opportunities (Mera Castillo, 2023; Zhai, 2022), yet its integration must respect pedagogical standards. This study reaffirms the indispensable role of teachers in guiding students, especially given the significant limitations of the current research design.

5.6 AI as an educational complement

These findings suggest that AI can serve as a valuable complement, particularly in routine or lower-complexity tasks, enabling students to focus on developing their analytical and adaptive skills. By leveraging AI to automate mechanical aspects of academic work, students can allocate greater cognitive resources to evaluating and refining content quality.

In conclusion, while AI tools may offer an initial framework for the design of learning activities, their effectiveness largely depends on the user's ability to critically interpret and adapt

the generated outputs. Tools such as Mathew exemplify how domain-specific AI can streamline certain instructional design processes, whereas more general tools often necessitate greater pedagogical mediation.

This study highlights the importance of equipping future educators with the analytical and pedagogical competencies required to engage meaningfully with AI technologies, harness their potential benefits, and uphold educational standards. Accordingly, the findings should be interpreted as exploratory and serve as a foundation for further empirical research in this area.

Finally, future directions may include expanding the implementation of this approach to additional courses, conducting ongoing evaluations of emerging learning modalities, and fostering students' critical capacities, particularly in content adaptation and personalization, as key factors in optimizing the role of AI in competency-based higher education.

5.7 Limitations and future directions

Despite the valuable findings obtained, this study presents several limitations that should be considered to guide future research. Firstly, the use of final course grades as the sole measure of academic performance limits the objectivity and generalizability of the results, as these grades, although based on detailed rubrics, do not constitute a standardized or independent indicator of learning. To overcome this limitation, it is recommended that future studies incorporate externally validated assessment instruments that allow for a more precise and comparative measurement of the impact of AI on student performance.

Secondly, the methodological design based on a post-test-only scheme without randomized group allocation, although justified by logistical and ethical considerations, restricts the ability to establish strong causal relationships. Future research should consider more rigorous experimental designs, such as randomized controlled trials, to enhance internal validity.

Additionally, the sample was selected for convenience, which limits the representativeness of the findings; therefore, it would be desirable to expand the sample to include different university contexts and disciplines to improve generalizability.

Another important limitation is that the learning situations designed were not implemented in real early childhood education classrooms, which prevents the evaluation of their applicability and effectiveness in authentic settings. Future studies should include implementation and observation phases in real-world environments to validate the pedagogical relevance of AI-generated proposals.

Moreover, students demonstrated limited ability to adequately adapt AI-generated content in key areas such as assessment, methodology, and contextual personalization, highlighting the need to strengthen teacher training in analytical, ethical, and pedagogical competencies for the critical use of these technologies.

Finally, although triangulation and peer review strategies were employed to strengthen the qualitative coding process, the absence of a quantitative measurement of formal double coding may have affected the reliability of the analysis. Consequently, it is recommended that future studies incorporate

more robust inter-coder validation procedures. Addressing these limitations will enable progress toward a deeper, more rigorous, and contextually grounded understanding of the role of AI in teacher education and its impact on educational quality.

Then, as future directions, it is proposed to overcome the limitations of this exploratory study through practical actions such as extending this experience to other courses, the continuous evaluation of these new forms of learning, and the strengthening of students' critical skills, particularly in the adaptation and personalization of content. These areas are considered key to maximizing the positive impact of AI on university-level, competency-based learning.

One of the main limitations of this study lies in the use of final course grades as the primary measure of academic performance. Although these grades were based on detailed rubrics aligned with the official learning outcomes and competences of the module, and were applied consistently by qualified teaching staff, we acknowledge that they do not constitute an independent or standardized measure of student achievement.

This choice was made due to logistical constraints and the need to assess performance within an authentic learning environment. Nevertheless, future research would benefit from the inclusion of externally validated instruments that allow for a more objective and generalizable evaluation of the impact of pedagogical interventions, particularly in contexts involving the integration of emerging technologies such as artificial intelligence.

Another limitation of the study is that the designed learning situations were not implemented in actual early childhood education classrooms. As a result, it is not possible to determine whether the adaptations made by the pre-service teachers, based on AI-generated materials, effectively respond to the real needs and dynamics of a preschool classroom. Furthermore, without classroom implementation, it remains unclear whether these proposals meet the legal and curricular requirements established for early childhood education in practice.

These limitations, while relevant, were addressed through practical safeguards to preserve the internal consistency and pedagogical coherence of the study.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for this specific study because it is part of a broader research project that has already been reviewed and approved by the Ethics Committee. All participants gave informed consent to take part in the larger project, which includes this sub-study. Given the minimal risk and the educational context of this particular analysis, no additional ethical review was deemed necessary. The study was conducted according to the guidelines of the Declaration of Helsinki. The

studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

PE: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing. PC-G: Formal analysis, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing. IM: Supervision, Validation, Visualization, Writing – review & editing. EC: Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research project, with the acronym DUAPIAT, was funded by the 11th Research Call of Camilo José Cela University, Madrid, Spain.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Correction note

30 September 2025 This article has been corrected with minor changes. These changes do not impact the scientific content of the article.

03 October 2025 This article has been corrected with minor changes. These changes do not impact the scientific content of the article.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Adarkwah, M. A. (2021). The power of assessment feedback in teaching and learning: a narrative review and synthesis of the literature. *SN Social Sciences*, 1(75). doi: 10.1007/s43545-021-00086-w
- Adeyale, M. D., Azeta, A., Abayomi-Alli, A., and Sambo-Magaji, A. (2024). Impact of artificial intelligence adoption on students' academic performance in open and distance learning: a systematic literature review. *Heliyon* 10.e40025. doi: 10.1016/j.heliyon.2024.e40025
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *Clear. House J. Educ. Strateg. Issues Ideas* 78, 218–223. doi: 10.3200/TCHS.78.5.218-223
- Alqahtani, N., and Wafula, Z. (2024). Artificial intelligence Integration: pedagogical strategies and policies at leading universities. *Innov. High. Educ.* 1:20. doi: 10.1007/s10755-024-09749-x
- Andersson, A. (2025). *Reassessing Academic Desert: Why Grades Can Be Deserved*. *Philosophia*. Available online at: <https://link.springer.com/article/10.1007/s11406-025-00879-w> (Accessed February, 14, 2025).
- Angenscheidt Bidegain, L., and Navarrete Antola, I. (2017). Actitudes de los docentes acerca de la educación inclusiva. *Cienc. Psicol.* 11, 233–243. doi: 10.22235/cp.v11i2.1500
- Bates, T., Cobo, C., Mariño, O., and Wheeler, S. (2020). Can artificial intelligence transform higher education? *Int. J. Educ. Technol. High. Educ.* 17:42. doi: 10.1186/s41239-020-00218-x
- Betancourth Zambrano, S. (2015). Desarrollo del pensamiento crítico en docentes universitarios. Una mirada cualitativa. *Rev. Virtual Univ. Catól. Norte* 44, 238–252.
- Bit, D., Das, K., Biswas, S., and Nag, M. (2024). The impact of artificial intelligence on the educational system. *Int. J. Sci. Res. Sci. Technol.* 10, 424–432. doi: 10.32628/IJSRST2411424
- Boaler, J. (2016). *Mathematical Mindsets: Unleashing Students' Potential Through Creative Math, Inspiring Messages, And Innovative Teaching*. San Francisco, CA: Jossey-Bass.
- Bolaño-García, M., and Duarte-Acosta, N. (2024). Una revisión sistemática del uso de la inteligencia artificial en la educación. *Rev. Colomb. Cirugía* 39, 51–63. doi: 10.30944/20117582.2365
- Cabrera, K. (2024). Transformando la Educación Básica: retos y perspectivas de la inteligencia artificial. *Vitalia* 5, 1–17. doi: 10.61368/r.s.d.h.v5i2.11
- Caldeiro, G., and Odetti, V. (2024). La transparencia como factor crítico en los usos educativos de la IA Virtualidad. *Educación y Ciencia. PsyAxiv preprint*. 28, 46–53.
- Chan, C. K. Y. (2023). A comprehensive ai policy education framework for university teaching and learning. *Comput. Sci. Comput. Soc.* 20:38. doi: 10.1186/s41239-023-00408-3
- Chu, H., Tu, Y., and Yang, K. (2022). Roles and research trends of artificial intelligence in higher education: a systematic review of the top 50 most-cited articles. *Austr. J. Educ. Technol.* 38, 22–42. doi: 10.14742/ajet.7526.
- Civil, B. (2023). ChatGPT can hinder students' critical thinking skills: artificial intelligence is changing how students learn to write. *Queen's J.* Available online at: <https://www.queensjournal.ca/story/2023-03-16/opinions/chatgpt-can-hinder-studentscritical-~thinking-skills/>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences 2nd Edn.*. Hillsdale, NJ: Lawrence Erlbaum.
- Cook, T. D., and Campbell, D. T. (1979). *Quasi-Experimentation. Design and Analysis Issues for Field Settings*. Boston, MA: Rand McNally.
- Cook, T. D., and Campbell, D. T. (1986). The causal assumptions of quasi-experimental practice. *Synthese* 28, 141–180.
- Creswell, J. W., and Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches 5th Edn.* Thousand Oaks, CA: SAGE Publications.
- Cristianini, N. (2016). Intelligence reinvented. *New Sci.* 232, 37–41. doi: 10.1016/S0262-4079(16)31992-3
- Crompton, H., and Burke, D. (2023). Artificial intelligence in higher education: the state of the field. *Int. J. Educ. Technol. High Educ.*, 20:22. doi: 10.1186/s41239-023-00392-8
- Essel, H., Vlachopoulos, D., Tachie-Menson, A., Johnson, E., and Baah, P. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. *Int. J. Educ. Technol. High. Educ.*, 19, 1–19. doi: 10.1186/s41239-022-00362-6
- Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *Methaodos. Rev.cienc. Sociales* 11:m231102a10. doi: 10.17502/mrcs.v11i2.710
- Gallent-Torres, C., Zapata-González, A., and Ortego-Hernando, J. L. (2023). El impacto de la inteligencia artificial generativa en educación superior: una mirada desde la ética y la integridad académica. *RELIEVE* 29:9134. doi: 10.30827/relieve.v29i2.29134
- García-Martínez, I., Fernández-Batanero, J. M., Fernández-Cerero, J., and León, S. P. (2023). Analysing the impact of artificial intelligence and computational sciences on student performance: systematic review and meta-analysis. *J. New Approach. Educ. Res.* 12, 171–197. doi: 10.7821/naer.2023.1.1240
- Goessmann, C., Idele, P., Jauer, K., Loinig, M., Melamed, C., and Zak, T. (2023). *Pulse of Progress: Mapping the State of Global SDG Data in 2023*. United Nations; ETH Zurich. doi: 10.3929/ethz-b-000630718
- González-González, C. (2023). El impacto de la Inteligencia Artificial en la Educación: Transformación de la forma de enseñar y aprender. *Rev. Curric.* 36, 51–60. doi: 10.25145/j.qurricul.2023.36.03
- Hashim, S., Omar, M. K., Ab Jalil, H., and Sharef, N. (2022). Trends on technologies and artificial intelligence in education for personalized learning: systematic literature review. *Int. J. Acad. Res. Progress. Educ. Dev.* 11, 884–903. doi: 10.6007/IJARPED/v11-i1/12230
- Holmes, W., Bialik, M., and Fadel, C. (2019). *Artificial intelligence in education: promises and implications for teaching and learning*. Boston, MA: Center for Curriculum Redesign.
- Hu, K. (2023). *ChatGPT Sets Record for Fastest-Growing User Base, Analyst Note*. Available online at: <https://www.reuters.com/technology/chatgpt-sets-record-fastestgrowing-user-base-analyst-note-2023-02-01/> (Accessed February 3, 2023).
- Hughes, J. E., and Roblyer, M. D. (2023). *Teaching with technology: Creating student-centered classrooms*. Boston, MA: Pearson.
- Hwang, G.-J., Xie, H., Wah, B. W., and Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Comput. Educ. Art. Intell.* 1:100001 doi: 10.1016/j.caeai.2020.100001
- Hwang, G. J. (2014). Definition, framework and research issues of smart learning environments - a context-aware ubiquitous learning perspective. *Smart Learn. Environ.* 1:4. doi: 10.1186/s40561-014-0004-5
- Hwang, G. J., and Tu, Y. F. (2021). Roles and research trends of artificial intelligence in mathematics education: a bibliometric mapping analysis and systematic review. *Mathematics* 9:584. doi: 10.3390/math9060584
- Jiao, D. (2024). AI-Driven Personalization in higher education: enhancing learning outcomes through adaptive technologies. *Adult High. Educ.* 6, 42–46. doi: 10.23977/aduhe.2024.060607
- Karpouzis, K. (2024). Plato's shadows in the digital cave: controlling cultural bias in generative AI. *Electronics* 13:1457. doi: 10.3390/electronics13081457
- Kasneji, E., Sessler, K., Betschart, P., and Kasneji, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Ind. Diff.* 103:102274. doi: 10.1016/j.lindif.2023.102274
- Katiyar, N. (2024). AI-driven personalized learning systems: enhancing educational effectiveness. *Educ. Admin. Theor. Pract.* 30, 11514–11524. doi: 10.53555/kuey.v30i5.4961
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2025.1604414/full#supplementary-material>

- Liang, J. C., Hwang, G. J., Chen, M. R. A., and Darmawansah, D. (2021). Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach. *Interact. Learn. Environ.* 31:8348. doi: 10.1080/10494820.2021.1958348
- Limna, P., Somporch, J., Sutitthep, S., Pichart, K., and Patcharavadee, S. (2022). A review of Artificial Intelligence (AI) in education during the digital era. *Adv. Knowl. Execut.* 1, 1–9. Available online at: <https://ssrn.com/abstract=4160798>.
- Loh, C. E., and Liew, W. M. (2016). Voices from the ground: the emotional labour of English teachers' work. *Teach. Teach. Educ.* 55, 267–278. doi: 10.1016/j.tate.2016.01.016
- Luckin, R., Holmes, W., Griffiths, M., and Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. London: Pearson.
- Macías, T., Plaza, L., Mina, W., and Chévez, Á. (2024). *Inteligencia artificial. Usos en la educación básica*. 593 Digital Publisher 9, 1167–1178.
- Martines Rizo, F. (2013). Dificultades para implementar la evaluación formativa: Revisión de literatura. *Perfiles educ. [online]* 35, 128–150. doi: 10.22201/iisue.24486167e.2013.139.35716
- Martins Ramos, A. S. (2023). Inteligência Artificial Generativa baseada em grandes modelos de linguagem - ferramentas de uso na pesquisa acadêmica. In *SciELO Preprints*. doi: 10.1590/SciELOPreprints.6105
- Mera Castillo, D. E. (2023). La influencia de la inteligencia artificial en la personalización del aprendizaje: perspectivas y desafíos en la educación. *Rev. Ingenio Global* 2, 28–39. doi: 10.62943/rig.v2n2.2023.64
- Monge-Vera, M. M., Villamagua, G. M., Aroca, C. E., Chico, B. A., and López, J. E. (2024). Personalización del proceso de aprendizaje mediante inteligencia artificial. *LATAM Rev. L. Am. Cienc. Soc. Human.* 5, 772–785. doi: 10.56712/latam.v5i3.2076
- Monika Singh, T., Kishor Kumar Reddy, C., Ramana Murthy, B. V., Nag, A., and Doss, S. (2025). AI and education: bridging the gap to personalized, efficient, and accessible learning," in *Internet of Behavior-Based Computational Intelligence for Smart Education Systems*, eds. M. Ouassia, M. Ouassia, H. Lamaazi, M. El Hamlaoui, and K. Reddy (Hershey, PA: IGI Global Scientific Publishing), 131–160. doi: 10.4018/979-8-3693-8151-9.ch005
- Navarrete, Z., and Manzanilla, H. (2023). Una perspectiva sobre la inteligencia artificial en la educación. *Perfiles Educ.* 15, 134–151. doi: 10.22201/iisue.24486167e.2023.Especial.61693
- Navas-Franco, L., Ortiz-Carrasco, W., Cabrera-Urbina, E., and Orna-Quintanilla, K. (2024). La efectividad de los materiales educativos en la personalización del aprendizaje. 593 Digital Publisher CEIT 9, 805–817. doi: 10.33386/593dp.2024.5.2688
- Nedungadi, P., Tang, K.-Y., and Raman, R. (2024). The transformative power of generative artificial intelligence for achieving the sustainable development goal of quality education. *Sustainability* 16:9779. doi: 10.3390/su16229779
- Octavio, M. M., Argüello, M. V. G., and Pujolà, J. (2024). ChatGPT as an AI L2 teaching support: a case study of an EFL teacher. *Technol. Lang. Teach. Learn.* 6, 1–25. doi: 10.29140/tl.v6n1.1142
- Pantelimon, F. V., Georgescu, T. M., and Păunescu, C. (2021). The Evolution of AI-Driven educational systems during the COVID-19 pandemic. *J. Educ. Technol. Soc.* 24, 45–59. doi: 10.3390/su132313501
- Perkins, M., Furze, L., Roe, J., and MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): a framework for ethical integration of generative AI in educational assessment. *J. Univ. Teach. Learn. Pract.* 21:36. doi: 10.53761/q3azde36
- Pervez, N., Mahmood, W., Akram, M., and Waqas, M. (2024). Analyzing the alignment between university curriculum and job market requirements in pakistan: challenges and opportunities. *Indus J. Soc. Sci.* 2, 397–410. doi: 10.59075/ijss.v2i2.269
- Quadrat-Ullah, H. (2025). *Enhancing Interdisciplinary Learning. Empowering Educational Leaders Using Analytics, AI, and Systems Thinking*. Hershey, PA: IGI Global.
- Rivero, A., Azcárate, P., Porlán, R., Martín del Pozo, R., Harres, J., and Solís, E. (2011). La progresión de las concepciones de los futuros profesores de primaria sobre la metodología de enseñanza. *Educ. Foco* 14, 169–206. doi: 10.24934/ef.v14i18.236
- Selwyn, N. (2019). *Should Robots Replace Teachers? Ai Ana The Future Of Education*. Polity.
- Seo, K., Tang, J., Roll, I., Fels, S., and Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *Int. J. Educ. Technol. High Educ.* 18:54. doi: 10.1186/s41239-021-00292-9
- Serrano, J. L., and Moreno-García, J. (2024). Inteligencia artificial y personalización del aprendizaje: innovación educativa o promesas recicladas? *Educat. Rev. Electrón. Tecnol. Educ.* 89, 1–17. doi: 10.21556/edutec.2024.89.3577
- Shukla, A. K., Janmajaya, M., Abraham, A., and Muhuri, P. K. (2019). Engineering applications of artificial intelligence: a bibliometric analysis of 30 years (1988–2018). *Eng. Appl. Art. Intell.* 85, 517–532. doi: 10.1016/j.engappai.2019.06.010
- Slimi, Z. (2023). the impact of artificial intelligence on higher education: personalization and automation during the pandemic. *Comput. Educ.* 10, 17–33. doi: 10.19044/ejes.v10no1a24
- Tiwari, R. (2023). The integration of AI and machine learning in education and its potential to personalise and improve student learning experiences. *Ind. Sci. J. Res. Eng. Manag.* 7:7645. doi: 10.55041/IJSREM17645
- Tomalá, M., Mascaró, E., Carrasco, C., and Aroni, E. (2023). Incidencias de la inteligencia artificial en la educación. *RECIMUNDO* 7, 238–251. doi: 10.26820/recimundo/7.(2).jun.2023.238-251
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind* 49, 433–460. doi: 10.1093/mind/LIX.236.433
- UNESCO (2019). *The Sustainable Development Goals Report*. Available online at: <https://bit.ly/34nbq60> (Accessed December 12, 2019).
- Vilar, A. (2024). La IA y los Recursos Humanos: transformando el panorama laboral en 2024. *RRHH Digital. Tecnología e Innovación*. Available online at: <https://www.rrhddigital.com/secciones/tecnologia-e-innovacion/751573/la-ia-y-los-recursos-humanos-transformando-el-panorama-laboral-en-2024/>
- Warschauer, M., Waverly, T., Soobin, Y., Thomas, W., Sharin, J., Qian, D., et al. (2023). The affordances and contradictions of AI-generated text for writers of english as a second or foreign language (October 31, 2023). *J. Sec. Lang. Writ.* 62:101071. doi: 10.1016/j.jslw.2023.101071
- Winkler-Schwartz, A., Bissonnette, V., Mirchi, N., Ponnudurai, N., Yilmaz, R., Ledwos, N., et al. (2019). Artificial intelligence in medical education best practices using machine learning to assess surgical expertise in virtual reality simulation. *J. Surg. Educ.* 76, 1681–1690. doi: 10.1016/j.jsurg.2019.05.015
- Wood, P. (2023). *Oxford and Cambridge ban AI language tool GPT-3 over fears of plagiarism*. inews.co.uk. doi: <https://inews.co.uk/news/oxford-cambridge-ban-chatgpt-plagiarism-universities-2178391>
- Yau, C., and Chan, K. (2023). *University of Hong Kong temporarily bans students from using ChatGPT, other AI-based tools in coursework*. South China Morning Post. Available online at: <https://www.scmp.com/news/hong-kong/education/article/3210650/university-hong-kong-temporarily-bans-students-using-chatgpt-other-ai-based-tools-coursework> (Accessed February 17, 2023).
- Yeh, H. C. (2024). The synergy of generative AI and inquiry-based learning: transforming the landscape of english teaching and learning. *Interact. Learn. Environ.* 1–15. doi: 10.1080/10494820.2024.2335491
- York, T. T., Gibson, C., and Rankin, S. (2015). Defining and measuring academic success. *Practical Assessment, Research, and Evaluation* 20:5. doi: 10.7275/hz5x-tx03
- Zawacki-Richter, O., Marin, V. I., Bond, M., and Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education— where are the educators? *Int. J. Educ. Technol. High. Educ.* 16:710. doi: 10.1186/s41239-019-0171-0
- Zhai, X. (2022). ChatGPT User Experience: Implications for Education. *SSR [online]* doi: 10.2139/ssrn.4312418
- Zhang, B., Zhu, J., and Su, H. (2023). Toward the third-generation artificial intelligence. *Sci. China Inf. Sci.*, 66:121101. doi: 10.1007/s11432-021-3449-x
- Zhou, Q., Suraworachet, W., Pozdniakov, S., Martinez-Maldonado, R., Bartindale, T., Chen, P., et al. (2021). "Investigating students' experiences with collaboration analytics for remote group meetings," in *Artificial Intelligence in Education: 22nd International Conference, AIED 2021*, eds. I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, and E. Millán (Utrecht, The Netherlands, Springer), 133–145.