education sciences

MDPI

Article

Preservice Early Childhood Teachers' AI Literacy and Self-Efficacy in a Structured "Vibe Coding" Design Task: An Exploratory Pre–Post Study

Stamatios Papadakis [1,*] **and Georgios Zacharis** [2]

[1] Department of Preschool Education, Faculty of Education, University of Crete, Rethymnon, 74100 Crete, Greece
[2] School of Early Childhood Education, Faculty of Education, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; gzacharis@nured.auth.gr
* Correspondence: stpapadakis@uoc.gr

Abstract

This exploratory study examined a structured Vibe Coding task in preservice early childhood teacher education and its association with self-reported AI literacy, self-efficacy, attitudes toward AI, and technology acceptance. Vibe Coding was used in a qualified sense: participants created educational applications through natural-language prompting, then reviewed and revised the outputs. The task was implemented in two undergraduate courses using a convergent mixed-methods, single-group pre–post design without a comparison group. Quantitative analysis included 32 matched cases, while qualitative analysis drew on 26 reflective logbooks. All 32 participants also completed a researcher-designed, non-validated self-assessment rubric, reported descriptively. Paired-samples *t*-tests showed significant pre–post increases across all six self-report scales after Holm–Bonferroni correction. Effect sizes were large for AI literacy, self-efficacy, attitudes, and perceived ease of use, and moderate for perceived usefulness and behavioural intention. These changes indicate perceptions rather than demonstrated competence and cannot be attributed causally to the task. Directed qualitative analysis identified three predefined categories: pedagogical empowerment, prompting and technical confidence, and professional identity reconfiguration. Participants described refining prompts, identifying errors, and justifying revisions against pedagogical criteria. The findings provide exploratory evidence of how preservice teachers learned to evaluate and improve AI-generated educational outputs through a structured design task.

Keywords: Vibe Coding; AI literacy; preservice teacher education; generative AI; self-efficacy

Check for updates

Academic Editor: Diego Vergara

Received: 1 August 2026
Revised: 19 September 2026
Accepted: 21 September 2026
Published: 27 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

As generative AI becomes common in professional practice, teacher education faces a clear problem: how do we prepare preservice teachers, particularly those entering early childhood contexts, to use AI tools purposefully, and in ways that genuinely serve young children's learning and development? This is a pressing issue. While AI is not the first classroom technology, it is unique in how it reduces the technical barrier between a teacher's pedagogical intention and a fully functional digital output.

In February 2025, software engineer and AI researcher Andrej Karpathy coined the term 'Vibe Coding' to describe a mode of software development in which the programmer communicates intent in natural-language, leaves the code to the AI, and simply 'sees

stuff, says stuff, runs stuff' (Karpathy, 2025). The term spread beyond software development because, The term spread beyond software development because, by 2025, widely used generative AI tools such as Google Gemini (Gemini 2.5 Flash), OpenAI ChatGPT (GPT-5), and Anthropic Claude (Claude Sonnet 4.5) had changed how intent becomes output. Specific tools and their capabilities will continue to change; the argument here concerns the prompting-based design process rather than any particular system. For educators without programming backgrounds, this shift is particularly significant: the ability to design a functional, interactive educational application no longer requires knowledge of syntax, logic structures, or debugging protocols.

Data from a 2025 RAND Corporation survey of 1586 public pre-K teachers in the United States illustrate the scale of the gap: while 69% of high school teachers report using generative AI tools, only 29% of preschool teachers report any use, with systematic weekly or daily use reaching just 9% (Berne et al., 2025). The same study found that although 70% of preschool educators had received training in how to use educational technology, only 37% had been trained to evaluate its pedagogical quality. This gap between operational skill and judgement motivates this study. Preservice preschool teachers are also a population with historically low technology self-efficacy (Kalogiannakis & Papadakis, 2019), operating in a professional context defined by Developmentally Appropriate Practice (DAP), a framework that centres children's developmental needs, play-based learning, and the teacher's relational role over the use of digital tools (NAEYC, 2020). Vibe Coding and DAP address different priorities, namely technical production and child-centred pedagogy. We treat this tension as a design choice: the act of designing an AI-generated preschool application through structured, affectively enriched prompting may support integrated AI literacy development, pedagogical reasoning, and professional identity formation.

Research on Vibe Coding as a pedagogical intervention with preservice early childhood teachers is limited. The broader literature on generative AI in teacher education is growing rapidly but remains focused on text-based tasks such as lesson planning, feedback generation, and content creation (see, e.g., Chen, 2025; OECD, 2025). The design and evaluation of AI-generated interactive applications, and the prompt-based process through which such applications are created, remain under-reported in AI literacy development for preservice teacher programmes.

This study examines a structured Vibe Coding intervention embedded within two undergraduate courses offered at the University of [Crete] and the [Aristotle] University of [Thessaloniki], [Greece], during the winter semester of 2025–2026. We employed a convergent mixed-methods design. We examined (RQ1) whether participation in the intervention was associated with significant changes in preservice preschool teachers' AI literacy, self-efficacy, attitudes toward AI, perceived ease of use, perceived usefulness, and behavioural intention to use AI; and (RQ2) how participants described their experience of AI-assisted app design in terms of pedagogical agency, prompt learning, and professional identity.

2. Theoretical Background and Literature Review

2.1. AI Literacy in Teacher Education

AI literacy has emerged as a central construct in educational technology research over the past five years, driven by the rapid diffusion of AI tools in everyday and professional contexts. Long and Magerko (2020) define AI literacy as a set of competencies that enable individuals to evaluate AI technologies, communicate and collaborate effectively with AI, and use AI as a tool. D. T. K. Ng et al. (2021) offer a complementary four-dimensional framework, including Know and understand AI, Use AI, Evaluate and

create AI, and AI ethics, that has been widely adopted in empirical studies of AI literacy in educational populations.

Empirical research on AI literacy in teacher education has grown considerably since 2021, though most studies focus on in-service teachers or on higher education students broadly rather than on preservice early childhood educators specifically. A consistent finding is that self-reported AI literacy and self-efficacy for AI use are positively associated with positive attitudes toward AI and with intentions to integrate AI tools into teaching practice (D. T. K. Ng et al., 2021; OECD, 2025). However, studies also consistently show that AI literacy gains from single, short-duration interventions, lectures, demonstrations, or guided exploration, are modest in effect size. This study examines whether a task-based, iterative, production-oriented intervention is associated with effects of different magnitude.

A relevant theoretical development is the emergence of Media and AI Literacy (MAIL) as an integrated construct, as framed by the OECD (OECD, 2026). MAIL proposes that media literacy and understanding of algorithmic systems are no longer separable. The citizen, and by extension the educator, who is genuinely AI-literate must be able to interrogate both the content that AI systems produce and the logic by which they produce it. In the preschool educational context, this integrated literacy translates into the cultivation of what the OECD terms 'human agency': the capacity to position AI tools as instruments under continuous human oversight rather than autonomous authorities.

2.2. Technology Acceptance in Educational Contexts

The Technology Acceptance Model (TAM), originally proposed by Davis (1989), provides the theoretical framework for the quantitative component of this study. TAM proposes that Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) are the primary determinants of technology adoption behaviour, mediated by Attitude (ATT) and Behavioural Intention (BI). In educational research, TAM has been extensively applied to predict and explain teachers' and students' adoption of digital tools, including AI-based systems (Scherer et al., 2019; Teo, 2011). The model's parsimonious structure makes it appropriate for examining pre–post changes in technology-related beliefs following an educational intervention. TAM captures acceptance beliefs but not pedagogical agency or judgement. In this study, TAM frames the four technology acceptance constructs measured before and after the task, PU, PEOU, ATT, and BI, while AI literacy and self-efficacy are treated as additional constructs; the questions of agency, prompt learning, and professional identity that the intervention raises are addressed through the qualitative strand rather than through TAM.

In the context of AI tools, TAM has been adapted to include AI-specific constructs such as AI trust, AI anxiety, and AI literacy (Sindermann et al., 2021). The present study employs a TAM-based instrument adapted for generative AI tools in preschool educational design contexts, combining the core TAM constructs (PU, PEOU, ATT, BI) with AI Literacy and Self-Efficacy scales drawn from D. T. K. Ng et al. (2021) and Compeau and Higgins (1995).

2.3. Vibe Coding: Definition, Origins, and Educational Potential

In this study we use Vibe Coding in a deliberately narrow sense: creating runnable applications through natural-language prompts, without manually editing the underlying code. We note at the outset that the task studied here is not pure vibe coding. Participants reviewed outputs, identified errors, and revised against pedagogical criteria, which Willison (2025) places outside the term. We retain "Vibe Coding" for the prompting-and-generation core of the task and treat the surrounding evaluation and revision as a separate, pedagogical layer. This aligns with Karpathy (2025) coinage of "Vibe Coding" to describe building

software by describing intent and letting the model produce the code, and with Willison (2025) clarification that vibe coding, where the code is not reviewed, should be distinguished from responsible AI-assisted programming in production contexts where code must be reviewed and tested.

Willison (2025) offers a precise interpretation relevant for educational applications: not all AI-assisted programming qualifies as vibe coding. The term is appropriately used in cases where the human does not engage with the underlying code. In educational contexts, this distinction is important. A teacher who uses Gemini to generate an HTML game for preschool education, iterates through prompts, documents errors, and makes pedagogical decisions throughout the process is engaged in more than passive Vibe Coding; they are exercising instructional design judgement through a generative AI interface. This study adopts this qualified interpretation. The three-stage prompt structure (naive, structured, affective) is treated as a scaffolded progression from vibe coding in the strict sense toward deliberate, pedagogically grounded AI use. The term therefore describes the entry point of the task, not the whole of it.

Empirical work on vibe coding is beginning to appear, though not yet in teacher education. In a preregistered study of 100 tertiary-level students, written-communication proficiency and computer-science achievement both predicted vibe coding performance, with prompt quality as the proposed mechanism (Thorgeirsson et al., 2026). A systematic review of 101 practitioner sources reports that users iterate, test, and revise rather than accepting first outputs, and identifies quality assurance as the central open problem (Fawzy et al., 2026). In health professions education, O. Ng and Chow (2026) describe educators without formal programming training building their own educational applications through conversational prompting, together with the pitfalls they encountered. Two features of this literature are relevant here: it is concentrated on higher education and professional training rather than early childhood teacher preparation, and it indicates that competence in prompting is bound up with general communication skill rather than technical knowledge alone.

Vibe Coding also intersects with a broader reconceptualisation of teacher knowledge in AI-rich environments. Ning et al. (2024) propose the AI-TPACK framework as an extension of Mishra and Koehler (2006) Technological Pedagogical Content Knowledge model. In AI-TPACK, artificial intelligence functions both as a medium and as an object of study. Teachers are expected not only to use AI tools, but also to develop pedagogical knowledge about teaching in contexts where AI-generated outputs are readily available. This dual requirement, using AI while reasoning pedagogically about its outputs, is central for preservice preschool teachers.

Two gaps follow from this literature. First, research on generative AI in teacher education has concentrated on text-based tasks such as lesson planning and feedback (Chen, 2025; OECD, 2025), rather than on the design of interactive applications through prompting. Second, where production-oriented coding has been studied in early childhood teacher education, it has used block-based environments such as ScratchJr (Louka & Papadakis, 2024), not generative-AI prompting. How preservice early childhood teachers experience AI-assisted application design, and how their self-reported AI literacy and self-efficacy change across such a task, has not been examined. The two research questions stated in Section 1 address these gaps. The related literature on computational thinking, creative coding, and no-code/low-code tools shows that production-oriented tasks tend to produce stronger gains in self-efficacy, attitudes, and intention to use technology than consumption-oriented tasks (Bers, 2022; Louka & Papadakis, 2024).

2.4. The Affective Dimension of AI Interaction in Educational Contexts

Evidence from classroom studies shows that AI-augmented, caring feedback can support motivation and self-efficacy; syntheses indicate that affective supports are most effective when combined with cognitive supports (Sung et al., 2025; Zhang et al., 2025). These responses may include excitement, anxiety, resistance, or combinations of these, and they can shape how educators interpret and adopt AI in practice.

In this study, affective prompting refers to providing AI with emotionally and contextually rich language that specifies the desired experiential qualities of an educational artefact, drawing on educators' strengths such as pedagogical empathy and developmental sensitivity (Mollick, 2024).

2.5. Developmentally Appropriate Practice and AI in Early Childhood Education

Developmentally Appropriate Practice (DAP) provides the professional framework within which early childhood educators in [Greece] and internationally are trained to make instructional decisions (NAEYC, 2020). DAP centres the physical, social, emotional, and cognitive development of children from birth through age eight, and prioritises play-based, child-initiated, and relationship-centred learning experiences. The integration of digital tools within this framework remains debated. Some scholars support that well-designed tools can support development, while others raise concerns about overuse and reduced opportunities for unstructured play (Christakis & Hale, 2025; Hirsh-Pasek et al., 2015).

This study does not take a position on the direct use of AI tools with preschool children. Instead, it focuses on how preservice teachers are prepared to evaluate such tools and make informed pedagogical decisions. This includes assessing AI-generated applications against DAP principles and engaging in the design process as active contributors. UNESCO (2024) identifies 'AI-assisted design of learning activities' as a core teacher competency, while the European Commission's DigCompEdu framework (Redecker, 2017) places digital content creation and pedagogical integration at the centre of teacher digital competence.

2.6. Professional Identity in Preservice Teacher Education

Professional identity, understood as the evolving sense of what it means to be a teacher, is a central concern in preservice teacher education (Beauchamp & Thomas, 2009; Sachs, 2005). Beauchamp and Thomas (2009) state that teacher identity is not a stable trait but a negotiated, contextually situated process, shaped by experience, reflection, and interaction with professional communities. Regarding technology, preservice teachers' identities are highly responsive to structured learning experiences (Mishra & Koehler, 2006).

The integration of AI tools introduces additional dimensions to this process. Beyond competence, it raises questions about professional agency and responsibility. As Sachs (2005) states, professional identity shapes how teachers act and interpret their work. For preservice preschool teachers, engaging with generative AI involves not only learning to use a tool ('can I use this tool?'), but also reflecting on how such use aligns with their emerging professional role ('what kind of teacher does using this tool make me?').

2.7. Positioning of the Frameworks

These bodies of literature play different roles in the study, and the analysis treats them differently. AI literacy is the central construct: the intervention targets it, and it organises both the quantitative scales and the qualitative reading of the logbooks. TAM supplies the four acceptance constructs (PU, PEOU, ATT, BI) measured before and after the task, but does not address pedagogical agency. Developmentally Appropriate Practice is the professional standard against which participants judged their applications. Professional identity is the lens for the qualitative analysis of how participants positioned themselves

in relation to AI. The MAIL and AI-TPACK frameworks informed the design of the task rather than its analysis and are not tested here. This keeps the framework aligned with the two research questions: RQ1 concerns the measured acceptance constructs and AI literacy, and RQ2 concerns agency, prompt learning, and identity as described by participants.

3. Methodology

3.1. Research Design

The study employed a single-group pre–post design to examine changes in preservice preschool teachers' AI literacy, self-efficacy, attitudes, and technology acceptance following a structured Vibe Coding intervention. This design is common in exploratory educational technology research in authentic course settings (L. Cohen et al., 2018). A comparison group was not formed within these compulsory course units, where all enrolled students completed the same assessed task. We treat the absence of a comparison group as a limitation of the design rather than as a methodological strength, and we return to it in Section 5.5. The study integrates quantitative (pre–post questionnaires) and qualitative (reflective coding logbooks) data, following a convergent mixed-methods approach (Creswell & Plano Clark, 2018).

3.2. Context and Participants

The study was conducted in two undergraduate courses at the University of [Crete] and the [Aristotle] University of [Thessaloniki], [Greece], during the winter semester of the 2025–2026 academic year. Both courses were compulsory for students in their third year or above. The questionnaire was administered at two time points through a link open to all students enrolled in the two courses rather than to a preregistered sample. Participants entered a self-chosen pseudonymisation code at each time point, and matching required the same code at both administrations. The post-intervention export contained 53 submissions. Two duplicate submissions were removed, leaving 51 unique post-intervention respondents. Nineteen post-intervention respondents could not be matched to a pre-intervention response using the self-entered pseudonymisation code. This left 32 complete pre–post pairs, which form the quantitative analytic sample; all inferential analyses are based on these 32 cases. Figure 1 presents the participant flow. Coding logbooks were available for 28 of these 32 participants. Of these 28, two contained insufficient reflective material in Part C, leaving 26 logbooks for directed qualitative content analysis. The qualitative sample is therefore a subset of the matched quantitative sample.

The matching procedure was re-run for this version of the manuscript. An earlier pass had matched 28 cases; a subsequent check against the full list of submitted codes identified four further pairs that the earlier pass had not matched, and the analytic sample is 32. All analyses reported here were recomputed on these 32 cases, and the matched dataset and analysis script are deposited so that the flow can be verified.

All 32 participants were female (100%), consistent with the demographic profile of early childhood education programmes in Greece. Most were third-year students (N = 24, 75.0%), followed by fourth-year students (N = 5, 15.6%) and final-year students (N = 3, 9.4%). Ages ranged from 19 to 50 years (M = 21.6, Mdn = 20), with the majority aged 20–21 (N = 25, 78.1%). Participants' prior AI use and their prior experience with generative AI as a production tool are reported in the Results (Section 4.1). This profile, frequent general use alongside limited experience in creating AI-generated artefacts, informed the design of the task.

Participation was voluntary. Ethical procedures included informed consent, GDPR compliance, and strict anonymity. Data were pseudonymised prior to analysis. The study

was conducted within the context of regular course instruction. Participation in the research component was voluntary and did not affect course assessment or grades.

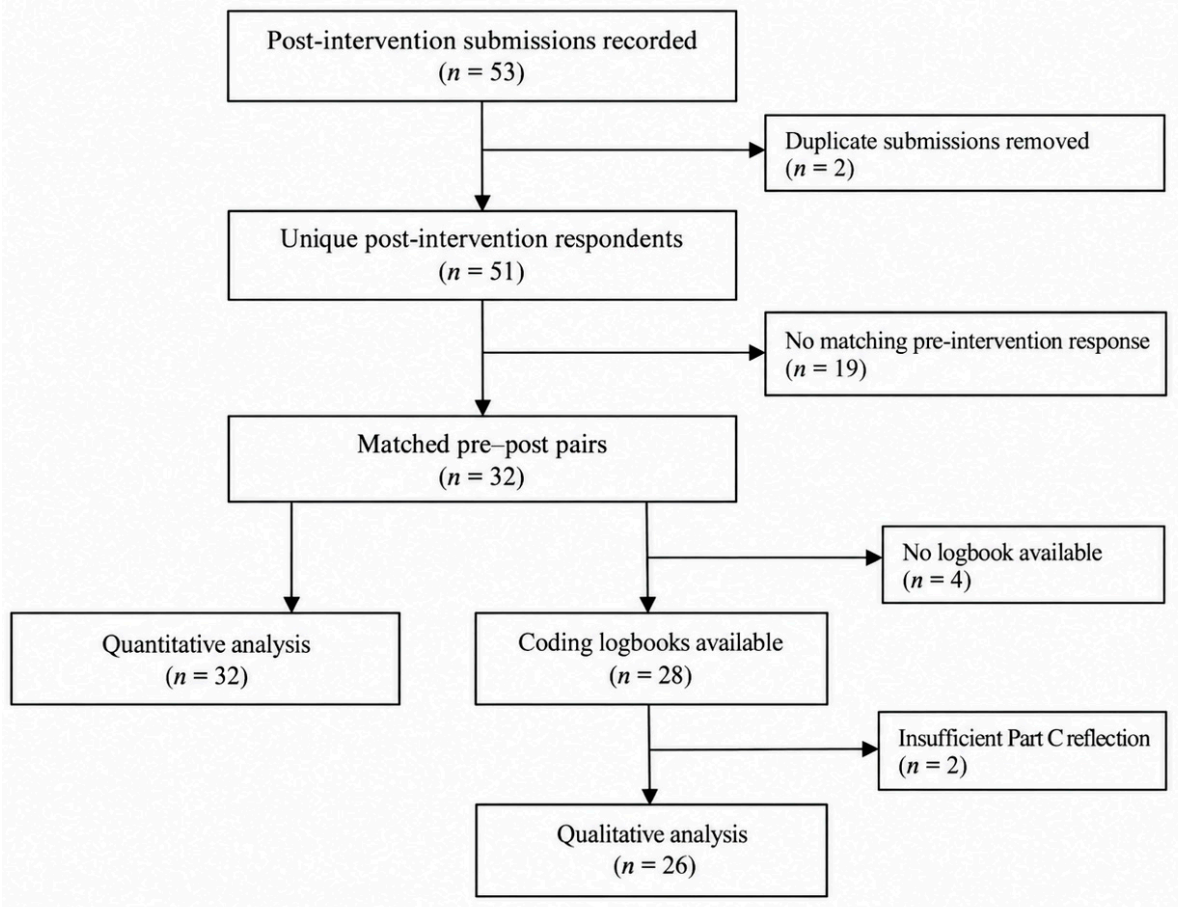


Figure 1. Participant flow. The quantitative analysis used 32 matched pre–post cases. Coding logbooks were available for 28 of these participants, and 26 contained sufficient Part C reflection for qualitative analysis.

3.3. The Vibe Coding Intervention

The intervention was an individual design task embedded in the course assessment. Each participant used Google Gemini to create an interactive web-based educational application (HTML) on a natural-science topic for children aged 3–6. The task assessed the quality of the prompts, the pedagogical appropriateness of the application for this age group, the reflective logbook, and the integration of the application into a full lesson plan; programming skill was not assessed. Participants were instructed not to write or edit code manually, and all outputs were produced through natural-language prompting, across supervised in-class sessions and independent work. Within the course this task was labelled "Assignment 2". The task followed a three-stage prompt sequence. Stage A (Naive/Unstructured) involved a free-form description without structural or pedagogical constraints. Stage B (Structured/Pedagogical) required explicit pedagogical objectives, target age group, and alignment with Developmentally Appropriate Practice (DAP). Stage C (Vibe/Affective) incorporated emotionally rich, context-specific language describing the intended experience of the application. This progression moved participants from unstructured interaction to intentional, pedagogically grounded, and affective use. Figure 2 summarises this three-stage structure and the review-and-revision loop at each stage.

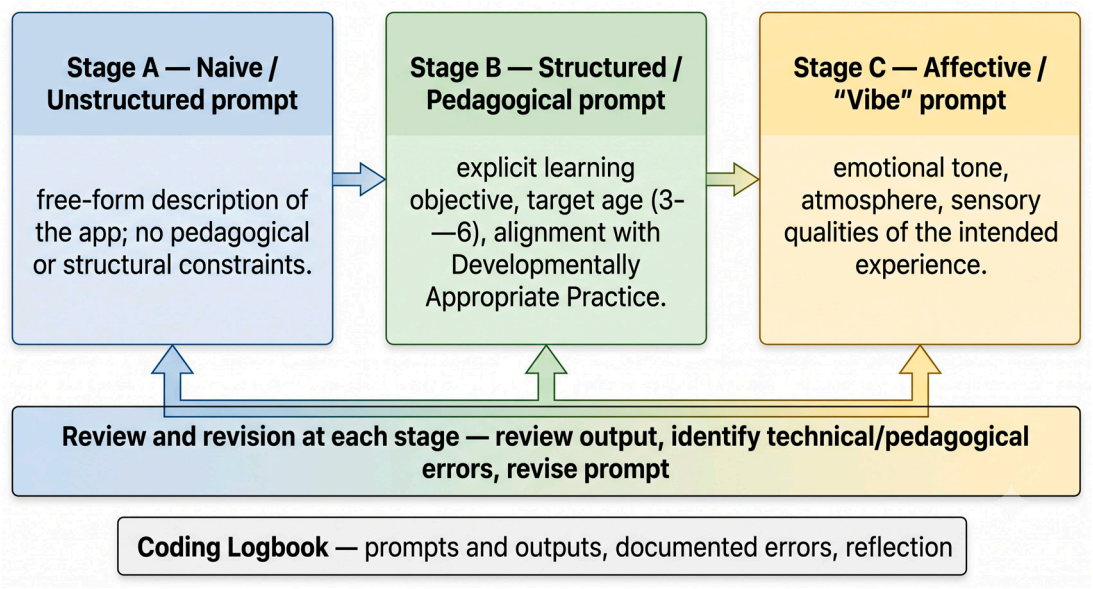


Figure 2. The three-stage prompt structure (naive → structured → affective). At each stage participants reviewed the AI output, identified errors, and revised the prompt; this evaluation layer distinguishes the task from vibe coding in the strict sense.

As part of the course assessment, students were required to submit three HTML application files, a structured coding logbook, and a lesson plan. The logbook included (a) prompts and outputs, (b) documentation of errors or hallucinations and their resolution, and (c) a reflective section on learning outcomes and personal contribution to the application's 'vibe'.

3.4. Data Collection

Quantitative data were collected via two self-report questionnaires administered online through Google Forms: a pre-intervention questionnaire completed at the start of the semester, before any AI-related instruction or task, and a post-intervention questionnaire completed immediately after the submission of the design task. Both questionnaires comprised 31 items across six scales. Pre–post responses were matched using the self-entered pseudonymisation code described in Section 3.2. After matching, cases were assigned study labels (e.g., PS1) for reporting. Qualitative data comprised the individual reflection sections (Part C) of the coding logbooks available for 28 of the 32 matched participants. Twenty-six contained sufficient Part C material for analysis; two were excluded for insufficient reflective content (see Section 3.5). The logbooks were collected as part of the standard course assessment and used for research purposes with participants' consent.

A supplementary task-specific self-assessment instrument, the PIQ rubric (Pedagogical, Interface, Quality), was developed by the research team and completed by all 32 participants at the end of the intervention. The PIQ was designed specifically for this study to capture participants' perceived quality of their AI-generated applications along three dimensions directly relevant to the Vibe Coding task: Pedagogical appropriateness (3 items: curriculum alignment, DAP compliance, feedback design), Interface and Vibe quality (3 items: atmospheric alignment, motor ergonomics, visual clarity), and Technical quality (2 items: functional reliability, responsiveness). Each item was rated on a 4-point scale. The complete PIQ rubric and its open-ended prompts are reproduced in Appendix A.

The instrument also included two open-ended items, one asking participants to describe the most significant AI error they encountered (the 'Vibe Gap') and one asking them to reproduce the corrective prompt they used, alongside a single creativity score (1–4) rating how creatively they had to frame their requirements for the AI to understand them.

As a researcher-designed instrument developed for this specific intervention context, the PIQ was not subjected to reliability or factor analysis and is therefore reported descriptively as supplementary evidence rather than included in the primary hypothesis-testing analysis. Future work may develop and validate the PIQ as a reusable evaluation rubric for AI-assisted educational design tasks.

3.5. Data Analysis

Quantitative analysis involved paired-samples *t*-tests comparing pre- and post-intervention mean scores for each of the six scales. Effect sizes were calculated as Cohen's d_z for paired samples (J. Cohen, 1988). With 32 matched pairs the analysis has limited statistical power, particularly for smaller effects, and the results are treated as exploratory. The paired *t*-test tolerates minor departures from normality at this sample size (Field, 2018); difference scores were inspected for normality and extreme outliers before testing. All quantitative analyses were conducted in Python 3.14 using the SciPy statistical library. The reflection sections (Part C) of the 26 usable logbooks were analysed using directed (deductive) qualitative content analysis (Hsieh & Shannon, 2005). Three categories were specified in advance from the aims of the intervention and the two research questions: pedagogical empowerment, prompting and technical confidence, and professional identity reconfiguration. The two authors coded the logbooks independently, assigning each logbook to the category that best characterised its reflection, and then compared their codings. No written codebook was used and no inter-rater agreement coefficient was calculated; disagreements were resolved through discussion until consensus. Each logbook was assigned to a single dominant category, so the category frequencies sum to the 26 analysed logbooks. We did not use reflexive thematic analysis, because the categories were defined a priori rather than generated from the data. Because the analysis was deductive, thematic saturation was not an applicable adequacy criterion: no new categories were being sought, so the point at which new categories cease to appear could not arise. Adequacy was instead assessed against category coverage and coding stability. Every analysed logbook could be assigned to one of the three predefined categories, no logbook required a category outside the framework, and the two coders reached consensus on all assignments through discussion. One participant had a single missing pre-intervention BI item; that participant's BI pre-score was computed from the remaining items.

The distributions of the six pre–post difference scores were inspected with the Shapiro–Wilk test and for outliers. Five were consistent with normality; the Self-Efficacy differences departed from normality ($W = 0.914$, $p = 0.014$), so this comparison was additionally evaluated with a Wilcoxon signed-rank test ($p < 0.001$), which was inferentially consistent with the paired *t*-test No difference score contained an outlier beyond $|z| = 3.29$.

3.6. Measurement Instruments

The questionnaire was developed by adapting and translating validated international instruments to the specific context of the study: preservice preschool teachers using generative AI tools (Gemini) for educational app design. All items used a 5-point response format. AI Literacy and the four TAM scales used an agreement anchor (1 = strongly disagree, 5 = strongly agree); the Self-Efficacy items used a confidence anchor (1 = not at all confident, 5 = completely confident), following the conditional-confidence format of Compeau and Higgins (1995).

AI Literacy (AIL, 12 items) was adapted from the frameworks of D. T. K. Ng et al. (2021) and Long and Magerko (2020), addressing four dimensions: conceptual understanding of AI, prompt writing ability, evaluation of AI output, and AI ethics awareness. Perceived Usefulness (PU, 4 items), Perceived Ease of Use (PEOU, 4 items), Attitude to-

ward AI (ATT, 3 items), and Behavioural Intention (BI, 3 items) were adapted from Davis (1989), contextualised for Gemini and vibe coding tasks. Self-Efficacy (SE, 5 items) was adapted from the computer self-efficacy instrument of Compeau and Higgins (1995), using conditional performance scenarios specific to AI-assisted app creation. The questionnaire was translated from English into Greek and checked by independent back-translation; discrepancies between the two versions were resolved by discussion before administration. The Greek version was piloted with 6 preservice teachers (3 per institution) who were not part of the main sample, to check item clarity and wording. We made minor wording changes after the pilot. The English-language version of the questionnaire is reproduced in Appendix B.

Internal consistency was satisfactory across all six scales. At pre-intervention: AIL $\alpha = 0.891$, PU $\alpha = 0.877$, PEOU $\alpha = 0.786$, ATT $\alpha = 0.703$, SE $\alpha = 0.866$, BI $\alpha = 0.846$. At post-intervention: AIL $\alpha = 0.882$, PU $\alpha = 0.929$, PEOU $\alpha = 0.751$, ATT $\alpha = 0.867$, SE $\alpha = 0.897$, BI $\alpha = 0.944$. The BI pre-intervention reliability was computed on 31 cases, as one participant had a missing BI item at pre-intervention and was excluded listwise from this scale's reliability estimate; all other coefficients are based on the 32 matched cases. All scales met the acceptable threshold of $\alpha \geq 0.70$ (George & Mallery, 2003). The pre-intervention ATT ($\alpha = 0.703$) and the PEOU scale ($\alpha = 0.786/0.751$) were the lowest; both met the 0.70 threshold but should be read with corresponding caution.

4. Results

4.1. Quantitative Results: Pre–Post Comparison

Of the 32 participants, general use of AI tools was frequent: 11 used them daily, 17 weekly, 3 rarely, and 1 never. Prior programming experience, including block-based tools such as Scratch and ScratchJr, was more limited: 2 reported none, 8 minimal (had seen or heard of it), 21 moderate (had made something simple), and 1 extensive. This profile, frequent general AI use alongside limited programming experience, is consistent with prior findings for this population (Kalogiannakis & Papadakis, 2019) and informed the design of a no-code production task. To address RQ1, paired-samples *t*-tests were conducted on the 32 matched pre–post cases to examine whether participation in the Vibe Coding intervention was associated with statistically significant pre–post changes across six scales. Effect sizes were calculated as Cohen's d_z for paired samples (mean difference divided by the standard deviation of the difference scores). J. Cohen (1988) benchmarks, which derive from between-subjects designs, are used as approximate reference points only. Descriptive statistics and test results are presented in Table 1. Because six comparisons were tested, *p*-values were adjusted with the Holm–Bonferroni procedure. All six remained significant at $\alpha = 0.05$ (adjusted $p \leq 0.004$). Figure 3 presents the effect sizes against Cohen's thresholds.

All six comparisons reached statistical significance ($p < 0.05$), with mean scores improving across every scale from pre- to post-intervention. The four largest effects, AIL ($d_z = 1.541$), SE ($d_z = 1.118$), PEOU ($d_z = 0.979$), and ATT ($d_z = 0.830$), all exceeded the conventional threshold for a large effect. These effect sizes are reported descriptively. In a single-group design with self-report measures, they index the size of the pre–post change in perception and should not be read as the causal magnitude of the intervention. These constructs share a common feature: they measure participants' relationship with AI as a tool. The two remaining constructs, Perceived Usefulness ($d_z = 0.562$, medium effect) and Behavioural Intention ($d_z = 0.597$, medium), showed smaller, though still statistically significant, improvements. The interpretation of this differential pattern is addressed in the Discussion. The confidence intervals are wide, which is expected with 32 matched pairs. For perceived usefulness and behavioural intention the lower bounds are consistent with

only a small change. The intervals should be read as the range of pre–post differences compatible with these data, not as bounds on an intervention effect.

Table 1. Pre–post comparison across six scales (N = 32).

Scale	Items	Pre M	Pre SD	Post M	Post SD	ΔM	95% CI	*t* (31)	*p*	*p* (Holm)	d_z	Effect
AI Literacy (AIL)	12	3.30	0.73	4.13	0.54	0.83	[0.64, 1.02]	−8.718	<0.001	<0.001	1.541	Large
Self-Efficacy (SE)	5	3.12	0.82	4.03	0.73	0.91	[0.62, 1.20]	−6.326	<0.001	<0.001	1.118	Large
Perceived Ease of Use (PEOU)	4	3.26	0.71	4.02	0.65	0.76	[0.48, 1.04]	−5.540	<0.001	<0.001	0.979	Large
Attitude toward AI (ATT)	3	3.25	0.75	3.99	0.84	0.74	[0.42, 1.06]	−4.696	<0.001	<0.001	0.830	Large
Behavioural Intention (BI)	3	3.64	0.90	4.17	0.91	0.53	[0.21, 0.85]	−3.378	0.002	0.004	0.597	Medium
Perceived Usefulness (PU)	4	3.77	0.85	4.28	0.77	0.51	[0.18, 0.84]	−3.179	0.003	0.004	0.562	Medium

Note: ΔM was calculated as post-intervention minus pre-intervention; the negative t values reflect the pre-intervention minus post-intervention subtraction order used in the paired-samples tests. Effect sizes are reported as absolute Cohen's d_z values. Confidence intervals are 95% intervals for the mean pre–post difference.

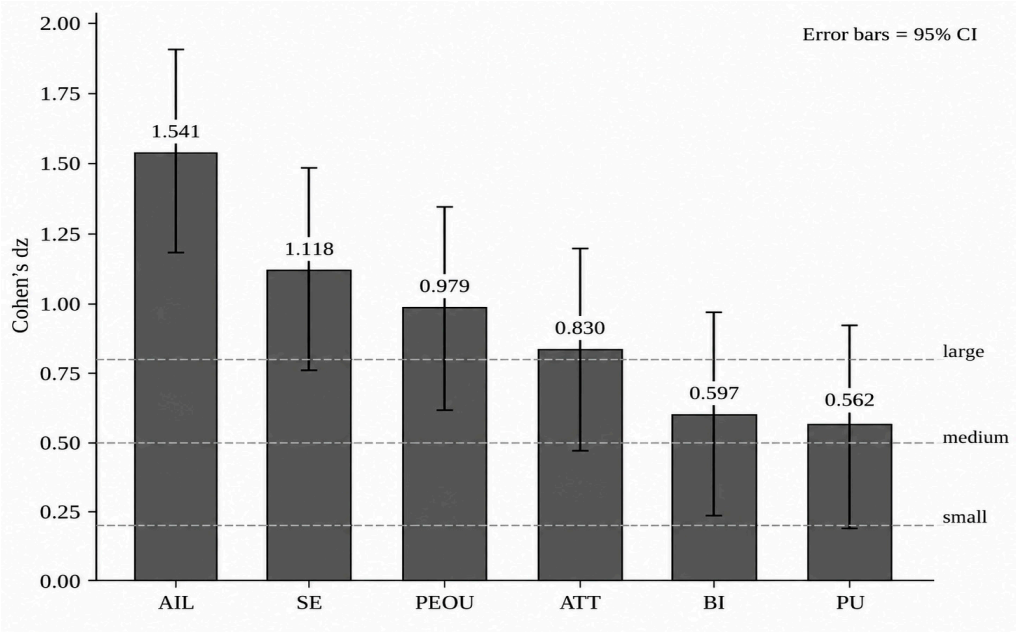


Figure 3. Pre–post effect sizes (Cohen's d_z) for the six scales (N = 32). Dashed lines mark Cohen's conventional small/medium/large thresholds, which derive from between-subjects designs and serve as approximate reference points. Error bars show 95% confidence intervals for Cohen's dz.

AI Literacy showed the largest pre–post change (ΔM = 0.83, d_z = 1.541). The iterative, three-stage structure can be interpreted as a form of applied AI-literacy practice because it required participants to prompt, evaluate, correct, and re-prompt. Self-Efficacy showed the second largest effect (ΔM = 0.91, d_z = 1.118).

4.2. Qualitative Findings: Directed Content Analysis of Coding Logbooks

To address RQ2, the reflection sections (Part C) of the 26 logbooks were analysed with the three predefined categories described in Section 3.5. The three categories and their frequencies are presented in Table 2.

Table 2. Category frequencies across 26 coding logbooks.

Category	Definition	N (%)
C1: Pedagogical Empowerment	Participants described an active pedagogical role, directing the AI toward a specific educational vision and framing their personal contribution in terms of developmental appropriateness and child-centred design.	10 (38%)
C2: Prompt Learning and Technical Confidence	Participants focused on developing the ability to communicate precisely with the AI, moving from vague, hopeful prompting toward structured, pedagogically specific instructions, often through iterative correction of AI errors.	8 (31%)
C3: Professional Identity Reconfiguration	Participants reflected on a shift, or unresolved tension, in their professional self-concept regarding AI: questioning what it means to be a preschool teacher who can create digital tools, and what limitations remain.	8 (31%)

Note: Two logbooks (PS12, PS22) were excluded from the directed content analysis because their Part C reflection sections contained insufficient reflective content. Both participants nevertheless completed the separate PIQ self-assessment, including its open-ended items, and are therefore included in the descriptive PIQ analysis reported below.

Supplementary evidence from the PIQ self-assessment (N = 32) is broadly consistent with the qualitative categories. When asked to identify the version of their application they considered best and wished to evaluate, 29 participants (90.6%) selected Version C (the affective/vibe prompt stage) and only 3 (9.4%) selected Version B (the structured prompt stage); no participant selected Version A (the naive prompt). This preference cannot be treated as an objective measure of application quality because participants evaluated the version they had already selected, introducing a potential confirmation effect. Nevertheless, the near-unanimous selection of the affective version is consistent with the Pedagogical Empowerment category: participants described the vibe-prompt stage as the point at which their educational vision was most fully realised. PIQ scores for the selected versions were broadly positive across all three dimensions: Pedagogical (M = 3.33, SD = 0.52), Interface/Vibe (M = 3.36, SD = 0.58), and Technical Quality (M = 3.41, SD = 0.85), yielding an overall PIQ mean of 3.36 (SD = 0.54) on a 4-point scale. Prompt creativity was rated M = 3.09 (SD = 0.69), indicating that most participants felt they needed to describe their requirements in moderately to highly creative terms for the AI to understand them.

The open-ended Vibe Gap responses, capturing the most significant AI errors encountered and the corrective prompts applied, support Category 2 (Prompt Learning and Technical Confidence). These accounts make visible the gap between participants' pedagogical intent and the AI's initial output, as well as the strategies used to bridge it. Recurring gap types included: inappropriate use of written text in an application designed for pre-literate children (reported by PS11, PS13, PS21, PS23, PS26, PS32); sound elements that failed to match the desired atmosphere (PS22: 'I asked for pleasant music during the game and it produced an irritating sound'; PS18: 'It did not add sound'; PS26: 'Add voice and images so they can understand on their own what they are asked to do'); interface elements too small or incorrectly positioned for children's motor skills (PS10, PS22, PS28, PS30); and pedagogically inaccurate visual representations (PS09: 'It used heart and ring

emojis to represent the planets Pluto, Uranus, and Saturn. I asked it to replace these with the correct planet emojis to be pedagogically accurate').

Corrective prompting strategies ranged along a continuum from minimal intervention to precise reformulation. At one end, participants relied on simple repetition (PS20: 'I repeated it'), while at the other, they produced highly specific, constraint-based prompts (PS08: 'Do exactly what you did but allow further attempts after an error and display a non-discouraging message'). These patterns align closely with the prompt iteration trajectory identified in Category 2 logbooks. This shift from vague to precise prompting was not driven by technical experimentation alone, but by the need to resolve concrete pedagogical misalignments. In this sense, prompt refinement functioned as a form of applied pedagogical reasoning, mediated through interaction with AI.

4.2.1. Category 1: Pedagogical Empowerment (N = 10, 38%)

The dominant construct across ten logbooks was agency: the clear perception that responsibility for the educational quality of the final product resided with the teacher, not the tool. Participants described a marked reorientation in which technical production became inseparable from pedagogical intentionality. Participant PS01 captured this with a directing metaphor: 'Through Vibe Coding, I realised that programming is not just logic but also directing'. PS30 described app creation as 'a continuous effort to transfer a specific vision and emotion through the code'. PS06 framed their contribution explicitly in DAP terms, noting that their personal input was 'to provide a reward system for the child to boost their morale'. PS13 described the intervention as a first experiment with developmentally appropriate digital design: 'I learned to give detailed and clear instructions to the AI and to patiently revise the results. This was the first time I experimented with creating a developmentally appropriate application'. These accounts show that participants took responsibility for pedagogical quality and shaped the generated outputs to match developmental goals.

4.2.2. Category 2: Prompt Learning and Technical Confidence (N = 8, 31%)

Eight participants described a clear learning trajectory from imprecise prompting toward deliberate, structurally informed interaction with the AI. The progression across the three prompt stages was experienced as a form of language acquisition, learning to make pedagogical intent legible to the system. Participant PS28 articulated a key insight: 'I realised that the AI responds better if you give specific instructions and that you can, and it is important to, give it a role (e.g., act as a kindergarten teacher, an expert programmer)'. This role prompting strategy, consistent with the recent literature (White et al., 2023), was not taught explicitly but emerged through practice. PS29 highlighted a metacognitive shift: 'I learned how to give correct and clear instructions from the beginning for a complete scientific game, without needing all three stages'. AI hallucinations, particularly the recurrent failure to implement audio-related instructions, were not experienced as barriers but as productive constraints. They functioned as boundary conditions that made the limits of natural-language interaction visible, thereby prompting refinement and deeper technical understanding.

4.2.3. Category 3: Professional Identity Reconfiguration (N = 8, 31%)

Eight participants reflected on shifts, or unresolved tensions, in their professional identity in relation to AI. These accounts ranged from expanded role awareness to explicit self-critique, indicating an ongoing process of professional becoming rather than a completed transformation. PS17 described an expanded professional role: 'My role expanded to testing and repeating the trial to get the desired results'. PS23 expressed an enhanced sense of professional capability: 'With a relatively simple method I can create quality

pedagogical material'. In contrast, PS26 offered a candid self-assessment: 'I learned that I need to be clearer and give detailed, as well as targeted, instructions. I can say I didn't fully manage to achieve this'. Collectively, these accounts reflect a dynamic process of identity negotiation, consistent with Beauchamp and Thomas (2009) conceptualisation of teacher identity as continuously constructed through experience. Participants did not present themselves as having 'arrived', but as actively repositioning their professional selves regarding AI-mediated practice.

5. Discussion

Before interpreting the results, one design constraint shapes every claim that follows. The study used a single group with no comparison condition, so the pre–post changes cannot be attributed to the Vibe Coding task alone. Testing effects, maturation over a semester, novelty, prior exposure to AI tools, social desirability in self-report, and general course learning are possible contributors. The interpretations below describe associations and participants' accounts, not effects of the intervention.

5.1. The Intervention as a Site of Applied AI Literacy Development

The largest pre–post effect was observed for AIL (d_z = 1.541), followed by SE (d_z = 1.118). Both exceed conventional thresholds for large effects, although those thresholds are only an approximate guide for within-subject d_z estimates. These are self-reported gains from an uncontrolled design, so they are read here as changes in how participants describe their use of AI, not as measured competence.

The qualitative data support this pattern (Category 2), which documents a progression from imprecise, exploratory prompting to deliberate, structurally informed interaction with the AI. The convergence across methods is informative: while quantitative data show a substantial increase in the overall AIL scale, qualitative reflections reveal how these gains were enacted as an iterative process of refining the articulation of pedagogical intent.

One practice was 'role prompting', independently identified by participant PS28, who perceived that assigning a professional persona to the AI improved the relevance of its output. Although not explicitly taught, this strategy aligns with established prompt engineering practices (White et al., 2023), indicating that participants developed tacit procedural knowledge through iterative engagement. AI hallucinations provided opportunities to learn, prompting participants to interrogate the limits of natural-language instruction and refine their inputs accordingly.

The substantial effect on self-efficacy (d_z = 1.118) is consistent with Bandura (1977) theory of mastery experience as the primary source of efficacy beliefs, as well as with the computer self-efficacy framework of Compeau and Higgins (1995). This finding is particularly relevant in early childhood teacher education, where preservice teachers have historically reported low confidence in using digital technologies (Kalogiannakis & Papadakis, 2019). The observed change suggests that structured AI-assisted design tasks warrant further investigation in this population.

5.2. From Technical Ease to Pedagogical Agency

The second cluster of quantitative findings, large effects on ATT (d_z = 0.830) and PEOU (d_z = 0.979), points beyond skill acquisition toward a shift in relational stance. These constructs capture participants' attitudes toward AI and their perceptions of the ease with which it can be used, rather than demonstrated ability; they also index how participants position themselves in relation to AI: as passive users or as active agents.

The affective prompting stage (Prompt C) is central to interpreting this shift. Empirical and review evidence treat educators' and learners' emotional responses as part of the design

space: affective feedback can improve self-efficacy, and integrated affective-plus-cognitive supports outperform affect-only approaches (Sung et al., 2025; Zhang et al., 2025). By requiring participants to articulate the desired 'vibe' of their applications, encompassing sensory qualities, emotional tone, and experiential characteristics for children aged 3–6, the intervention defined this affective dimension within the design process. The directing metaphor used by PS01, and the multiple participants who described their personal contribution in terms of emotional character and atmosphere, are consistent with the intention behind Prompt C: participants reported drawing on pedagogical and affective knowledge they already held and giving it a place in the design of the artefact. Mollick (2024) framing of AI interaction as a space for specifically human capacities, not despite the AI's technical power but because of it, resonates strongly with this pattern.

Category 1 (Pedagogical Empowerment) supports this interpretation. Participants frequently described their contribution in terms of atmosphere, emotional resonance, and developmental appropriateness. The 'directing' metaphor used by PS01 ('programming is not just logic, but also directing') captures a shift from technical execution to intentional orchestration. In this sense, Prompt C activated forms of pedagogical and affective expertise that participants already possessed but had not previously integrated into digital design.

Within the TAM framework (Davis, 1989), perceived ease of use is commonly treated as an antecedent of perceived usefulness and behavioural intention. The present study did not test these structural relationships; it examined only pre–post changes in the individual constructs. The larger change in PEOU than in PU or BI therefore describes a differential pattern of change rather than evidence for or against the relationships proposed by TAM. In the qualitative data, higher perceived ease of use coincided with accounts of a move from technical anxiety to pedagogical intent. Hudson (2026) distinction between automation and augmentation provides a useful interpretive lens. Rather than delegating tasks to AI (automation), participants described using AI to extend their own pedagogical capacities (augmentation). The accounts in Category 1, emphasising direction, emotional design, and developmental sensitivity, reflect a 'human-first, human-last' orientation toward AI use. This supports Long and Magerko (2020) definition of AI literacy as inherently critical and agentic (D. T. K. Ng et al., 2021; Long & Magerko, 2020).

Further supporting this, several participants explicitly framed their contributions in DAP terms, such as integrating motivational reward systems (PS06) or experimenting with age-appropriate digital design (PS13). In these accounts, participants described applying pedagogical knowledge they already held. References to general enthusiasm for AI were less frequent.

5.3. The Open Question: Professional Identity and Future Intention

The smallest effects were observed for PU (d_z = 0.562) and BI (d_z = 0.597), both of which require careful interpretation. These constructs had the two highest pre-intervention means (PU: M = 3.77; BI: M = 3.64), indicating that participants already held relatively positive views regarding AI's utility and future use. These relatively high starting points left less room for further increase, which may partly account for the smaller observed changes.

A theoretically substantive explanation emerges from the qualitative data (Category 3). Participants' reflections indicate that perceptions of usefulness and intentions for future use are not resolved within a single intervention but evolve over time through experience. As Beauchamp and Thomas (2009) argue, teacher identity is a dynamic, negotiated construct, while Sachs (2005) emphasizes its role in shaping professional action as 'how to be' in the profession. Participants who engaged most deeply with identity-related reflection did not describe completed transformations but ongoing processes. For example, PS17 framed

her role as expanding into testing and redesign, while PS26 acknowledged partial failure as part of learning. In these accounts, participants described identity development as still in progress. System-level evidence (Berne et al., 2025) reinforces this interpretation, highlighting a persistent gap between operational competence and pedagogical judgement in AI use. This study suggests that production-oriented, task-based interventions may help bridge this gap, not by prescribing solutions, but by generating the kinds of experiences through which meaningful professional questions emerge.

These findings imply that PU and BI are particularly sensitive to intervention duration and contextual embedding. Longitudinal research, especially following school placement experiences, would be necessary to determine whether these constructs stabilise, increase, or decline over time.

5.4. Integration of the Quantitative and Qualitative Strands

Bringing the two strands together, the results converge on the relational rather than the utilitarian aspects of AI use. The larger gains in AI literacy, self-efficacy, ease of use, and attitude correspond to Categories 1 and 2, where participants described taking a directing role and learning to prompt deliberately. The smaller gains in perceived usefulness and behavioural intention align with Category 3, where participants described their professional relationship with AI as still forming. One tension stands out: participants reported confidence and agency while remaining uncertain about future use, which suggests that a single task can shift self-perception faster than it settles intention.

Two divergences are as informative as the convergences. First, the quantitative strand records an increase in behavioural intention while the qualitative strand shows participants describing their professional relationship with AI as unsettled; a rising score on a three-item intention scale and an unresolved professional stance are not in conflict, but they are not the same finding. Second, the near-unanimous preference for the affective version in the PIQ is not matched by any independent evidence that those versions were pedagogically better; the two strands agree on what participants valued and are silent on whether that valuation was warranted. Integration here identifies where the strands speak to different objects, not only where they agree.

5.5. Limitations

Several limitations should be acknowledged. First, the single-group pre–post design does not permit causal inference: effects may reflect maturation, course exposure, or testing rather than the Vibe Coding intervention itself (L. Cohen et al., 2018). Second, the sample is small (N = 32), drawn from two institutions in a single country, and entirely female. The all-female composition reflects the actual demographic profile of Greek early childhood education programmes rather than a sampling decision, but it means the findings describe this population and cannot be extended to mixed-gender preservice cohorts, to other national teacher-education systems, or to preservice teachers preparing for other age phases. Restricting the analytic sample to fully matched complete cases may additionally have introduced selection bias, since students who completed both questionnaires may differ systematically from those who did not. The findings should be read as descriptive of this cohort. Third, the quality of qualitative data varied, with some logbooks offering limited reflective depth. Fourth, the psychometric evidence for the adapted instrument is limited to internal consistency. The scales were translated into Greek, contextualised for generative AI and preschool application design, and in some cases shortened, so the factor structure reported in the source instruments cannot be assumed to hold here. Confirmatory factor analysis was not performed: with 32 cases and 31 items the case-to-parameter ratio falls far below any conventional requirement for stable estimation, and a model fitted on

this sample would not be interpretable. Internal consistency was acceptable across all six scales, but consistency is not evidence of construct validity. The pre-intervention ATT scale ($\alpha = 0.703$) and the PEOU scale ($\alpha = 0.786/0.751$) were weakest and should be read with corresponding caution. Establishing the factor structure of this adapted instrument requires a separate validation study with a sample large enough to support it.

Fifth, the task treated prompting as the entry point to AI-assisted design, and the observed changes are correspondingly bounded by that framing. Prompt proficiency is a narrow competence: eliciting a usable output from a model is not the same as knowing when the output should be used, whether the underlying task was worth automating, or how the tool reshapes professional practice over time. Participants' accounts of pedagogical agency in this study are accounts of directing a generation process, not evidence of professional growth in a broader sense. A task built around prompting can make the limits of AI output visible, as it did here, but sustained professional development plausibly requires engagement with the institutional, curricular and ethical conditions of AI use that a single design assignment cannot address (Dourvas et al., 2026; Khodi et al., 2026). Intervention evidence points in the same direction: in a six-month professional development programme for university teachers, gains in AI technological knowledge outpaced gains in AI technological pedagogical knowledge, which the authors attribute to operational AI skills such as prompting having a low entry threshold while pedagogical transformation is gradual (Tan et al., 2025). The qualitative analysis was deductive: the three categories were defined in advance from the intervention's aims, so the analysis examined whether these constructs appeared in participants' reflections rather than allowing other patterns to surface (Hsieh & Shannon, 2005). This orientation is confirmatory by design and limits the interpretive openness of the qualitative findings.

6. Conclusions

This study examined pre–post changes in self-reported AI literacy, self-efficacy, attitudes, and technology acceptance beliefs among preservice preschool teachers enrolled in two undergraduate courses, alongside participants' reflections on pedagogical agency, prompt learning, and professional identity. Adopting a convergent mixed-methods design, it combined quantitative pre–post comparisons across six adapted scales, all with acceptable internal consistency in this sample, with directed qualitative content analysis of reflective coding logbooks (N = 26; from 28 collected).

Quantitative results showed statistically significant pre–post increases on all six self-report scales, which remained significant after Holm–Bonferroni correction. Large effect sizes were observed for AIL ($d_z = 1.541$), SE ($d_z = 1.118$), ATT ($d_z = 0.830$), and PEOU ($d_z = 0.979$), while PU ($d_z = 0.562$) and BI ($d_z = 0.597$) showed medium effects. Qualitative findings converged around three categories, Pedagogical Empowerment, Prompt Learning and Technical Confidence, and Professional Identity Reconfiguration, offering a coherent account of how these changes were experienced in practice.

The results suggest that when generative AI lowers the syntactic barrier of programming, participants described taking on a more deliberate design role. This is an account of how they positioned themselves, not a measured change in decision-making. The 'directing' metaphor that emerged in participant reflections captures this shift succinctly: the AI generates the code, while the teacher provides the educational direction.

For teacher educators, Vibe Coding tasks, when scaffolded through progressive prompt design and supported by structured reflection, may be useful for AI literacy work in preservice programmes and warrant testing in controlled designs. The three-stage prompting sequence provided opportunities for iterative, practice-based engagement with-

out requiring prior programming expertise, and may suit populations with historically low technology self-efficacy, though this requires direct comparison to confirm.

These conclusions are exploratory and bounded by the design. The single-group pre–post structure does not support causal inference, so the changes may reflect testing, maturation, novelty, or general course effects rather than the task. The sample is small, all-female, and drawn from two institutions, which limits transfer. The outcomes are self-reports, not measured competence or assessed product quality. The qualitative analysis used consensus coding, which carries interpretive subjectivity. Future research should incorporate controlled designs, larger and diverse samples, and longitudinal follow-up to examine whether observed shifts in professional identity persist and transfer to classroom practice.

As generative AI tools enter early childhood settings, preparing educators to engage with them purposefully and in developmentally appropriate ways is a critical priority. This study offers an initial evidence-based contribution toward that goal and highlights directions for future research.

Author Contributions: Conceptualization, S.P. & G.Z.; methodology, S.P. & G.Z.; software, S.P. & G.Z.; validation, S.P. & G.Z.; formal analysis, S.P. & G.Z; investigation, S.P. & G.Z.; resources, S.P. & G.Z. data curation, S.P. & G.Z.; writing—original draft preparation, S.P. & G.Z.; writing—review and editing, S.P. & G.Z.; visualization, S.P. & G.Z.; supervision, S.P. & G.Z.; project administration, S.P. & G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Institutional Review Board Statement: The study was conducted within the context of regular course instruction. The study was conducted in accordance with the Declaration of Helsinki, and reviewed and approved by the Ethics Committee of the Department of Preschool Education, School of Education, University of Crete, and by Decision No. 664/18-02-2026 of the Assembly of the Department of Preschool Education, University of Crete, following the recommendation of the Departmental Ethics Committee (approval date: 18 February 2026). Participation in the research component was voluntary and did not affect course assessment or grades.

Informed Consent Statement: Participation was voluntary. Informed consent was obtained from all individual participants, and data were pseudonymised prior to analysis to preserve strict anonymity.

Data Availability Statement: All data will be available at OSF: https://osf.io/exr3n/overview?view_only=c96470453b3a4824ad28f73264275e97 (accessed on 19 September 2026). Full student essays cannot be shared due to privacy restrictions; summary data and code are provided for replication. The quantitative data file uses neutral analysis identifiers (R01–R32) that are not linked to the pseudonyms used for logbook quotations, to protect participant anonymity.

Acknowledgments: Participants consented to the use of their submitted coding logbooks for research purposes.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. PIQ Self-Assessment Instrument (Researcher-Designed; Exploratory, Non-Validated; Reported Descriptively)

Completed once by all participants at the end of the intervention.

Part A, Selection of best version. Participants chose which of their three application versions to evaluate: Version A (basic prompt), Version B (prompt with role/constraints), or Version C (full Vibe prompt). They also gave the pseudonym, the application title, and a short written justification for their choice.

Part B, PIQ rubric. Each row rated on a four-point scale (1 = inadequate, 2 = adequate, 3 = very good, 4 = excellent).

Pedagogical (P):

- Curriculum goal: the application serves a clear learning objective (e.g., classification, phonological awareness) and is not merely a game.
- DAP (age-appropriateness): content and logic suit preschool children (e.g., no long text, direct action–reaction).
- Feedback: praise is encouraging ("Well done!") and errors are handled with humour or sound without discouraging the child.

Interface and Vibe (I):

- Vibe alignment: the atmosphere fits the theme (e.g., "magic garden", "calm sea") and is neither dull nor frightening.
- Ergonomics (motor skills): buttons and objects are large enough for a child's fingers and spaced apart.
- Visual media: large emojis or CSS shapes are used instead of broken internet images.

Quality, technical (Q):

- Functionality: the application runs without bugs and responds to every click.
- Responsiveness: it adjusts correctly to the browser window without losing elements.

Part C, Prompt engineering and vibe correction.

- Vibe Gap (open): the largest initial error the AI made in aesthetics or function (e.g., "Click Here" text, small buttons).
- Vibe Correction (open): the exact natural-language prompt used to fix that problem.
- Prompt Creativity (1–4): how creatively the requirement had to be described for the AI to understand it (1 = not at all, 4 = very).

Appendix B. Full Questionnaire (Administered Pre and Post, Identical at Both Points)

Items were completed in Greek. The questionnaire used two response anchors: a five-point agreement anchor (1 = strongly disagree, 5 = strongly agree) for AI Literacy and the four TAM scales, and a five-point confidence anchor (1 = not at all confident, 5 = completely confident) for Self-Efficacy. Items are grouped below by the scale to which they were assigned for scoring.

Background items: pseudonymisation code; gender; age; year of study; frequency of AI-tool use (never/rarely/weekly/daily); prior programming experience (none/minimal, have seen or heard of it/moderate, have made something simple/extensive).

AI Literacy (AIL, 12 items)

Conceptual understanding:

1. I can explain in simple terms what artificial intelligence is.
2. I broadly understand how a language model such as Gemini produces its responses.
3. I know the basic limitations of free AI tools (e.g., outdated data, "hallucinations").

Functional use:

4. I can write suitable prompts so that the AI creates educational material for preschool-age children.
5. I can evaluate the quality and usefulness of the text or code the AI produces.
6. I can combine AI with other tools (e.g., ScratchJr, LearningApps) to create an educational activity.

Critical and ethical understanding:

7. I can recognise when an AI output may be inaccurate or biased.
8. I can explain why it is important to check AI responses before using them in the classroom.
9. I am aware of privacy and ethics issues when AI tools are used in education.

Pedagogical application:

10. I can judge whether an AI output is pedagogically appropriate for preschool-age children.
11. I can use AI to design activities that promote children's creativity.
12. I can use AI tools to support my own professional learning and reflection.

Perceived Usefulness (PU, 4 items)

1. Using AI (Gemini) will help me create educational material more quickly.
2. The ability to create applications (Vibe Coding) will improve the quality of my teaching.
3. I consider the ability to develop web apps a valuable skill for the contemporary preschool teacher.
4. Overall, I find AI useful for my professional development.

Perceived Ease of Use (PEOU, 4 items)

1. It is (or seems) easy for me to learn to operate Gemini to create applications.
2. The process of giving natural-language instructions (prompts) is clear and understandable.
3. I can easily get the AI (Gemini) to produce exactly what I have in mind.
4. Interacting with this tool does not require much mental effort. (Displayed in the form under the Attitude heading; scored as PEOU by content.)

Attitude toward Using (ATT, 3 items)

1. Using Vibe Coding to create material is a good idea.
2. I find the process of programming through dialogue (AI) enjoyable.
3. I like using Gemini to make games. (Displayed in the form under the Self-Efficacy heading; scored as ATT by content.)

Self-Efficacy (SE, 5 items; confidence anchor)
Stem: How confident do you feel that you can create a functional educational application. . .

1. . . .if there is no one around to tell me what to do (to give me the prompt).
2. . . .if I have only the notes or the software's built-in help guide.
3. . . .if someone first shows me how to do it (modelling).
4. . . .even if I face a technical problem or error (debugging) and have to fix it.
5. . . .with only a general idea of the lesson or application I want to design.

Behavioural Intention (BI, 3 items)

1. I intend to use ChatGPT/Gemini-type tools when I work in a school.
2. I would recommend that my colleagues learn this method of creating material.
3. I will keep experimenting with creating applications with AI after this course ends.

References

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191. [CrossRef] [PubMed]

Beauchamp, C., & Thomas, L. (2009). Understanding teacher identity: An overview of issues in the literature and implications for teacher education. *Cambridge Journal of Education*, *39*(2), 175–189. [CrossRef]

Berne, J., Doss, C., & Shapiro, A. (2025). *American public school pre-K teacher survey 2025*. RAND Corporation.

Bers, M. U. (2022). *Beyond coding: How children learn human values through programming*. MIT Press.

Chen, B. (2025). Beyond tools: Generative AI as epistemic infrastructure in education. *arXiv*, arXiv:2504.06928. [CrossRef]

Christakis, D. A., & Hale, L. (Eds.). (2025). *Handbook of children and screens: Digital media, development, and well-being from birth through adolescence*. Springer.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, *19*(2), 189–211. [CrossRef]

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. [CrossRef]

Dourvas, I., Kokkonis, G., & Kontogiannis, S. (2026). Reconceptualizing prompt engineering as reflective professional practice: A framework for teacher development. *Electronics*, *15*(5), 930. [CrossRef]

Fawzy, A., Tahir, A., & Blincoe, K. (2026). Vibe coding in practice: Motivations, challenges, and a future outlook—A grey literature review. In *Proceedings of ICSE 2026, software engineering in practice, Rio de Janeiro, Brazil, April 12–18* (pp. 212–223). ACM Digital Library.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Allyn & Bacon.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in 'educational' apps: Lessons from the science of learning. *Psychological Science in the Public Interest*, *16*(1), 3–34. [CrossRef]

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. [CrossRef] [PubMed]

Hudson, E. (2026, January 4). *Four priorities (not predictions) for 2026*. Learning on Purpose. Available online: https://erichudson.substack.com/p/four-priorities-not-predictions-for (accessed on 19 September 2026).

Kalogiannakis, M., & Papadakis, S. (2019). Pre-service kindergarten teachers' acceptance of ScratchJr as a tool for learning and teaching computational thinking and science education. *The Journal of Emergent Science*, *15*, 31–34.

Karpathy, A. (2025, February 3). *The year of vibe coding* [Post]. X (formerly Twitter). Available online: https://x.com/karpathy/status/1886192184808149383 (accessed on 19 September 2026).

Khodi, A., Curle, S., & Parra-Guinaldo, V. (2026). Pedagogical prompting rather than technical mastery: Generative AI use by English and English-medium instruction teachers. *Frontiers in Education*, *11*, 1901680. [CrossRef]

Long, D., & Magerko, B. (2020, April 25–30). *What is AI literacy? Competencies and design considerations* [Conference session]. 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA. [CrossRef]

Louka, K., & Papadakis, S. (2024). Enhancing computational thinking in early childhood education through ScratchJr integration. *Heliyon*, *10*(10), e30482. [CrossRef]

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record: The Voice of Scholarship in Education*, *108*(6), 1017–1054. [CrossRef]

Mollick, E. (2024). *Co-intelligence: Living and working with AI*. Portfolio/Penguin.

NAEYC. (2020). *Developmentally appropriate practice* (4th ed.). National Association for the Education of Young Children.

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

Ng, O., & Chow, M. (2026). Empowering health professions educators: Developing educational tools with AI-assisted vibe coding. *Medical Science Educator*, *36*(1), 47–52. [CrossRef]

Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge elements. *Sustainability*, *16*(3), 978. [CrossRef]

OECD. (2025). *Digital education outlook 2025: Artificial intelligence and education*. OECD Publishing.

OECD. (2026). *PISA 2029 media & artificial intelligence literacy (MAIL)*. OECD Publishing.

Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu*. Publications Office of the European Union.

Sachs, J. (2005). Teacher education and the development of professional identity: Learning to be a teacher. In P. Denicolo, & M. Kompf (Eds.), *Connecting policy and practice: Challenges for teaching and learning in schools and universities* (pp. 5–21). Routledge.

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, 13–35. [CrossRef]

Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI–Künstliche Intelligenz*, *35*(1), 109–118. [CrossRef]

Sung, G., Guillain, L., & Schneider, B. (2025). Using AI to care: Lessons learned from leveraging generative AI for personalized affective-motivational feedback. *International Journal of Artificial Intelligence in Education*, *35*, 1913–1952. [CrossRef]

Tan, X., Cheng, G., & Ling, M. H. (2025). Enhancing teachers' AI competency: A professional development intervention study based on intelligent-TPACK framework. *Computers and Education: Artificial Intelligence*, *9*, 100521. [CrossRef]

Teo, T. (2011). Factors influencing teachers' intention to use technology: Model development and test. *Computers & Education*, *57*(4), 2432–2440. [CrossRef]

Thorgeirsson, S., Weidmann, T. B., & Su, Z. (2026, April 13–17). *Computer science achievement and writing skills predict vibe coding proficiency* [Conference session]. 2026 CHI Conference on Human Factors in Computing Systems, Barcelona, Spain. [CrossRef]

UNESCO. (2024). *AI competency framework for teachers*. UNESCO. [CrossRef]

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. *arXiv*, arXiv:2302.11382.

Willison, S. (2025, March 19). *Not all AI-assisted programming is vibe coding (but vibe coding rocks)*. Simon Willison's Weblog. Available online: https://simonwillison.net/2025/Mar/19/vibe-coding/ (accessed on 19 September 2026).

Zhang, H., Liu, Y., Jiang, M., Chen, J., Wang, M., & Paas, F. (2025). Emotional artificial intelligence in education: A systematic review and meta-analysis. *Educational Psychology Review*, *37*, 106. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.