

Article

AI-Supported Design of Teaching Units for English to Young Learners: A Case Study in Initial Teacher Education

Cecilia Lazzeretti 

Faculty of Education, Free University of Bozen-Bolzano, 39042 Brixen-Bressanone, Italy; cecilia.lazzeretti@unibz.it

Abstract

While generative artificial intelligence (GenAI) is increasingly used by university students for writing support, less is known about its role in discipline-specific professional tasks. This study examines how pre-service primary teachers integrate and conceptualise GenAI when designing Teaching Units for English for Young Learners (EYL), with a focus on whether AI is positioned as a substitute for pedagogical reasoning or as a support within teacher decision-making. The qualitative study involved 75 fifth-year pre-service teachers at the Free University of Bozen-Bolzano (Italy), working in 23 groups. Data included 23 Teaching Units and 10 AI Use Reports, analysed through document analysis and thematic coding. GenAI was used mainly for material production (visual and text generation, idea generation, and text revision) and resource adaptation, with limited evidence of use for macro- or micro-planning decisions (objectives, sequencing, assessment). Prompts were often underspecified, but reports described iterative refinement and critical adaptation to improve age appropriateness and reduce lexical overload. Overall, within a transparent course framework, pre-service teachers retained pedagogical ownership while using GenAI as a supplementary resource, underscoring the need to develop pedagogically grounded AI literacy (prompt design, evaluation, and disclosure).

Keywords: generative artificial intelligence (GenAI); English as a foreign language (EFL); English for young learners (EYL); pre-service primary teachers; teaching unit design

1. Introduction

The public availability of Generative Artificial Intelligence (GenAI) tools has rapidly altered the landscape of educational practice. Beyond their technical novelty, such systems challenge fundamental assumptions about expertise, creativity, and professional judgement. In higher education, the debate has often centred on academic integrity and authorship (Perkins, 2023; McDonald et al., 2025). Yet in professional degree programmes, a different question emerges: what happens when AI intersects not merely with academic writing, but with the development of professional competence?

In the Italian context, this shift has unfolded alongside rapidly evolving policy and institutional positioning. At national level, the *Italian Strategy for Artificial Intelligence 2024–2026* frames AI as a cross-sector priority and includes education and training among the domains requiring coordinated development (Presidenza del Consiglio dei Ministri, 2024). In higher education, universities have increasingly issued guidance that foregrounds transparency, critical verification of outputs, and user responsibility for accuracy and ethics (University of Bologna, 2024; University of Trento, 2024). Taken together, these developments signal a move from ad hoc reactions to more formalised expectations regarding when and how GenAI may be used in academic work.



Academic Editor: Beryl Exley

Received: 4 March 2026

Revised: 7 April 2026

Accepted: 8 April 2026

Published: 11 April 2026

Copyright: © 2026 by the author.

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.

This broader policy momentum is also visible in the school sector. The Ministry of Education and Merit has issued national guidance to support the informed and safe adoption of AI in schools, emphasising human oversight, transparency, and data protection as key principles (Ministero dell'Istruzione e del Merito, 2025). For language education, these principles matter because EFL teaching in primary school is framed by national curriculum guidance that stresses age-appropriate progression, inclusivity, and coherent learning objectives (Ministero dell'Istruzione, dell'Università e della Ricerca, 2012). In such a setting, GenAI may offer efficiencies in material preparation, but it also raises pedagogical questions about linguistic control (e.g., lexical load) and the professional judgement required to adapt resources for young EFL learners (Bland, 2015; Shin & Crandall, 2014).

Teacher education represents a particularly sensitive terrain in this regard. Learning to teach is not limited to acquiring theoretical knowledge; it involves cultivating the capacity to design learning experiences that are developmentally appropriate, pedagogically coherent, and contextually responsive. At the heart of this professional formation lies planning, which has long been described as a core component of teacher effectiveness and instructional competence (Anderson & International Institute for Educational Planning, 1991; Romiszowski, 2016). Designing a teaching unit requires future teachers to translate curricular goals into structured sequences of activities, anticipate learners' needs, select suitable materials, and balance cognitive, linguistic, and affective dimensions of learning.

The arrival of GenAI tools introduces both opportunities and tensions into this process. On the one hand, AI systems can generate texts, visuals, songs, and activity ideas with remarkable speed. On the other hand, planning for young learners—particularly in the context of English as a Foreign Language (EFL)—demands a nuanced understanding of developmental stages, linguistic proficiency levels, and classroom realities that may not be automatically embedded in AI outputs (Bland, 2015; Shin & Crandall, 2014). The critical issue is therefore not simply whether AI can produce materials, but how future teachers engage with those materials: Do they delegate pedagogical reasoning to the machine, or do they position AI as a subordinate tool within their professional decision-making?

This question is especially pertinent in primary education. Teaching English to young learners involves complex micro- and macro-planning decisions: sequencing input, limiting lexical load, designing multimodal tasks, and ensuring alignment with curricular standards. For pre-service teachers, navigating these decisions already constitutes a demanding cognitive task. The introduction of AI adds another layer of mediation—requiring them not only to design instruction, but also to formulate prompts, evaluate outputs, and critically revise generated content.

The present case study investigates this emerging intersection between AI use and professional planning within initial teacher education. Conducted in Italy in a multilingual university context where AI use was permitted under conditions of transparency, the research examines how pre-service primary teachers incorporated (or chose not to incorporate) GenAI tools in the design of Teaching Units for English for young learners.

Rather than focusing on abstract ethical debates or technical capabilities, this study centres on situated pedagogical practice. It asks how future teachers negotiate AI within a concrete professional task that lies at the core of their identity formation.

The research is guided by the following question:

RQ: How do pre-service primary school teachers integrate and conceptualise the use of Generative AI tools in the design of Teaching Units for English as a Foreign Language for young learners?

By examining AI use within the authentic practice of lesson planning, this study contributes to a more nuanced understanding of how digital technologies are reshaping professional learning in teacher education.

2. Literature Review

2.1. Generative AI in Higher Education: Prohibition or Regulation?

The rapid diffusion of GenAI tools such as large language models (LLMs) has profoundly reshaped higher education. Since the public release of conversational AI systems, universities worldwide have been compelled to reconsider long-standing assumptions about authorship, assessment, academic integrity, and digital literacy (Duah & McGivern, 2024; Idris et al., 2024; McDonald et al., 2025; Peláez-Sánchez et al., 2024; Perkins, 2023; Michel-Villarreal et al., 2023).

The literature suggests that institutional responses have largely converged around two dominant approaches. The first approach advocates restriction or prohibition (de Fine Licht, 2024; Perkins, 2023). From this perspective, GenAI tools are perceived primarily as a threat to academic integrity, originality, and the development of critical thinking skills. Concerns include plagiarism, ghostwriting, erosion of disciplinary knowledge, and the outsourcing of cognitive effort. Institutions adopting this stance tend to limit or ban AI use in assessments, emphasising detection technologies and academic misconduct policies.

The second approach favours regulation through guidelines and transparency mechanisms (Guizani et al., 2025; Jin et al., 2025; McDonald et al., 2025). Rather than framing AI as inherently problematic, this perspective recognises its inevitability and pedagogical potential. Scholars in this camp argue for explicit policies that define acceptable use, require disclosure statements, and integrate AI literacy into curricula. Under this model, students may use AI tools, but must document when, how, and for what purposes they have done so, thereby fostering accountability and critical engagement.

Increasingly, the literature suggests that outright prohibition may be unsustainable, given the accessibility and rapid evolution of AI systems. As a result, attention is shifting toward structured integration, ethical frameworks, and the development of AI-related competencies (Memarian & Doleck, 2023). Within this broader debate, however, disciplinary and contextual differences remain underexplored—particularly in professional programmes such as teacher education.

2.2. Main Uses of Generative AI by Students

Empirical research on students' AI practices reveals a range of common applications, most of which cluster around efficiency and linguistic support (Kim et al., 2025). One of the most widespread uses is basic proofreading, including spelling correction, grammar checking, and stylistic refinement. Students frequently rely on AI systems to improve clarity, coherence, and formality in written assignments (Barrett & Pack, 2023). Closely related is the use of AI for text modification, such as reducing the length of a text to meet a word limit or paraphrasing content to improve conciseness. These practices reflect a utilitarian orientation, in which AI is perceived as an editorial assistant (Koltovskaia et al., 2024). Another common application concerns initial idea generation and research scaffolding. Students often use AI tools to brainstorm potential arguments, outline essay structures, generate research questions, or obtain preliminary summaries of unfamiliar topics. In such cases, AI functions as a cognitive catalyst rather than a final content producer (Koltovskaia et al., 2024). Automatic translation is also frequently reported, particularly in multilingual contexts. Students may translate entire texts or specific segments to facilitate comprehension or to produce drafts in a target language (Koltovskaia et al., 2024).

Across these applications, a recurring theme in the literature is the predominance of AI as a support tool for writing-related tasks. Less attention has been devoted to how students in professional programmes—such as teacher education—use AI to design pedagogical materials or plan instructional sequences. This relative gap is particularly evident in the context of primary education and English for Young Learners (EYL).

2.3. Prompt Engineering as an Emerging AI Literacy Skill

Defined as “the skill of crafting, refining, and optimising prompts to facilitate effective interactions between users and GenAI, with the aim of generating desired outputs from LLMs” (Hwang et al., 2025, p. 1164), prompt engineering has emerged as a central component of AI literacy. Its primary goal is to identify best practices that enable large language models (LLMs) to generate accurate, relevant, and context-sensitive outputs. However, the literature reveals significant uncertainty regarding how this skill should be operationalised in educational contexts. Much of the existing knowledge about effective prompting remains anecdotal and case-based (Schmidt et al., 2024). Some scholars argue that students should be explicitly trained through exposure to model prompts and guided practice (Schmidt et al., 2024). In contrast, qualitative research suggests that students may spontaneously develop sophisticated prompting strategies through iterative experimentation, facilitated by the intuitive interface and high usability of conversational AI systems (Sawalha et al., 2024).

Further investigation is necessary to clarify the obstacles students encounter when developing prompt engineering skills. Their proficiency likely depends on a variety of influences, such as the belief that generative AI systems comprehend language as humans do, coupled with a limited grasp of how these technologies actually function (Han et al., 2023; Knoth et al., 2024; Mollick & Mollick, 2023; Schmidt et al., 2024). For instance, studies indicate that students frequently engage with ChatGPT as if they were conversing with another person, incorporating polite expressions and adopting a conversational tone (Sawalha et al., 2024). However, there is currently no firm evidence to suggest that this manner of interaction improves the relevance or accuracy of the model’s replies.

Despite these contributions, prompt engineering has rarely been examined within discipline-specific professional tasks, such as lesson planning. In teacher education, prompting is not merely a technical activity; it requires pedagogical precision, awareness of learner characteristics, and alignment with curricular standards. Whether student teachers naturally incorporate such pedagogical constraints into their prompts remains an open question.

2.4. Micro- and Macro-Planning in Primary Education and the Role of AI

Planning lies at the core of professional teaching practice (Anderson & International Institute for Educational Planning, 1991). In primary education, in particular, lesson planning is a central component of pedagogical competence.

Planning operates at two interconnected levels (Romiszowski, 2016). Macro-planning refers to long-term instructional design, including curriculum alignment, definition of learning objectives, sequencing of units, and coherence across lessons. Micro-planning concerns the detailed organisation of individual lessons, including activity timing, classroom management strategies, materials selection, differentiation, and assessment procedures. In the context of EYL, planning assumes additional complexity. Teachers must account for learners’ developmental stage, limited literacy skills, short attention spans, and the gradual acquisition of communicative competence. Activities must be linguistically accessible (often pre-A1 to A1 level), cognitively appropriate, and emotionally engaging (Bland, 2015; Shin & Crandall, 2014).

Recent research has begun to explore the potential of AI tools to support teacher planning. Proposed benefits include generating lesson ideas, producing differentiated materials, aligning activities with learning standards, and reducing preparation time (Alwaqdani, 2025; Ruiz-Rojas et al., 2023). However, much of this research remains exploratory and often focuses on secondary or tertiary education contexts. Very few studies have examined AI-assisted planning in primary education (Hsu et al., 2024; Sakamoto et al., 2024), and even fewer have addressed the specific domain of English for Young Learners (Han & Li,

2024). Moreover, existing research rarely analyses how pre-service teachers integrate AI into both macro- and micro-planning processes, or how they reflect on the pedagogical adequacy of AI-generated outputs.

This gap is significant. If planning constitutes the professional core of teaching, understanding how future teachers negotiate AI support within this domain is crucial for shaping responsible and pedagogically sound integration strategies in initial teacher education.

2.5. Theoretical Framing for GenAI Integration in EFL Teacher Education

To interpret how student teachers incorporate GenAI into lesson and unit design, this review draws on three complementary lenses. First, from a sociocultural perspective, tools are not neutral add-ons but mediational means that can reshape what learners attend to and how they organise activity; in teacher education, GenAI can therefore be understood as a resource that mediates planning, prompting, evaluation, and revision rather than as an autonomous “planner” (Lantolf, 2000; Vygotsky, 1978). Second, the Technological Pedagogical Content Knowledge (TPACK) framework (Schmidt et al., 2009) highlights that meaningful technology integration depends on the interplay among content knowledge, pedagogical knowledge, and technology knowledge; for EFL/EYL, this interplay is visible in decisions about linguistic grading, multimodality, sequencing, and age-appropriateness (Mishra & Koehler, 2006; Shin & Crandall, 2014). Third, reflective practice foregrounds how teachers learn by evaluating consequences of instructional decisions and iterating on design; this is particularly relevant when GenAI outputs require verification, adaptation, and justification (Gibbs, 1988). Together, these lenses frame GenAI use as a situated process of mediated pedagogical reasoning, where the key analytic issue is not whether AI is used, but how it is orchestrated within professional judgement.

2.6. Advantages and Limitations of GenAI Tools in Teaching EFL to Young Learners

Across recent studies, reported advantages of GenAI in language education include rapid generation of draft materials (stories, dialogues, worksheets, images), brainstorming activity ideas, and supporting differentiation through alternative versions of the same input (e.g., simplified texts or multiple task variations), potentially reducing preparation time and widening the repertoire of classroom resources (Alwaqadani, 2025; Ruiz-Rojas et al., 2023). In higher education settings, students also report using GenAI for linguistic scaffolding such as paraphrasing, translation, and revision, which can be repurposed in teacher education as a way to prototype classroom language and evaluate register or clarity (Barrett & Pack, 2023; Kim et al., 2025; Koltovskaia et al., 2024). In EYL contexts, these affordances are most valuable when they remain subordinate to pedagogical constraints such as controlled lexical load, age-appropriate content, and alignment with learning objectives (Bland, 2015; Shin & Crandall, 2014).

At the same time, the literature points to limitations that are pedagogically consequential for EFL. First, GenAI systems can produce plausible but inaccurate content and may obscure sources, requiring verification and careful teacher oversight (Duah & McGivern, 2024; McDonald et al., 2025). Second, outputs may embed cultural bias or stereotyped representations and may generate language that is mismatched to learners’ proficiency level unless prompts specify constraints clearly (Hwang et al., 2025; Memarian & Doleck, 2023). Third, the use of GenAI raises governance issues around transparency, authorship, and assessment validity, which institutions have begun to address through disclosure policies and usage guidelines (Jin et al., 2025; Perkins, 2023). Finally, there is a risk of overreliance that shifts attention from pedagogical reasoning to surface-level production, particularly if AI is treated as a shortcut rather than a tool for iterative design and reflective evaluation (de Fine Licht, 2024; Perkins, 2023). For teacher education, a balanced view

therefore requires emphasising both the potential for productive support and the need for AI literacy practices—prompting, critical evaluation, and documentation—grounded in the specific demands of EFL teaching with young learners.

3. Materials and Methods

3.1. Research Design

The study adopts a qualitative exploratory design aimed at investigating how future primary school teachers integrate and conceptualise the use of AI tools in the design of Teaching Units for EFL. Data were generated through document analysis of student-produced artefacts, namely the completed Teaching Units and the accompanying AI Use Reports.

3.2. Research Setting

The study is situated within a naturalistic educational setting, as the data emerged from regular coursework activities rather than from an externally imposed experimental design. This ecological validity allows for an authentic understanding of AI integration practices within initial teacher education.

A reflective-analytical lens was adopted to interpret students' AI Use Reports. In particular, the reports were examined in light of the reflective dimensions by Gibbs (1988). Although students were not formally instructed to structure their reports according to Gibbs' Reflective Cycle, many of their reflections spontaneously addressed elements corresponding to its stages: description of use, evaluation of outcomes, identification of limitations, and consideration of possible improvements. The model therefore provided a useful heuristic framework for organising and interpreting the reflective data.

The study was conducted during the 2025/2026 academic year at the Free University of Bozen-Bolzano (unibz), a multilingual institution located in the bilingual (German–Italian) Autonomous Province of Bolzano–South Tyrol, Italy. The university's distinctive plurilingual model, explicitly embedded in its statute, requires instruction across three languages: German, Italian, and English. Unibz enforces rigorous language entry and exit requirements. At undergraduate level, most programmes require a B2 level in two of the three institutional languages upon entry and at least a B1 level in the third language by the second semester of the first year. By graduation, students must demonstrate C1 proficiency in two languages and B2 in the third. At Master's level, requirements vary slightly across faculties; some programmes adopt more stringent criteria, while others are offered entirely in English and require proficiency solely in that language. These requirements are supported by the university's Language Centre, which provides modular language courses in German, Italian, and English.

The present study involved students enrolled in the fifth year of the Master's Degree in Primary Education. This programme places particular emphasis on preparing future teachers to manage linguistic and cultural diversity in nursery and primary school settings and to engage effectively with families and the broader local community. The curriculum includes courses in pedagogy, psychology, anthropology, and subject didactics, including English language education. Students complete two English courses (50 h each) in the fourth and fifth years and must meet the institutional language certification requirements described above. Upon graduation, they are qualified to teach English in kindergarten and primary school contexts.

3.3. Participants and Course Design

Participants were 75 fifth-year students attending the compulsory English laboratory course during the 2025/2026 academic year. The laboratory functions as a practice-oriented space in which teacher trainees operationalise theoretical knowledge through hands-on

activities related to EFL instruction for young learners. The laboratory adopts a project-based approach: students work collaboratively to design and produce a complete Teaching Unit for preschool or primary school learners. Each Teaching Unit includes clearly defined language learning objectives; a sequence of lesson plans; instructional materials and classroom activities; and assessment tools aligned with intended learning outcomes. Planning requirements explicitly address both macro- and micro-planning dimensions, including learner age and developmental stage; selection of lexis and grammar; timing and sequencing of activities; classroom setting; teaching–learning approaches; materials selection and adaptation; alignment with CEFR descriptors for young learners; and compliance with provincial and national curriculum guidelines.

Inclusion criteria were: (a) enrolment in the compulsory fifth-year English laboratory course in 2025/2026, and (b) participation in a group that submitted a complete Teaching Unit for course assessment. Exclusion criteria were limited to materials that were not produced as part of the course assignment. Students were informed that anonymised coursework and reflective reports could be used for research purposes. The assignment was completed in groups, resulting in 23 groups and 23 Teaching Units for evaluation. Assessment focused on pedagogical coherence, methodological soundness, and the appropriateness of the designed units, rather than solely on the quality of the materials produced.

3.4. Integration of Artificial Intelligence Tools

For the first time in the 2025/2026 edition of the laboratory, students were encouraged to explore the use of AI tools—such as ChatGPT and similar chatbots—to support the preparation of their Teaching Units. AI tools were presented as optional aids for generating classroom resources (e.g., reading texts, vocabulary lists, flashcards, activity ideas, or visual prompts). However, explicit guidelines clarified that the pedagogical design of the Teaching Unit—including its structure, learning objectives, sequencing, and assessment strategies—had to be independently conceived by the students. Any AI-generated content required critical evaluation and revision to ensure linguistic accuracy, age appropriateness, and cultural sensitivity. Transparency was mandatory and, in cases where AI tools were used, each group was required to submit an AI Use Report appended to the project. Submission of this report was conditional on documented AI use; groups that did not employ AI in their assignment were not obliged to provide such a report. The report documented whether AI tools were used and at which stages of the project. It also required the exact prompts entered, the nature of the generated outputs, and a short reflective commentary on the usefulness and limitations of the tool. This documentation served both as a data source for the study and as a reflective exercise aimed at fostering critical digital literacy among future teachers.

3.5. Data Collection and Analysis

The dataset consisted of 23 Teaching Units (one per group) and 10 AI Use Reports submitted by groups who declared AI integration. The Teaching Units were analysed to identify explicit references to AI-generated materials and to categorise the types of outputs incorporated (e.g., visuals, rhymes, dialogues, activity ideas, revised texts). The AI Use Reports were examined through qualitative thematic analysis.

Operationalisation of analytic categories. The unit of analysis was each group's Teaching Unit and, where applicable, its AI Use Report. The four typologies of AI use were operationalised as follows: (1) *visual generation*—prompts requesting the creation of images/visuals for classroom use; (2) *text generation*—prompts requesting the creation of new linguistic input (e.g., rhymes, stories, dialogues, songs); (3) *idea generation*—prompts requesting activity ideas, lesson ideas, or brainstorming suggestions; and (4) *text revision*—

prompts requesting translation, proofreading, or rewriting of existing text. When a report contained multiple prompts, each prompt was coded by typology; group-level counts were then derived from whether at least one prompt of a given typology occurred in the report.

Interpreting participant–AI interaction. Because the study draws on documents (Teaching Units and AI Use Reports) rather than on recordings of real-time prompting, “interaction” is interpreted analytically as the sequence of prompts, outputs, and reflective statements reported by each group. Indicators such as prompt iteration (e.g., “we had to change the prompt several times”) are therefore treated as evidence of an iterative human–AI workflow as documented by participants, while recognising that the full interactional process (e.g., intermediate prompts/outputs not included in the report) cannot be reconstructed.

1. Identification and categorisation of AI uses, leading to the four typologies presented in the Results (visual generation, text generation, idea generation, and text revision).
2. Prompt analysis, focusing on structure, specificity, the presence or absence of pedagogical constraints (e.g., learner proficiency level, EFL context), and evidence of iterative refinement.
3. Reflective content analysis, examining how students evaluated AI outputs, described limitations, and articulated learning processes.

Particular attention was paid to indicators of emerging professional judgement, such as awareness of lexical overload, age appropriateness, and alignment with curricular guidelines. The absence of an AI Use Report in 13 out of 23 groups was also treated as analytically meaningful data. While it cannot be determined whether these groups did not use AI or chose not to disclose it, this pattern was considered in the interpretation of transparency practices and perceived legitimacy of AI use.

To enhance credibility, the analysis followed a systematic coding procedure, with iterative rereading of the corpus and constant comparison across groups. The study does not claim statistical generalisability; rather, it offers an in-depth examination of AI integration practices within a specific institutional and pedagogical context.

3.6. Ethical Considerations

The study was conducted within a regular university laboratory course and involved the analysis of students’ coursework and reflective reports produced as part of routine academic activities. All data were fully anonymised prior to analysis, ensuring that individuals could not be identified directly or indirectly.

4. Results

4.1. Disclosure and Non-Disclosure of AI Use

Among the 23 groups, 10 submitted an AI Use Report and explicitly documented the integration of AI tools in the preparation of their Teaching Unit. The remaining 13 groups did not submit an AI Use Report. Given that the report was only required when AI tools were used, the absence of a report indicates either non-use or non-disclosure. In the submitted reports, groups typically documented the tool(s) used, their prompts, the generated outputs, and short reflective notes on usefulness and limitations.

4.2. Typologies of GenAI Use in Teaching-Unit Production

Analysis of the 10 AI Use Reports allowed the identification of four main typologies of requests: visual generation, text generation, idea generation, and text revision.

(A) Visual Generation

Four groups reported using AI tools to generate images to support classroom activities. A representative prompt was:

“Create an image. A hidden-object scene in a chaotic elf workshop. Many elves are working on gifts. Christmas packages everywhere. There are tables, shelves, and work surfaces. Place the following objects only once each: a book, a teddy bear, a toy car, a basketball, a chocolate bar, a doll, an electric guitar, a toy airplane, a toy motorcycle, a pianola, a jacket, a tablet, a tennis racket, a toy train, a wristwatch, and a radio.” (#R6)

This example illustrates a relatively high degree of procedural specificity regarding visual content (e.g., the constraint that each object should appear only once). AI was therefore used instrumentally to produce engaging visual material that would otherwise require advanced graphic design skills or extensive preparation time.

(B) Text Generation

Four groups used AI to generate rhymes, stories, dialogues, or songs. One prompt reads:

“Please generate a rhyme for the topic emotion, using the proposed words: happy, scared, angry and sad. The rhyme should be linked to the book *The Colour Monster*. Use simple language for children.” (#R2)

Another example involved a more generic request:

“Write a story related to the topic ‘Birthday’ for primary school children.” (#R7)

In these cases, AI functioned as a creative linguistic resource. However, the prompts often lacked key pedagogical specifications, such as learners’ proficiency level or their status as EFL learners. As a result, some outputs contained vocabulary deemed too complex. One group reflected that the generated story included “some difficult words, which require explanations,” indicating a mismatch between the output and the intended learner profile.

(C) Idea Generation

Only one group explicitly used AI to generate activity ideas:

“Give us ideas for activities to carry out in class related to the topics animals, means of transportation and clothing.” (#R1)

In this dataset, AI use for ideational support was documented in one report.

(D) Text Revision and Proof-Editing (1 group)

Similarly, only one group reported using AI for translation and formal revision:

“Please translate the German words into English and rewrite the sentences if needed to make them clear and formal.” (#R4)

Proofreading and stylistic revision are frequently reported uses of GenAI in higher education; in this dataset, however, AI use was documented predominantly for producing classroom-facing materials rather than for academic-text optimisation.

Across the groups, ChatGPT was the most frequently cited tool (6 groups). One group used Microsoft Copilot, while two groups employed more specialised tools such as Sora (for image generation) and Easymusic (for music generation). Overall, AI tools were primarily used to generate supplementary teaching materials rather than to replace macro- or micro-planning decisions. None of the reports indicated that AI had been used to structure parts of the Teaching Unit, formulate learning objectives, or design assessment frameworks.

4.3. Prompt Features, Iteration, and Pedagogical Constraints

A key finding concerns the generally limited sophistication of prompting. With few exceptions, prompts were short and omitted pedagogically relevant constraints. Most did not specify, for example, the learners' proficiency level (e.g., pre-A1 or A1), that the target group consisted of EFL learners rather than native speakers, or constraints related to vocabulary, length, or alignment with curricular objectives.

For example, the prompts:

"Write a dialogue about jobs and professions." (#R8)

"Write a story related to the topic 'Birthday' for primary school children." (#R7)

did not clarify that the dialogue was intended for young, pre-A1 EFL learners. As a result, the generated output required vocabulary that exceeded learners' expected competence. Similarly, the birthday story prompt omitted any reference to linguistic level or lexical boundaries, which likely contributed to the presence of difficult vocabulary. These issues were addressed during classroom sessions. Through short inputs and hands-on exercises, students were guided to recognise the importance of tailoring materials and prompts to young EFL learners and to the specific level of instruction. This practical work reinforced the need to embed pedagogical constraints in prompts when using AI tools.

Despite the initial lack of specificity in many prompts, students frequently demonstrated emergent pedagogical awareness in their reflections. Several groups recognised mismatches between AI-generated output and the linguistic needs of young EFL learners.

For instance, one group explained:

"First, we revised the text of the book with the help of AI tools. Since the original version was linguistically too demanding, we made adjustments to the vocabulary. In addition, we slightly modified the storyline so that it better suited our teaching unit." (#R3)

This excerpt shows that AI was used not only to generate new content but also to support the adaptation of an existing authentic text. At the same time, the group retained control over narrative coherence and pedagogical alignment by modifying the storyline independently. Similarly, another group noted:

"The bingo cards [generated by ChatGPT] were really nice [. . .], but we had to cut some pictures out, because we thought that there were too many new words." (#R7)

This comment reveals sensitivity to lexical overload and cognitive load. In another reflection, a group reported:

"We often had to adapt or change our prompt to get the desired results. Especially when we tried to create the rhyme, we had to change the prompt several times, so that we ended up changing words and expressions on our own." (#R2)

Seven out of the ten reporting groups indicated that they had to reiterate or refine their prompts before obtaining satisfactory results. The interaction with AI was therefore not linear or automatic; it involved iterative negotiation, critical filtering, and pedagogical adjustment.

Many prompts employed polite forms such as "Please create . . ." or "Please generate . . .", although there is no empirical evidence that such politeness markers enhance output quality. This tendency may reflect anthropomorphisation of the tool rather than strategic optimisation.

4.4. Evidence of Emerging Professional Judgement in Adapting GenAI Outputs

Despite the overall limited sophistication in prompt formulation, several cases demonstrate emerging professional judgement and creative adaptation. A particularly innovative

example concerned privacy constraints. Students were unable to use actual photographs of children due to data protection regulations but required visual representations for envisioning activities. To address this limitation, they asked AI tools to generate cartoon-like portraits inspired by each child, thereby maintaining the pedagogical objective while respecting privacy restrictions.

“We have created images depicting three moments from the three lessons of our didactic unit. The children in the images are not real, which is why their faces are visible and have not been censored.” (#R9)

This case illustrates that AI was not merely used as a convenience tool but also as a means of solving context-specific professional problems.

Overall, across the 10 AI Use Reports, GenAI was used primarily for generating or revising supplementary teaching materials (visuals, short texts, activity ideas) rather than for structuring the Teaching Unit itself (e.g., sequencing, objectives, assessment). The excerpts also show that groups frequently evaluated and modified outputs (e.g., reducing lexical load, revising storylines, refining prompts) before incorporation into the final Teaching Units.

5. Discussion

5.1. Transparency, Disclosure, and Normative Uncertainty

The relatively low rate of declared AI use (10/23 groups) can be read in two non-exclusive ways. Some groups may have used GenAI but opted not to disclose it, suggesting continuing uncertainty about what counts as legitimate “acceptable” AI use in assessed coursework, even when transparency is formally required. Alternatively, some groups may have decided that GenAI was not particularly suitable for a task that foregrounded pedagogical coherence and professional responsibility. In either case, the pattern aligns with recent analyses showing that institutions are still converging on workable disclosure norms and that students often navigate AI use in a context of shifting expectations regarding authorship and integrity (Duah & McGivern, 2024; Jin et al., 2025; McDonald et al., 2025; Perkins, 2023).

5.2. What GenAI Was Used for (and What It Was Not Used for)

Consistent with work showing that students commonly use GenAI for language-related support (proofreading, paraphrasing, translation, and drafting) (Barrett & Pack, 2023; Kim et al., 2025; Johnston et al., 2024), the reports in this study also documented text-focused uses (e.g., rhyme/story generation and revision). However, the dominant pattern was the production of classroom-facing materials (visuals, short texts, printable resources) rather than the outsourcing of planning decisions. This complements emerging evidence in teacher education that GenAI can support lesson-planning workflows while leaving teachers responsible for pedagogical reasoning (Hsu et al., 2024; Sakamoto et al., 2024). Importantly, the present findings suggest that, in an EYL task with explicit constraints (age, proficiency level, lexical load), pre-service teachers tended to keep macro- and micro-planning (sequencing, objectives, assessment) under their control and used GenAI mainly as a supplementary generator of artefacts.

5.3. Prompting, Iteration, and Pedagogically Grounded AI Literacy

The results show that many initial prompts were underspecified with respect to pedagogically critical constraints (e.g., pre-A1/A1 level, lexical limits, EFL learner profile), a pattern that resonates with research positioning prompt engineering as a developing component of AI literacy rather than an intuitive skill (Hwang et al., 2025; Schmidt et al., 2024). At the same time, several groups documented iterative refinement and selective

editing—adjusting prompts, changing words, and trimming vocabulary—before accepting outputs. This aligns with findings that students often learn prompting through experimentation and revision cycles (Knoth et al., 2024; Sawalha et al., 2024). For EYL, where linguistic grading and age-appropriateness are central (Bland, 2015; Shin & Crandall, 2014), these iterations can be interpreted as early indicators of professional judgement: GenAI becomes useful not as an authority on “what to teach,” but as a resource whose products must be evaluated against child-centred and curriculum-aligned criteria. Accordingly, initial teacher education may benefit from treating GenAI use as an opportunity to make pedagogical constraints explicit (e.g., specifying target lexis, discourse functions, task length), to practise systematic evaluation of outputs, and to normalise transparent documentation of AI contributions.

5.4. Summary of Contributions and Implications in Relation to Prior Work

Taken together, these findings extend research on students’ GenAI practices, which has largely focused on writing-related applications such as proofreading, paraphrasing, translation, and drafting (Barrett & Pack, 2023; Deep & Chen, 2025; Kim et al., 2025; Johnston et al., 2024), by illustrating how GenAI is appropriated within a discipline-specific professional task: designing EFL Teaching Units for young learners. They also complement early work in teacher education, showing that GenAI can support lesson-planning workflows while leaving pedagogical reasoning with teachers (Hsu et al., 2024; Sakamoto et al., 2024; Han & Li, 2024). Finally, the results add nuance to discussions of AI literacy by showing that prompt specificity is often initially limited but can develop through iterative refinement and critical evaluation (Hwang et al., 2025; Sawalha et al., 2024; Knoth et al., 2024). As an exploratory case study, these contributions should be read as a starting point for further work that replicates and tests the emerging patterns across contexts and with additional data sources.

6. Conclusions

This case study shows that, within a course framework requiring transparency and critical evaluation, pre-service primary teachers tended to use GenAI as a supplementary resource for producing and adapting classroom-facing materials (e.g., visuals and short texts) rather than as a substitute for pedagogical reasoning in macro- and micro-planning. A second inference is that pedagogically relevant AI literacy in EYL is closely tied to the ability to specify constraints (e.g., proficiency level, lexical limits, task length) and to iteratively evaluate and revise outputs against age-appropriate and curriculum-aligned criteria. Overall, the findings suggest that teacher education can leverage GenAI to support design work while reinforcing professional judgement, accountability, and disclosure practices.

Limitations. This study is limited by its single-institution setting and context-specific task design, which restrict its statistical generalisability. Although the participant cohort comprised 75 students, the primary unit of analysis was group-produced artefacts (23 Teaching Units and 10 AI Use Reports), which constrains the breadth of observable AI practices and warrants cautious interpretation of frequencies and patterns. Evidence of GenAI use was based on submitted AI Use Reports and document analysis; therefore, non-submission may reflect either non-use or non-disclosure, and the study cannot estimate the prevalence of undeclared use. In addition, the analysis focuses on artefacts and written reflections rather than on real-time interaction with tools (e.g., prompting processes captured via screen recordings) or classroom enactment of the Teaching Units.

Future research. Future studies could replicate this design across multiple institutions and national contexts and include mixed-methods data (e.g., log data, screen capture, interviews) to better trace how prompts evolve and how evaluative decisions are made.

Research could also examine whether and how GenAI-supported materials translate into classroom practice and learner outcomes in EYL settings and which instructional supports (e.g., prompt templates, constraint checklists, disclosure rubrics) most effectively develop pedagogically grounded AI literacy in initial teacher education.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted within a regular university laboratory course and involved the analysis of students' coursework and reflective reports produced as part of routine academic activities. All data were fully anonymised prior to analysis, ensuring that individuals could not be identified directly or indirectly. In accordance with the EU General Data Protection Regulation (GDPR, Regulation 2016/679) and the Italian Codice in materia di protezione dei dati personali (D.Lgs. 196/2003, updated), anonymised educational data processed with appropriate safeguards do not require formal approval from an Ethics Committee or Institutional Review Board.

Informed Consent Statement: Verbal informed consent was obtained from the participants. Verbal consent was obtained rather than written because the study was conducted within the context of a regular university laboratory course and involved the analysis of students' coursework produced as part of normal assessment activities. No additional data collection procedures were implemented beyond routine teaching practice, and all materials were fully anonymised prior to analysis.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy reasons.

Acknowledgments: During the preparation of this manuscript, the author used Microsoft Copilot for proofreading and revision. The author reviewed and edited the output and takes full responsibility for the content of this publication.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CEFR	Common European Framework of Reference
EFL	English as a Foreign Language
EYL	English for Young Learners
GenAI	Generative Artificial Intelligence
LLM	Large Language Models

References

- Alwaqdani, M. (2025). Investigating teachers' perceptions of artificial intelligence tools in education: Potential and difficulties. *Education and Information Technologies*, 30(3), 2737–2755. [\[CrossRef\]](#)
- Anderson, L. W., & International Institute for Educational Planning. (1991). *Increasing teacher effectiveness*. UNESCO.
- Barrett, A., & Pack, A. (2023). Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, 20, 59. [\[CrossRef\]](#)
- Bland, J. (2015). *Teaching English to young learners*. Bloomsbury Publishing.
- Deep, P. D., & Chen, Y. (2025). The role of AI in academic writing: Impacts on writing skills, critical thinking, and integrity in higher education. *Societies*, 15, 247. [\[CrossRef\]](#)
- de Fine Licht, K. (2024). Generative artificial intelligence in higher education: Why the “banning approach” to student use is sometimes morally justified. *Philosophy & Technology*, 37, 113. [\[CrossRef\]](#)
- Duah, J. E., & McGivern, P. (2024). How generative artificial intelligence has blurred notions of authorial identity and academic norms in higher education, necessitating clear university usage policies. *The International Journal of Information and Learning Technology*, 41, 180–193. [\[CrossRef\]](#)
- Gibbs, G. (1988). *Learning by doing: A guide to teaching and learning methods*. Further Education Unit.
- Guizani, S., Mazhar, T., Shahzad, T., Ahmad, W., Bibi, A., & Hamam, H. (2025). A systematic literature review to implement large language model in higher education: Issues and solutions. *Discover Education*, 4, 35. [\[CrossRef\]](#)

- Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, 126, 103502. [CrossRef]
- Han, J., Yoo, H., Kim, Y., Myung, J., Kim, M., Lim, H., Kim, J., Lee, T. Y., Hong, H., Ahn, S.-Y., & Oh, A. (2023). RECIPE: How to integrate ChatGPT into EFL writing education. In *Proceedings of the tenth ACM conference on learning @ scale* (pp. 416–420). Association for Computing Machinery. [CrossRef]
- Hsu, H. P., Mak, J., Werner, J., White-Taylor, J., Geiselhofer, M., Gorman, A., & Capurro, C. T. (2024). Preliminary study on pre-service teachers' applications and perceptions of generative artificial intelligence for lesson planning. *Journal of Technology and Teacher Education*, 32, 409–437. [CrossRef]
- Hwang, M., Jeens, R., & Lee, H.-K. (2025). Exploring learner prompting behavior and its effect on ChatGPT-assisted English writing revision. *The Asia-Pacific Education Researcher*, 34, 1157–1167. [CrossRef]
- Idris, M. D., Feng, X., & Dyo, V. (2024). Revolutionizing higher education: Unleashing the potential of large language models for strategic transformation. *IEEE Access*, 12, 67738–67757. [CrossRef]
- Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, 8, 100348. [CrossRef]
- Johnston, H., Wells, R. F., Shanks, E. M., Boey, T., & Parsons, B. N. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity*, 20, 2. [CrossRef]
- Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies*, 30, 1265–1300. [CrossRef]
- Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, 6, 100225. [CrossRef]
- Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of ChatGPT for academic text revision: Behavioural, cognitive, and affective engagement. *Journal of Second Language Writing*, 65, 101130. [CrossRef]
- Lantolf, J. P. (Ed.). (2000). *Sociocultural theory and second language learning*. Oxford University Press.
- McDonald, N., Johri, A., Ali, A., & Collier, A. H. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*, 3, 100121. [CrossRef]
- Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, 5, 100152. [CrossRef]
- Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences*, 13, 856. [CrossRef]
- Ministero dell'Istruzione, dell'Università e della Ricerca. (2012). *Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione*. Available online: https://www.mim.gov.it/documents/20182/51310/DM+254_2012.pdf (accessed on 7 April 2026).
- Ministero dell'Istruzione e del Merito. (2025). *Linee guida per l'utilizzo dell'intelligenza artificiale a scuola*. Available online: https://www.mim.gov.it/documents/20182/0/MIM_Linee+guida+IA+nella+Scuola_09_08_2025-signed.pdf/b70fdc45-4b75-1f7e-73bf-eab12989b928 (accessed on 7 April 2026).
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108, 1017–1054. [CrossRef]
- Mollick, E. R., & Mollick, L. (2023). Assigning AI: Seven approaches for students, with prompts. *SSRN Electronic Journal*. [CrossRef]
- Peláez-Sánchez, I. C., Velarde-Camaqui, D., & Glasserman-Morales, L. D. (2024). The impact of large language models on higher education: Exploring the connection between AI and Education 4.0. *Frontiers in Education*, 9, 1392091. [CrossRef]
- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2), 7. [CrossRef]
- Presidenza del Consiglio dei Ministri. (2024). *Strategia italiana per l'intelligenza artificiale 2024–2026*. Presidenza del Consiglio dei Ministri.
- Romiszowski, A. J. (2016). *Designing instructional systems: Decision making in course planning and curriculum design*. Routledge.
- Ruiz-Rojas, L. I., Acosta-Vargas, P., De-Moreta-Llovet, J., & Gonzalez-Rodriguez, M. (2023). Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. *Sustainability*, 15, 11524. [CrossRef]
- Sakamoto, M., Tan, S., & Clivaz, S. (2024). Social, cultural and political perspectives of generative AI in teacher education: Lesson planning in Japanese teacher education. In *Exploring new horizons: Generative artificial intelligence and teacher education* (p. 178). AACE—Association for the Advancement of Computing in Education.
- Sawalha, G., Taj, I., & Shoufan, A. (2024). Analysing student prompts and their effect on ChatGPT's performance. *Cogent Education*, 11, 2397200. [CrossRef]
- Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK) the development and validation of an assessment instrument for preservice teachers. *Journal of research on Technology in Education*, 42(2), 123–149. [CrossRef]
- Schmidt, D. C., Spencer-Smith, J., Fu, Q., & White, J. (2024). Towards a catalog of prompt patterns to enhance the discipline of prompt engineering. *ACM SIGAda Ada Letters*, 43, 43–51. [CrossRef]

- Shin, J. K., & Crandall, J. (2014). *Teaching young learners English: From theory to practice*. National Geographic Learning.
- University of Bologna. (2024). *Guidance on the use of generative AI tools in teaching, learning, and assessment*. Available online: <https://www.unibo.it/en/university/statute-standards-strategies-and-reports/artificial-intelligence> (accessed on 7 April 2026).
- University of Trento. (2024). *Guidelines for the responsible use of generative AI in university study and teaching*. Available online: <https://www.unitn.it/en/artificial-intelligence> (accessed on 7 April 2026).
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

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.