Computers and Education Open 10 (2026) 100325

Contents lists available at ScienceDirect

Computers and Education Open

journal homepage: www.sciencedirect.com/journal/computers-and-education-open

Pre-service teachers' agency during their interactions with generative AI while designing for learning – a process view on Intelligent-TPACK

Kristina Krushinskaia *, Jan Elen, Annelies Raes

Faculty of Psychology and Educational Sciences, Centre for Instructional Psychology & Technology (CIP&T), Dekenstraat 2, Leuven 3000, Belgium

ARTICLE INFO

Keywords:
Instructional design
Design for learning
Generative artificial intelligence
Human agency
Teacher-AI interaction
Pre-service teachers
Intelligent-TPACK

ABSTRACT

Agency is considered a core societal value, yet it remains unknown how human agency is affected in human-AI interactions. This study is among the first to use process-level indicators to examine teacher agency during interactions with generative AI (GenAI) in the context of instructional design. While instructional design is regarded as a key teacher role, research shows that teachers often avoid this cognitively demanding task. Today, GenAI offers new opportunities to support teachers in the design process. One approach involves using expert bots customized with one instructional design model, which scaffold the design process and offer ready-made components (e.g., learning goals). However, it is unclear how interactions with an expert bot versus basic ChatGPT influence teacher agency. In this research, we analyzed process data and qualitative feedback from an experiment involving 78 pre-service teachers randomly assigned to one of two conditions: (1) the expert bot condition, where participants interacted with a custom GenAI bot; and (2) the ChatGPT condition, where participants could request any type of support. Three indicators of agency were examined: (1) length of interaction, (2) ownership, and (3) collaborative problem-solving behaviors. The first two indicators suggested higher agency in the expert bot condition, as participants interacted longer and used significantly more self-generated prompts. However, the third indicator revealed reduced agency in both conditions. Instead of providing meaningful feedback on the bot's suggestions, participants in the expert bot condition often agreed with its output. This suggests that the material contexts of ChatGPT and the expert bot both have unique constraints on teachers' agency.

1. Introduction

Agency, broadly defined as the capacity to act and make one's own choices [1], is considered a core value in our societies. This concept is deeply embedded in moral and political practices [2], as well as in the foundational principles of philosophy, education, and psychology [3]. For example, Bandura's [4] theory of human development relies on the concept of agency, emphasizing that people are active contributors to their life circumstances rather than mere products of them. Fostering learners' agency is argued to be a necessary component of self-regulated learning [5], and an increased level of agency is considered a desirable outcome of education.

Given its importance, one of the most intriguing questions in modern educational research has become: *How is human agency affected when individuals interact with generative artificial intelligence (GenAI)?* This question arises because GenAI is increasingly conceptualized as an *agent* [6] or an *infrastructure* [7,8], "not a mere passive tool" [9, p. 7]. This conceptualization highlights that GenAI also enacts a form of agency, expressed through its ability to make choices during automatic task execution. However, it remains unknown what implications for human agency arise when two agents, humans and GenAI, work together on the same task, what power dynamics emerge, and what short- and long-term consequences this may entail.

The current AI-in-education literature frequently emphasizes the potential threat to human agency in human-AI interactions [10,11]. One reason for this concern is that GenAI can be used for *cognitive offloading* [7], meaning that humans can delegate tasks to GenAI with potentially detrimental consequences for their agency. As a response, the prevailing ambition in the literature has become to preserve human agency [12–15]. For example, UNESCO's AI competency framework for teachers [16] encourages educators to adopt a human-centered mindset when interacting with AI, emphasizing the prioritization of human rights and needs.

For teachers, preserving human agency requires new forms of

* Corresponding author.
E-mail address: kristina.krushinskaia@kuleuven.be (K. Krushinskaia).

https://doi.org/10.1016/j.caeo.2025.100325
Received 11 July 2025; Received in revised form 29 November 2025; Accepted 12 December 2025
Available online 13 December 2025

teacher knowledge to ensure that GenAI is integrated in ways that support rather than replace human roles. One framework that specifies new knowledge elements for AI integration into teaching is Intelligent-TPACK [17], which extends the TPACK framework [18] by adding an ethics dimension and new knowledge indicators. Interestingly, Intelligent-TPACK [17] was developed before the emergence of publicly available GenAI with the intention to measure teachers' technological, pedagogical, content, and ethical knowledge needed to integrate AI-based tools into instruction [17]. Although some of its indicators relate to teacher's agency (e.g., *"I can understand the justification of any decision made by an AI-based tool"*), the framework does not explicitly address how the use of GenAI may influence teachers' agency. In other words, while Intelligent-TPACK provides a valuable conceptual starting point, further empirical work is needed to examine which additional agency-related knowledge components are required in the GenAI era.

In this study, we examine pre-service teachers' agency in the context of instructional design, also referred to as the design of learning environments, when using GenAI. Instructional design corresponds to the reflective and systematic process of iteratively framing and reframing a problem in learning, identifying learner needs [19], and selecting and sequencing learning tasks, tools to be used, and support mechanisms [20]. Instructional design (otherwise also called design for learning) is gaining increasing recognition in research and is proposed as a central element of teaching [19,21–24]. However, the full potential of designing, which is quality education where learners are supported in the best possible ways to effectively reach learning goals, is far from being realized in real classrooms. This may be explained by the nature of teachers' design processes. According to empirical studies, these processes are usually enacted in an intuitive rather than systematic manner and are rarely grounded in a relevant learning or instructional theory [25,26]. These findings highlight the need to support teachers during design, which becomes even more prominent when teachers are inexperienced, such as pre-service teachers [23]. Recent efforts have been concentrated on developing technology to support teachers during the design process [21,24,27]. For example, Laurillard [24] introduced the Learning Designer as a tool designed to help educators plan, represent, and share learning activities. Asensio-Pérez et al. [21] developed an Integrated Learning Design Environment, which guides teachers through the full cycle of learning design, from conceptualization to implementation. While learning design and instructional design stem from different traditions, both refer to the practice of designing learning environments [28]. In this study, we use the term "instructional design" to emphasize a systematic approach to the design process.

The mentioned projects can be conceived as descendants of earlier systems developed to support teachers during the design process [29,30], and many related issues, including the problem of agency, were raised long before the advent of GenAI. In the 1990s, for example, research placed high expectations on expert systems developed to guide novices in instructional design [31]. Yet the implications of expert systems for human agency raised serious concerns, as pointed out by Salomon et al. [32]: *"What if the technological component contains sufficient expertise actually to decrease the intellectual share of the human partner, as is nowadays the case with expert piloting systems that supplant the human pilot during dangerous landings? What if expert systems for medical diagnosis become so smart that they reduce the novice physician to the role of data feeder?"* (p. 5). Nevertheless, even in highly structured environments, there remains room to exercise agency. While such environments suggest a particular mode of engagement, they do not *dictate* it [33]. This implies that interactions between novices and expert systems do not inherently limit agency; rather, the impact depends on *how* the system is used.

Today, GenAI has shown the potential to perform many instructional design tasks [34,35], and it has become possible to build GenAI tools customized with instructional design theory, resembling earlier expert systems. Recent research revealed that using GenAI-enabled expert systems in the context of instructional design was associated with higher outcome quality [36]. However, concerns about potential risks to teachers' agency persist. Understanding how expert systems and non-customized GenAI, such as the well-known ChatGPT, affect teachers' agency is important, as this knowledge can inform future directions in developing GenAI-enhanced educational technology to support teachers (both pre-service and in-service) as instructional designers. If teachers' agency is significantly reduced through the use of such tools, this would suggest that the chosen approach may not be suitable for the target group. Accordingly, this study explores how interactions with an expert GenAI bot and a non-customized GenAI bot influence pre-service teachers' agency. The research questions are as follows: (RQ1) "What is the impact of interacting with an expert GenAI bot versus a non-customized GenAI bot on pre-service teachers' agency during design for learning, as indicated by agency-related indicators?"; (RQ2) "What issues concerning pre-service teachers' agency emerged in the open-ended responses?"

2. Conceptual framework

2.1. Professional agency

Agency is a multi-dimensional concept that carries different meanings across and within various fields [3], and it is challenging to define without being overly reductionistic. In this study, we rely on the definition provided by Eteläpelto et al. [37], who describe professional agency as a practice "when professional subjects and/or communities exert influence, make choices, and take stances in ways that affect their work and/or their professional identities" (p. 61). We intentionally chose the concept of *professional agency*, as our main interest lies in the agency of pre-service teachers as professional designers for learning.

Eteläpelto et al. [37] outlined several conditions that shape professional agency from a subject-centered socio-cultural perspective. First, professional agency is practiced when individuals make choices and exert influence on their work and/or professional identities. Second, agency is always purpose-oriented and shaped by socio-cultural and material contexts. Third, the practice of professional agency is closely linked to professional identity, which consists of commitments, ideals, motivations, interests, and goals. Fourth, individuals' personal backgrounds, such as prior knowledge, experience, and skills, serve as affordances for exercising professional agency. Fifth, in the exploration of professional agency, individuals and social entities are analytically separated but considered "mutually constitutive of each other" (p.62). Sixth, individuals maintain discursive, practical, and embodied relations to work, situated within its temporal context. Lastly, professional agency is essential for taking initiative, engaging in professional development, and transforming work practices. Taken together, these conditions present professional agency as a multifaceted concept situated within its sociomaterial context, which is particularly relevant for technology-mediated environments that both provide affordances and constraints to human agency.

In this study, we focus on the professional agency of pre-service teachers during their interactions with different kinds of GenAI bots (expert vs. non-customized) while designing for learning. This contexts provide many opportunities for participants to exert influence on the final outcome and make choices during the process: for example, deciding how long to collaborate with a bot and when to end the interaction, what prompts to write and how to formulate them, whether to generate prompts themselves or simply copy text from the task description. Most importantly, participants determine the nature of their interaction, ranging from more proactive engagement (e.g., providing information to GenAI, giving feedback on its suggestions) to more reactive approaches (e.g., offloading the task and copy-pasting the output). Moreover, the second condition proposed by [37] is particularly relevant to this study, as it highlights that the material context also shapes agency. When pre-service teachers interact with an expert GenAI bot that guides them step-by-step through the design process and

suggests design components, their agency is shaped and constrained by the structure imposed by the system. In contrast, when pre-service teachers interact with a regular ChatGPT, where they decide what kind of support they need, the material setting affords more freedom and fewer constraints. However, as stated in the fourth condition [37], individuals' backgrounds, such as prior knowledge, also serve as affordances and constraints in the exercise of agency. Hence, some pre-service teachers with higher levels of prior knowledge about GenAI or design for learning may exhibit high agency even in the expert GenAI condition, whereas others with less prior knowledge may demonstrate lower agency in the less structured setting.

2.2. Research on teachers' agency and generative AI

The current state of research on teacher's agency in GenAI-mediated environments can be characterized as being in its infancy. As the concept of agency has become a "hot" and widespread topic, it is now sometimes used in studies without explicit definition or theoretical elaboration. For example, Lan and Chen [38], in their paper "*Teachers' agency* in the era of LLM and generative AI: Designing pedagogical AI agents", referred to the concept of "agency" without providing a definition or exploring how teachers' agency might be affected, as their primary focus was on the design of AI agents. Similarly, Zhai [39], in the study "Transforming teachers' roles and *agencies* in the era of generative AI: perceptions, acceptance, knowledge, and practices", proposed a framework of four teacher roles based on their integration of GenAI into teaching: observer, adopter, collaborator, and innovator. However, the question of how teachers' agency changes across these roles was also left implicit, with the focus placed on describing teacher roles that involve more active engagement with AI. Table 1 gives examples of publications where teacher agency is discussed extensively; however, no explicit definition is provided.

Much of the GenAI-in-education research appears to overlook prior studies on agency and educational technology in terms of theoretical framework adoption. Nevertheless, earlier research has much to offer. For instance, Brod et al. [3] explored the concept of agency in educational technology. For them, the central question is how to determine the "right amount" of agency in human-computer interactions. According to the authors, a balance must be found between preserving human agency and ensuring effective outcomes, since young learners, for example, often lack prior knowledge and self-regulation skills and may experience ineffective learning if their agency is not constrained [3]. Therefore, Brod et al. [3] argue that the level of agency should be assigned adaptively, taking users' background characteristics into account.

Conceptual work on human agency in the context of interaction with GenAI is only beginning to emerge. Krakowski [43] examined "human-AI agency in the age of GenAI" and proposed a model of human-AI complementarity, emphasizing that AI can augment rather than replace human capabilities. Focusing on task distribution between humans and GenAI, the model suggests that GenAI can be adopted for problem exploration to catalyze creativity and generate novel outcomes. Frøsig and Romero [44] focused on teachers' agency in GenAI-assisted learning design and proposed that "hybrid intelligence" may serve as a means of preserving human agency in teacher-AI interactions. It is worth noting that both studies suggest that certain forms of human-AI interaction can help safeguard human agency; however, this raises further questions about the meaning of teacher agency and what it entails to maintain it. A recent scoping review [45] on learner and teacher agency in GenAI-enabled environments identified three main themes in the current academic discussion on agency: control in digital spaces, variable engagement and access, and changing notions of agency. Control in digital spaces refers to the redefinition of power and authority relations in GenAI-supported environments. Variable engagement and access highlights that the use of GenAI (or, conversely, the rejection of its use) may widen the digital divide among humans. Changing notions of agency refers to the idea that agency becomes shared between human and non-human actors, raising the question of the desirable level of agency to assign to GenAI.

Empirical work on agency in human-AI interaction is also in its early stages. Currently, most studies on agency focus on indirect indicators of agency, such as users' *perceptions* of agency. For instance, Ortega-Arranz et al. [14] explored how teachers perceived their autonomy, defined closely to agency as "the teachers' willingness, capacity and freedom to take control of their own teaching and learning" (p. 895), when interacting with a GenAI system. They found that the ability to configure the GenAI system according to their preferences resulted in high perceived autonomy. Another study [46] explored how students perceived their agency during human-GenAI collaborative problem-solving. In their research, undergraduate students were tasked with collaborating with ChatGPT to apply computational thinking skills and design a 3D maze; afterwards, they were surveyed on their experienced agency. The study identified three types of collaboration related to agency: human-led (44.30 %), evenly shared contribution (32.91 %), and AI-led (15.19 %). A key concern that emerged was the potential loss of agency during problem-solving, understood as a sense of control over one's own thinking and environment. Didion et al. [47] explored how different characteristics of AI-generated images influenced users' perceived agency using two indicators: temporal binding and magnitude estimation (see 2.4 Indicators of agency). They found that participants' implicit sense of agency remained strong even when the AI produced abstract images, but their explicit judgments of agency were higher for realistic, expected images. Darvishi et al. [48] examined how AI assistance impacts student agency. Unlike the other studies, this research employed objective, process-oriented indicators of agency (see 2.4 Indicators of agency) to examine whether students learn from AI assistance or become dependent on it. They found that while AI support improved the quality of students' peer feedback, its removal led to a decline in quality, indicating reduced student agency.

2.3. Research gaps and the focus of this study

Although agency has become a frequently mentioned concept in the AI-in-education literature [10,11,38,39,43], empirical research on teachers' agency remains limited and tends to examine indirect indicators of agency, such as participants' perceived agency, usually understood as a sense of control or ownership of the outcome. To our knowledge, few studies have employed direct process-related indicators of agency rather than participants' perceptions; only Darvishi et al. [48] used several of them (see 2.4 Indicators of agency), while Didion et al. [47] included one. While perceptions are valuable for understanding how participants experience agency when interacting with AI, studying them also has limitations, as they may differ from actual behavior. In this study, we build on Darvishi et al.'s [48] approach of using direct process-level indicators of agency to empirically examine agency enactment through interaction data. Specifically, we investigate how

Table 1
Examples of works discussing teacher agency without providing an explicit definition.

Publication	Authors
Guidance for generative AI in education and research	UNESCO [40]
The Manifesto for Teaching and Learning in a Time of Generative AI: A Critical Collective Stance to Better Navigate the Future	Bozkurt et al. [10]
Teachers' agency in the era of LLM and generative AI	Lan & Chen [38]
Crafting a Teaching Compass for the Generative AI Age: Nurturing Educator Agency, Wellbeing, and Competencies	Mishra et al. [41]
Educational platformization with Generative AI: impacts on teacher autonomy	Radtke-Bederode & Meireles-Ribeiro [42]

interactions with an expert GenAI bot versus a non-customized GenAI influence pre-service teachers' agency using a combination of observable indicators that signal high or low levels of agency. Additionally, we complement these with qualitative data on pre-service teachers' perceptions of agency to enhance the interpretative depth.

2.4. Indicators of agency

Previous research has employed various indicators to better understand participants' levels of agency. For example, Darvishi et al. [48] used a set of indicators of student agency, including the length of students' comments, the average time spent writing comments, the similarity score, and other measures. The length of students' comments suggested that longer comments were associated with higher agency and greater control over the interaction with AI. The average time spent writing comments was interpreted as a measure of student involvement and engagement, with longer times indicating higher agency [48]. The similarity score referred to the average similarity between a student's comments and their previous comments; a high similarity score suggested that students tend to reuse previous comments and thus demonstrate lower agency. Didion et al. [47] employed two indicators of users' agency: temporal binding, an implicit measure in which shorter perceived delays between action and outcome indicate a stronger sense of control, and magnitude estimation, an explicit measure in which participants evaluate the extent to which they felt they created an image. In the study by Hong et al. [49], teachers' perceived agency was measured with the Professional Agency in Teacher Community (PATC) survey, which includes five dimensions: transformative practices, collective efficacy, positive interdependence, mutual agreements, and active help-seeking. Inspired by the approach of Darvishi et al. [48], we selected three process-level indicators of teachers' agency: length of interaction (Table 2), ownership (Table 2), and collaborative problem-solving (CPS) behaviors (Table 3).

The first indicator, **length of interaction**, was operationalized in two ways: (1a) the total number of *prompts* written by a participant per session and (1b) the total number of *words* written by participant per session. A prompt is a single message written by the teacher during the conversation with the AI, defined by OpenAI as "a text input that initiates a conversation or triggers a response from the model" [50], and may consist of one or multiple utterances. Regarding (1a), in both conditions (expert bot and non-customized GenAI), participants were free to decide when to end the interaction and, consequently, how many prompts they considered sufficient. A higher number of prompts written by participants may indicate a more active role, as it can reflect greater participant engagement. This indicator is supported by literature on collaborative learning, where turn-taking is considered an important marker of active engagement. Turn-taking refers to the alternation of speaking turns during which only one person holds the conversational floor [51]. D'Angelo and Rajarathinam [52] analyzed turn-taking behaviors in small-group work with teaching assistant interventions and found that a student's maximum turn duration can indicate active participation in collaboration. Based on this, we selected (1a) the total number of prompts per session as one of the indicators of high participant agency. However, this measure should be interpreted with caution and cannot serve as the sole indicator of agency. This is because a participant may produce many prompts consisting of only a single word that do not meaningfully contribute to the dialogue (e.g., "ok," "next," "fine"). The second operationalization, (1b) the total number of words written by participant per session, is inspired by [48], who used the length of students' comments as an indicator of high agency. A higher word count may reflect greater effort to elaborate and explain, indicating a more proactive role in the interaction. Wordier prompts are more likely to suggest that participants provided additional information to GenAI or responded meaningfully to its output. However, if a participant copied text from the task description or guiding questions (see Ownership), the word count may be high even though only a small portion was self-written. Therefore, this indicator alone cannot reliably capture agency and should be interpreted in combination with other indicators.

Table 2
Two indicators of pre-service teachers' agency, each with two operationalisations.

	Prompts	Words
(1) Length of interaction	(1a) Total number of prompts written by a participant per session	(1b) Total number of words written by a participant per session
(2) Ownership	(2a) Proportion of self-generated prompts written by a participant per session	(2b) Proportion of self-generated words written by a participant per session

Table 3
Adapted framework of CPS behaviors [58].

	Original CPS behaviors (Sun et al., 2022)	Adapted CPS behaviors	Assumptions about the level of teacher agency
A	Establishing, constructing and maintaining shared knowledge and understanding	Providing information to GenAI to establish common ground, maintaining shared understanding	High
B	Negotiating and coordinating for task completion and problem solving	Giving feedback on GenAI's suggestions, making choices based on its proposals, and monitoring whether it considers previous input	High
C	Maintaining team function and organization	Greeting, confirming, expressing gratitude, encouraging, asking to proceed to the next step, or inquiring about the number of steps	Low
D		Requesting GenAI to perform a task, giving orders	Low

The second indicator, **ownership**, was also operationalized in two ways: (2a) as the proportion of self-generated *prompts* per session and (2b) as the proportion of self-generated *words* per session. This indicator is more qualitative, as it focuses on the nature of participants' input. Participants were provided with a task description, background information about their students, and a script containing nine guiding questions intended to help them stay focused on designing for learning rather than developing materials (e.g., *"What is the overarching aim of this lesson?", "What are the learning objectives?"*). However, some participants primarily copied these questions into the chat with no additional input. While copying utterances may be a pragmatic strategy for interacting with GenAI (which is why this indicator should also be interpreted with caution), a higher proportion of self-written prompts and words reflects a more proactive and agentic approach than simply copying the provided materials, as participants take greater ownership of their input.

The third indicator of agency in this study is the relative frequency of different **collaborative problem-solving (CPS) behaviors** demonstrated by participants. While the CPS literature does not directly link specific CPS behaviors to levels of participant agency, in this research we assume that some CPS behaviors reflect higher agency than others (see below).

CPS is defined as the "capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution" [53]. While "multiple agents" have traditionally been understood as humans, GenAI can also be considered one of the agents involved in CPS [54], and

many previous studies have already focused on CPS in interactions between humans and computer agents [55,56].

The original CPS framework was developed by the OECD [57]. In the adaptation by Sun et al. [58], the CPS framework comprises three categories of behaviors: (1) constructing shared knowledge, (2) negotiation and coordination, and (3) maintaining team function. Buseyne et al. [59] adopted the Sun et al.'s [58] framework to code the utterances of team members during a CPS task in order to explore how these behaviors were distributed across participants. In our study, however, we focused solely on coding the utterances written by pre-service teachers, excluding those generated by GenAI. This is because our focus lies in exploring teachers' agency, not the functioning of GenAI; therefore, the teachers' input is our primary interest.

To adapt Sun et al.'s framework [58] to the context of human-GenAI interaction, we operationalized the CPS behaviors to reflect the dynamics of this interaction (Table 3). Additionally, we identified a fourth behavior specific to human-GenAI interaction, "requesting GenAI to perform a task", which was not present in the original framework. We propose that these four CPS behaviors (Table 3) reflect different levels of pre-service teachers' agency.

Behavior A, "providing information to GenAI to establish common ground and maintain shared understanding", and behavior B, "giving feedback on GenAI's suggestions, making choices based on its proposals, and monitoring whether it considers previous input", are interpreted as high-agency behaviors. These behaviors require participants to proactively provide relevant information and critically evaluate GenAI's output in order to provide meaningful feedback. According to Eteläpelto et al. [37], professional agency is exercised when individuals make choices and exert influence. Demonstrating behaviors A or B involves making various design-related decisions: for example, determining what kind of information is relevant at a given stage, which aspects should be prioritized or omitted, and how the current design component (e.g., a learning objective) could be improved. Moreover, by engaging in these behaviors, participants draw on their own conceptions of what constitutes a high-quality design product and what is necessary to achieve it. This aligns with the fourth condition of professional agency, which highlights the role of individuals' prior knowledge as an affordance for exercising agency.

In contrast, behavior C, "greeting, confirming, expressing gratitude, encouraging, asking to proceed to the next step, or inquiring about the number of steps", and behavior D, "requesting GenAI to perform a task, giving orders", are interpreted as low-agency reactive behaviors. Behavior C involves minimal design-related decision-making. For example, asking GenAI to proceed to the next step or simply responding with "ok" reflects limited engagement and a lower level of contribution compared to behaviors A and B. Behavior D could be viewed as high agency if agency were defined as control over task execution. However, since this study adopts Eteläpelto et al.'s [37] conceptualization of agency as the capacity to make choices and exert influence, we argue that behavior D is more appropriately interpreted as low agency. When participants request GenAI to complete a task, GenAI makes the majority of design decisions, while the user contributes only a brief instruction (e.g., "design a lesson"). However, such requests may also be detailed and carefully prompt-engineered. This is one reason why our analysis of CPS behaviors is conducted at the level of individual utterances (see Method). For example, if a single prompt includes three utterances (one requesting GenAI to perform a task and two specifying the task), the first is coded as behavior D, while the latter two are coded as behavior A.

We acknowledge that these assumptions about participants' agency reflect our own interpretation and may be understood differently under other conceptual frameworks. More research is needed to investigate the relationship between CPS behaviors and human agency.

2.5. Hypotheses

Our general hypothesis is that the expert GenAI would foster higher agency among pre-service teachers, who are novices in designing for learning. Additionally, for each indicator, a corresponding hypothesis was formulated.

Regarding (1) length of interaction, we hypothesized that (1a) interactions in the expert GenAI condition would be longer than those in the ChatGPT condition, as the expert bot follows the steps of a specific instructional design (ID) model and prompts participants for additional input on proposed ID components (see Materials, GenAI support). Furthermore, for (1b), we hypothesized that the expert GenAI condition would be associated with a higher word count. This is based on the assumption that the expert GenAI would foster deeper engagement and reflection during the design process, which would be reflected in more elaborate and wordier participant prompts.

For the (2) ownership indicator, we hypothesized that participants in the expert GenAI condition would demonstrate higher ownership than those in the ChatGPT condition. This expectation stems from the assumption that the expert bot poses structured questions, making it easier for participants to respond without needing to initiate or regulate the interaction themselves. Because this reduces cognitive load, participants may produce more self-generated prompts and words in this condition.

Finally, for the (3) distribution of CPS behaviors, we hypothesized that participants in the expert GenAI condition would exhibit higher proportions of high-agency behaviors (A and B) and lower proportions of low-agency behaviors (C and D) compared to the ChatGPT condition. This expectation is based on the assumption that the expert bot's design would encourage participants to more frequently respond to its questions (behavior A – providing information) and offer meaningful feedback on its suggestions (behavior B).

3. Method

3.1. Research design

This study adopts a between-subjects experimental design to investigate how different types of GenAI support (expert bot vs. non-customized GenAI) influence pre-service teachers' agency. Participants were randomly assigned to one of two conditions: (1) the expert bot condition (called "Co-instructional designer"), in which participants interacted with a GenAI bot pre-customized with one instructional design model to scaffold their design process; and (2) the ChatGPT condition (non-customized GenAI), in which participants interacted with the free version of ChatGPT. The independent variable is the type of GenAI support; the dependent variable is pre-service teachers' agency. All participants completed the same instructional design task and received the same instructions and guiding questions to ensure comparability across the two groups.

3.2. Participants and sampling strategy

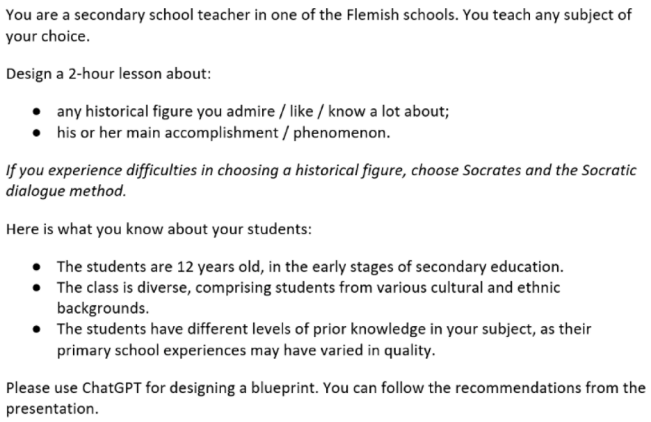
In this study, purposeful sampling [60] was used to target a specific population of pre-service teachers, as they align with the study's aim of examining teachers' agency in interactions with different types of GenAI bots. Pre-service teachers were selected because they are expected to possess teaching-related qualities such as content knowledge, pedagogical knowledge, and technological knowledge required for conducting teaching activities. Many have completed internships in schools, so they are likely familiar with real classroom contexts. Moreover, as our participants were pre-service teachers from Master of Teaching programs, they were expected to possess advanced subject knowledge. According to the curriculum they followed, none had received formal training in instructional design, a situation typical for in-service teachers as well. However, a potential advantage of pre-service teachers compared to in-service teachers is that they may be more open to engaging in instructional design with GenAI, as they are still learning and their teaching practices have not yet become "automatic".

A total of $N = 78$ pre-service teachers participated in the study. To be included, participants had to be enrolled in one of the Master of Teaching programs at a large research university in the center of Europe. According to the pre-test, their mean age was $M = 26.46$ (SD = 6.12, Mdn = 24), 58.97 % were female, and most participants planned to teach behavioral/social sciences (39 %), physics (33 %), mathematics (21 %), or architecture (13 %) at the secondary school level. Most participants had completed two or more teaching internships (34 %), and another 29 % had completed one internship. A similar proportion (32 %) reported gaining teaching experience in contexts other than a formal internship. Only 5 % reported having no teaching internship experience.

In the pre-test, participants were also asked to self-evaluate their instructional design competence, familiarity with ChatGPT, and competence in using ChatGPT (Table 4), as ChatGPT was the most well-known GenAI bot at the time of the experiment. The items were not taken from validated instruments but were specifically developed for this study, as they were intended only to provide descriptive background information about participants and not to serve as outcome measures. The pre-test revealed that participants were generally positive about their design competence and familiarity with ChatGPT but rated their competence in using ChatGPT slightly lower overall (Table 4). We conducted a one-way ANOVA to examine the differences between the conditions (see GenAI support) for each construct and found no significant differences in instructional design competence ($p = .855$), familiarity with ChatGPT ($p = .270$), and competence in using ChatGPT ($p = .910$). This indicates that participants in both conditions obtained roughly similar scores on the three constructs. Noteworthy, although participants self-reported relatively high levels of competence and familiarity regarding instructional design and GenAI, these self-assessments are not objective indicators of actual competence and should be interpreted only as contextual descriptors of the sample. For example, in the open-ended question on the pre-test ("Alignment is very important in designing instruction. What are the elements that need to be aligned?"), 38 % of participants were unable to answer this basic instructional design question. This suggests that their actual knowledge of instructional design was limited despite their relative confidence in it.

3.3. Materials

Instructional design task. Participants were asked to design a learning environment for a two-hour lesson about a historical figure of their choice, focusing on any phenomenon associated with that figure (e.g., Socrates and the Socratic dialogue, Galileo and gravity, Archimedes and his principle of buoyancy) (Fig. 1). Although the task may appear unusual at first glance, it was selected to ensure applicability across different subject domains, as each domain has its own key figures and associated phenomena. The task was also intentionally designed to be open, allowing participants to formulate their own learning goals, objectives, and other instructional design components. In the task description, participants were provided with background information about their students, including age and other relevant characteristics.


You are a secondary school teacher in one of the Flemish schools. You teach any subject of your choice.

Design a 2-hour lesson about:

- any historical figure you admire / like / know a lot about;
- his or her main accomplishment / phenomenon.

If you experience difficulties in choosing a historical figure, choose Socrates and the Socratic dialogue method.

Here is what you know about your students:

- The students are 12 years old, in the early stages of secondary education.
- The class is diverse, comprising students from various cultural and ethnic backgrounds.
- The students have different levels of prior knowledge in your subject, as their primary school experiences may have varied in quality.

Please use ChatGPT for designing a blueprint. You can follow the recommendations from the presentation.

Fig. 1. Experimental task.

Participants were also orally instructed that they could specify additional details about their students (e.g., the number of students) if such information was relevant to their design.

Script with guiding questions. Given that pre-service teachers are novices in instructional design, there was a risk that they might deviate from the intended task of *designing* a learning environment and immediately begin *developing* teaching materials. To mitigate this risk, we developed a script containing nine guiding questions, such as: *"What is the overarching aim of this instruction?", "What specific outcomes do your students have to achieve?"*, and "*Who are the learners, and what do we know about their characteristics, backgrounds, and learning needs?"* (see Appendix 1). These questions were based on the Systems Approach instructional design model [61] and were intended to help pre-service teachers reflect on their design decisions.

GenAI support. Participants completed the experimental task by interacting with one of two types of GenAI support, depending on their assigned condition:

- *ChatGPT condition* – Participants were free to request any type of support from ChatGPT (version GPT-4), as no specific instructions on how to use it were intentionally provided in the task ("Please use ChatGPT for designing a blueprint"). Thus, they could use it as a tool to automate the design process, as a partner to discuss some design decisions, as a tutor who teaches them to design, etc. This type of interaction is participant-led, as it relies on what the participant considers important to accomplish the task.
- *Expert bot condition* – Participants interacted with an expert GenAI bot designed to scaffold the instructional design process. Specifically, the bot was instructed to follow one ID model (Systems approach, [61]) step by step, obtain relevant information from the user needed to design an ID component, generate ID components (e.g., a learning goal), request feedback from the teacher, incorporate it and proceed to the next step of the ID model after user confirmation. In contrast to the participant-led interaction with ChatGPT, the interaction with

Table 4
Pre-test.

Construct	Item	Scale	Results per condition						Condition effect (ANOVA)
			ChatGPT			Expert bot			
			M	Mdn	SD	M	Mdn	SD	
Instructional design competence	How competent do you feel at designing instruction?	1 - not competent, 6 - very competent	4.22	4.00	0.80	4.26	4.00	0.81	$p = .855$
Familiarity with ChatGPT	How familiar are you with ChatGPT?	1 - not familiar, 6 - very familiar	4.07	4.00	1.27	4.43	4.00	0.99	$p = .270$
Competence in using ChatGPT	How competent do you feel using ChatGPT?	1 - not competent, 6 - very competent	4.04	4.00	1.14	4.00	4.00	1.09	$p = .910$

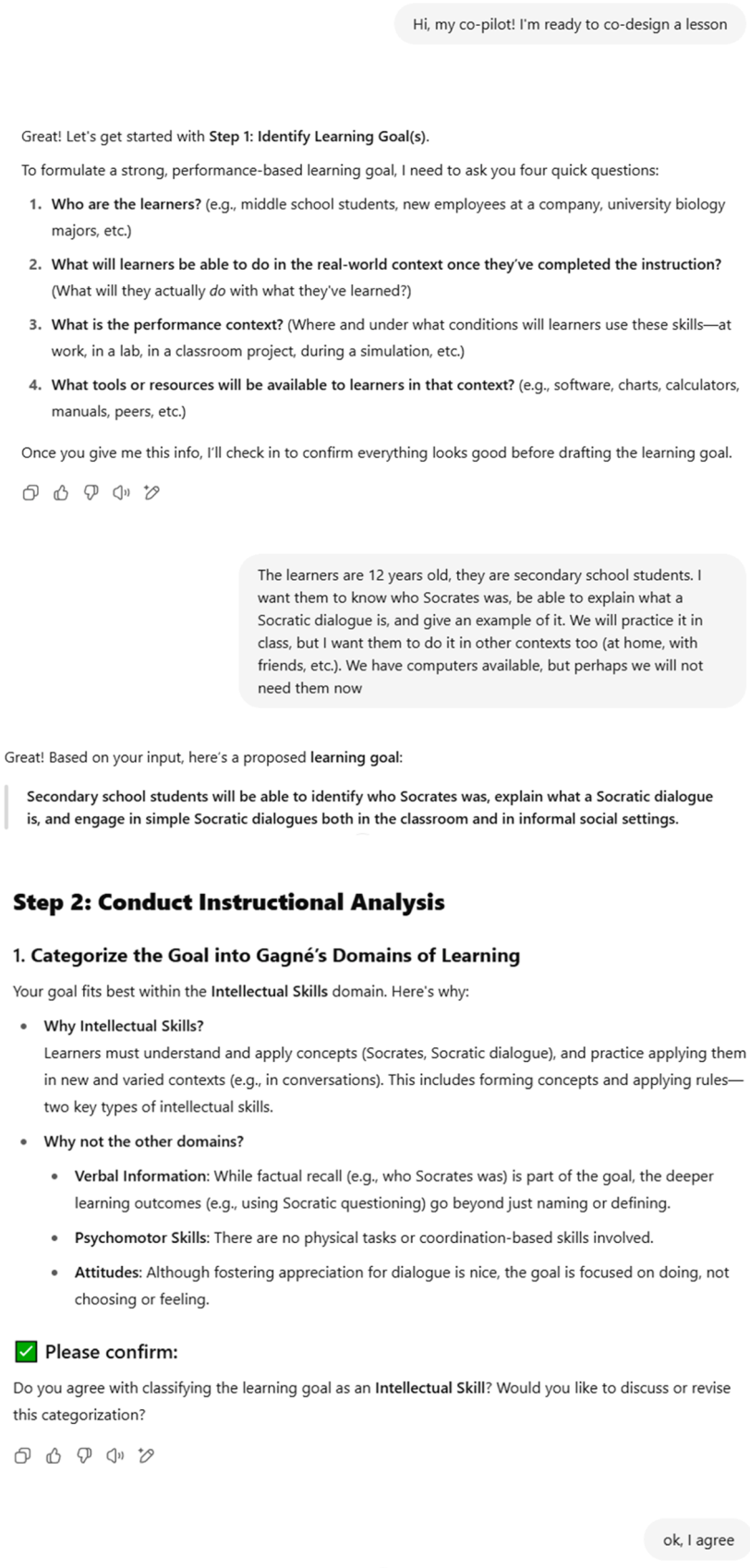


Fig. 2. Example of the start of the interaction with the Co-instructional designer bot.

Great! Now let's break the goal down into the **specific tasks a learner would perform** to achieve it.

Step 2 (continued): Instructional Analysis – Task List

Here are the **key tasks** a student would perform to meet the learning goal:

1. Recognize who Socrates was and describe key facts about his life.
2. Define what a Socratic dialogue is.
3. Identify the purpose and features of Socratic questioning (e.g., open-ended questions, probing assumptions).
4. Analyze a short example of a Socratic dialogue.
5. Generate their own example of a Socratic dialogue based on a chosen topic.
6. Participate in a simple Socratic dialogue in class.
7. Reflect on how Socratic dialogue can be used outside of class.

Now I'll represent these steps as a flowchart.

Fig. 2. (*continued*).

the expert bot was highly scripted. The bot functioned as a more knowledgeable design partner, guiding the teacher through each step of the model. However, because it was still based on the ChatGPT training data and relied on probabilistic output, its suggestions were also prone to inaccuracies. Participants were instructed to comment on its suggestions and provide feedback.

The expert bot is available online: https://chat.openai.com/g/g-VfkELjNBX-co-instructional-designer. The following link provides an example of one interaction: https://chat.openai.com/share/00dd6a92-e9e4–4e9a-b676-be6c93653fd2. Below, we also provide an example of the beginning of one interaction (Fig. 2).

Expert bot development. The expert bot was created through a no-code approach, using only the functionality of ChatGPT Builder (ChatGPT Plus, version GPT-4). A detailed description of the design and development process is provided in [36]. Similar to the architecture of earlier expert systems [31], the bot's development relied on two components: a knowledge base and rules for its application. The knowledge base was a summary of the Systems Approach instructional design model [61]. Fig. 3 presents the steps of the Systems Approach, also known as the Dick and Carey model [61].

The knowledge file uploaded into the ChatGPT Builder summarized information on each step of the model. For every step, the summary included the step name, a description of the expected outcome, two to three examples of well-designed outcomes, and additional remarks, all drawn from [61]. Fig. 4 illustrates this summary for the first step, Identify learning goals.

The rules in the ChatGPT Builder's "Instructions" menu defined how the knowledge base was to be applied (Fig. 5).

The bot was validated by three experts in ID who reviewed its logic and functioning. In addition, pilot tests with pre-service teachers were conducted, which led to refinements aimed at enhancing usability.

3.4. Procedure

Data collection took place as part of the didactic component of the Master of Teaching programs. For the pilot studies, we created a website to attract pre-service teachers from our university. Additionally, we contacted all program directors at the university and obtained approval from several programs (behavioral sciences, architecture, and physics) to conduct the experiment. The data collection was conducted offline and included two pilot sessions, three main data collection days within each program, and one additional online session for participants who wished to take part outside the scheduled sessions.

Before the experiment, participants completed a pre-test survey in which they provided background information and signed informed consent and data management forms. They were then introduced to the experiment, including the study's goal, a definition of instructional

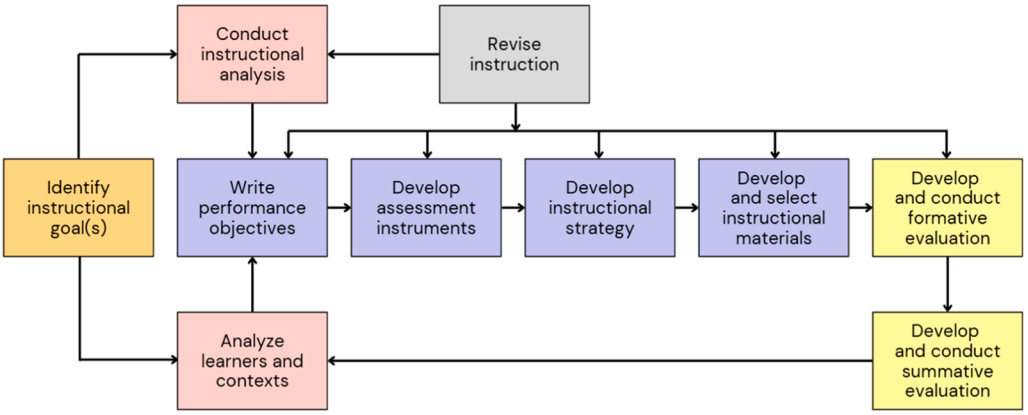


Fig. 3. Systems Approach model (Dick et al., 2015). This image was created by the authors and is also used in [36].

Step 1. Identify Learning Goal(s)

A complete goal statement should describe the following:

- The learners
- What learners will be able to do in the performance context
- The performance context in which the skills will be applied
- The tools that will be available to the learners in the performance context

Example 1: The Acme call center operators will be able to use the Client Helper Support System to provide information to customers who contact the call center.

Example 2: Personnel will demonstrate courteous, friendly behavior while greeting customers, transacting business, and concluding transactions.

Example 3: Students will punctuate a variety of simple sentences using periods, question marks, and exclamation points.

Learning goals must be formulated clearly. A *fuzzy* goal is generally some abstract statement about an internal state of the learner, such as *appreciating, having an awareness of, and sensing*. These kinds of terms often appear in goal statements, but the designer does not know what they mean because there is no indication of what learners would be doing if they achieved this goal.

Fig. 4. Summary of the Systems approach model as the knowledge base for the expert bot.

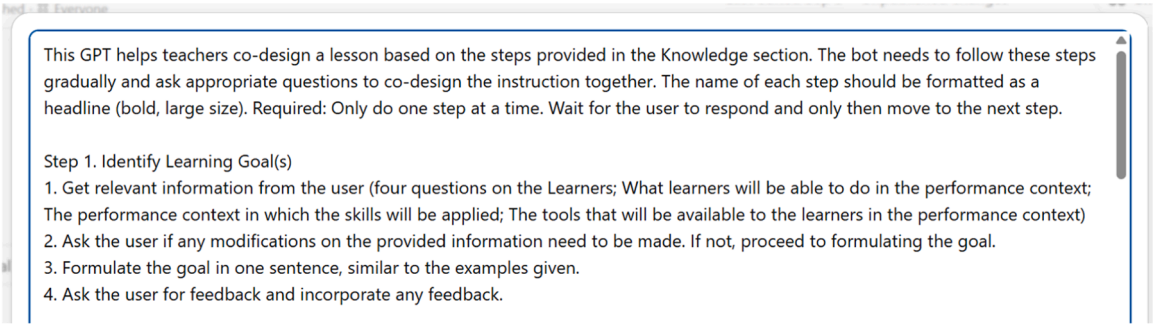


Fig. 5. Rules for using the knowledge base, uploaded to the "Instructions" menu in ChatGPT Builder.

design, and an explanation of why teachers can be considered instructional designers (20 min) (Appendix 3). Next, participants across all conditions were introduced to ChatGPT (20 min) (Appendix 3). Based on OpenAI's prompt engineering guide [62], we demonstrated different strategies for interacting with ChatGPT and techniques for effective prompt engineering. After this, participants were introduced to the experimental task and had the opportunity to ask questions. Participants were then randomly assigned to one of the experimental conditions. In the expert bot condition, future teachers received additional instructions on how to access the bot and interact with it effectively. All participants were given two hours to complete the task, although they were allowed to finish earlier if they wished. The data collected for this study included links to participants' interactions with GenAI and screen recordings of teacher-GenAI interactions. After completing the task, participants were asked to fill out a brief post-experimental survey (5 min). It included two open-ended questions that were important for the current study on teachers' agency: *"In your opinion, does ChatGPT ⟨expert bot⟩ have added value for designing instruction? Why or why not?"* and *"Do you have any other comments?"*

3.5. Ethical considerations

Before collecting any data, we received ethical approval to conduct research involving human participants (G-2023–7051) from the university's Privacy and Ethics Platform and the Social and Societal Ethics Committee.

3.6. Data analysis

Independent variable. Type of GenAI support

This independent variable was manipulated using a between-subjects design, with participants assigned to either the ChatGPT condition or the expert bot condition (see GenAI support).

Dependent variable. Pre-service teachers' agency

Three indicators of teachers' agency were used in this study, all of which were extracted from teacher-GenAI interactions:

- **Length of interaction,** operationalized as (1a) the total number of prompts written by a participant per session, and (1b) the total number of words written by a participant per session. These were calculated by summing the number of prompts (1a) and the number of words (1b).
- **Ownership,** operationalized as (2a) the proportion of self-generated prompts per session, and (2b) the proportion of self-generated words per session. A self-generated prompt was defined as a prompt composed entirely of original content written by the participant, containing no copied utterances. Similarly, self-generated words refer to words within utterances that were not copied and were fully produced by the participant.

- **CPS behaviors**, operationalized as the relative frequency of CPS behaviors (A, B, C, D) among all self-generated utterances written by a participant.

Statistical analysis. To examine the effect of GenAI condition (ChatGPT vs. expert bot) on pre-service teachers' agency, we conducted a one-way Multivariate Analysis of Variance (MANOVA) using the four operationalizations (1a, 1b, 2a, 2b) of two indicators: length of interaction and ownership. A single MANOVA was conducted because these variables were significantly correlated (see Findings) and conceptually represent related dimensions of agency.

Before conducting the MANOVA, assumptions of multivariate normality and homogeneity of covariance matrices were tested. While Box's M test indicated no violation of homogeneity ($p = .062$), the Shapiro-Wilk test revealed a violation of multivariate normality ($p < .001$). Therefore, Pillai's Trace was used as the test statistic, as it is more robust to violations of these assumptions. Following the significant MANOVA result (see Findings), follow-up ANCOVAs were conducted to identify which specific agency indicators differed between the conditions. ANCOVAs were used instead of ANOVAs to control for the influence of instructional design competence, familiarity with ChatGPT, and competence in using ChatGPT as covariates. All statistical analyses were conducted using JASP (version 0.19.2). Because ANCOVAs served as the primary follow-up analyses, we only report ANOVA results when the ANOVA was significant and the corresponding ANCOVA was not. When ANCOVAs were significant, ANOVA results were not reported.

The third indicator, the distribution of CPS behaviors, was excluded from the MANOVA because it is based on categorical variables (proportions of behavior types) rather than continuous variables, as is the case for indicators (1) and (2). To analyze the relationship between condition and the distribution of CPS behaviors across the four categories (A, B, C, D), a separate procedure was applied. The unit of analysis was participants' self-generated utterances. To identify these, we first reviewed the entire teacher-GenAI conversation and flagged teacher utterances that were identical to previous GenAI responses, the task description, or the script with guiding questions (Fig. 6). Copied utterances were excluded as they were not considered illustrative of participants' CPS behaviors.

When scanning teacher utterances to label them as "copied", we typically encountered no problems and applied this label strictly only to utterances that were exactly identical to the text in the task description, guiding questions, or GenAI responses. The only exception occurred when an utterance had the same meaning as the provided script but differed slightly in wording. For instance, the guiding script included the question, "How will *you* measure and evaluate the learners' progress and achievement of the learning objectives?", and one participant asked ChatGPT, "How will *I* measure and evaluate the learners' progress and achievement of the learning objectives?" Because the participant did not add any new content, this utterance was also flagged as "copied", even though it was not entirely identical.

Drawing on the adapted framework of [58], we developed a codebook to guide the coding process of self-generated utterances (Table 5). Each self-generated utterance was coded into one of the four CPS behavior categories. To ensure coding reliability and transparency, two independent raters coded a random 10 % of all teacher-GenAI conversations. Inter-rater reliability was calculated using Cohen's kappa and resulted in a score of $\kappa = 0.70$, indicating substantial agreement [63]. The remaining data was then coded by one rater using the codebook.

To ensure consistency in the coding process, several principles were developed and followed: (1) we first reviewed the entire teacher-GenAI conversation; (2) each self-generated utterance was labeled with only one category; and (3) the categories were operationalized as follows: A – giving new information that had not been discussed before or asking questions aimed at declarative knowledge; B – commenting on what GenAI had proposed, asking for meaningful modifications, or giving feedback; C – utterances aimed at supporting social interaction; D – orders and requests, usually expressed with an imperative verb.

Some conflicting cases were discussed within the research team, and we always took into account the context of interaction to arrive at a decision. For instance, the utterance *"Can you give me a detailed schedule of the two-hour lesson?"* could have been interpreted as "A" (merely asking a question about AI's capabilities), but because GenAI responded with a generated schedule and the question appeared to be more of a request, we labeled it as "D" – requesting GenAI to perform a task. Another example is *"Is there another scientist whose work is or seems somewhat less complicated to twelve-year-old students?"* This utterance was labeled as "B" rather than "A" because it followed a previous discussion in which another historical figure had been mentioned, and the participant therefore asked about someone "less complicated". Finally, we always aimed to stay as close to the original meaning as possible. For example, the utterance *"In your lesson plan, reconsider the evaluation method in order to best capture the progress the students made"* was labeled as "B" rather than "D", since the participant provided feedback indicating that the evaluation method should be changed. In contrast, the utterance *"From your lesson plan, extract a list of activities and brief content that are applicable to it"* was labeled as "D" rather than "B", because it contains a request without offering any meaningful feedback.

For each participant, we calculated the relative frequency (proportion) of each CPS behavior by dividing the number of utterances coded as A, B, C, or D by the total number of self-generated utterances. These individual-level proportions were then aggregated to generate descriptive statistics per experimental condition. Before conducting the analysis, we verified that the assumptions for the test were met. We then conducted a chi-square test of independence to determine whether the distribution of CPS behaviors differed significantly between the expert bot and ChatGPT conditions.

Qualitative analysis of issues concerning agency. To analyze the issues concerning teachers' agency, we conducted a thematic analysis [60] of participants' open-ended responses to two post-experimental survey questions on teachers' agency (see 3.4 Procedure). Using an inductive coding approach [60], we first identified all issues raised by participants relating to their agency during the interaction. The codes were then organized into broader categories to capture overarching themes [60].

4. Findings

Before conducting the MANOVA, we examined the Pearson correlations among the four continuous dependent variables: (1a) total number of prompts, (1b) total number of words, (2a) proportion of self-generated prompts, and (2b) proportion of self-generated words. The results indicated a moderate positive correlation between (1a) total number of prompts and (1b) total number of words ($r = 0.438$, $p < .001$), suggesting that participants who wrote more prompts also tended to write more words overall. A strong positive correlation was found between (2a) proportion of self-generated prompts and (2b) proportion of self-generated words ($r = 0.727$, $p < .001$), indicating that participants who wrote more self-generated prompts also produced more self-generated words. Interestingly, (1b) total number of words was negatively correlated with (2a) proportion of self-generated prompts ($r = –.329$, $p = .014$) and with (2b) proportion of self-generated words ($r = –.631$, $p < .01$). This suggests that participants who produced more words overall were more likely to copy content rather than generate it themselves. No significant correlations were found between (1a) total number of prompts and (2a) proportion of self-generated prompts or between (1a) and (2b) proportion of self-generated words. Since most of the dependent variables were correlated, we proceeded with a one-way MANOVA.

A one-way MANOVA revealed a significant multivariate effect of GenAI condition on agency indicators, Pillai's Trace = 0.262, $F(4, 50) = 4.43$, $p = .004$. Below, we present the findings for each indicator.

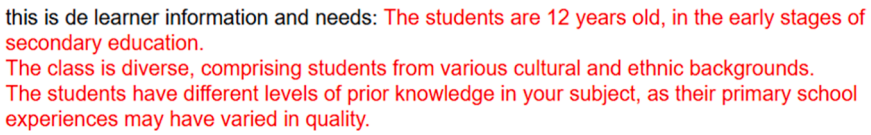


Fig. 6. Example of utterances copied from GenAI output.

Table 5
Codebook for coding participants' self-generated utterances as CPS behaviors.

CPS behavior	Meaning	Example utterances
A	Providing information to GenAI to establish common ground, maintaining shared understanding	• "The lesson will be about Erikson and the 8 stages of development" • "The lesson should last 120 min" • "So by scientist I mean physicist." • "What is the most important invention in physics from Blaise Pascal?" • "What are instructional methods?"
B	Giving feedback on GenAI's suggestions, making choices based on its proposals, and monitoring whether it considers previous input	• "Can you adapt teaching and learning activities to take into account different cultural and ethnic backgrounds of students?" • "Add more accent on Understanding Historical Context" • "Keep this text, but add which scientific principles they should be familiar with, and what historical context you are talking about" • "This is not chronological in time, please put this chronologically in time." • "The introduction on atoms should be more elaborated, since some of the students don't have any prior knowledge on this topic"
C	Greeting, confirming, expressing gratitude, encouraging, asking to proceed to the next step, or inquiring about the number of steps	• "Hi!" • "Allright!" • "Yes I am happy with this" • "It's good" • "Yes" • "No, next step" • "This analysis looks good to me"
D	Requesting GenAI to perform a task, giving orders	• "Can you provide a starting point for a lesson plan around Sigmund Freud for 12-year-olds?" • "Can you suggest an interesting scientist?" • "Can you make it about Marie Curie?" • "Cite your sources" • "Can you provide some interesting reading activity for the class in connection with Anne Frank?"

(1) Length of interaction

We conducted separate ANCOVAs for (1a) total number of prompts written by a participant per session and (1b) total number of words written by a participant per session, with condition (ChatGPT / expert bot) as the fixed factor and instructional design competence, familiarity with ChatGPT, and competence in using ChatGPT as covariates (Type III SS). After adjustment, significant differences were revealed in (1a) total number of prompts, $F(1, 41) = 8.20$, $p_{(holm)} = p_{(bonf)} = 0.007$, $\eta^2 = 0.126$, 95 % CI [.002, 0.325] (Table 6), with participants in the expert bot condition producing more prompts per session. Two covariates were also significant: instructional design competence, $F(1, 41) = 5.44$, $p = .025$, $\eta^2 = 0.084$, 95 % CI [.000, 0.273], and familiarity with ChatGPT, $F(1, 41) = 8.42$, $p = .006$, $\eta^2 = 0.129$, 95 % CI [.003, 0.329], whereas competence in using ChatGPT was not significant, $F(1, 41) = 1.96$, $p = .169$, $\eta^2 = 0.030$, 95 % CI [.000, 0.190]. Regarding (1b) total number of words, ANCOVA revealed no significant difference between conditions, $F(1, 41) = 0.29$, $p = .595$, $\eta^2 = 0.004$, 95 % CI [.000, 0.114] (Table 6). Each covariate showed a significant association with total words: instructional design competence, $F(1, 41) = 5.29$, $p = .027$, $\eta^2 = 0.070$, 95 % CI [.000, 0.255]; familiarity with ChatGPT, $F(1, 41) = 17.77$, $p < .001$, $\eta^2 = 0.237$, 95 % CI [.048, 0.438]; competence in using ChatGPT, $F(1, 41) = 10.75$, $p = .002$, $\eta^2 = 0.143$, 95 % CI [.007, 0.344].

(2) Ownership

A univariate ANOVA revealed a significantly greater (2a) proportion of self-generated prompts per session in the expert bot condition than in the ChatGPT condition, $F(1, 53) = 4.76$, $p = .034$, $\eta^2 = 0.082$ (Table 7). However, an ANCOVA controlling for instructional design competence, familiarity with ChatGPT, and competence in using ChatGPT found no effect of condition, $F(1, 41) = 0.55$, $p_{(holm)} = p_{(bonf)} = 0.461$, $\eta^2 = 0.013$, 95 % CI [.000, 0.150]. One potential explanation is missing data on some covariates, which reduced the analytic sample from 55 to 46 and thus lowered power. None of the covariates were significant (instructional design competence: $F(1, 41) = 0.09$, $p = .763$, $\eta^2 = 0.002$, 95 % CI [.000, 0.101]; familiarity with ChatGPT: $F(1, 41) = 0.08$, $p = .782$, $\eta^2 = 0.002$, 95 % CI [.000, 0.097]; competence in using ChatGPT: $F(1, 41) = 0.81$, $p = .375$, $\eta^2 = 0.019$, 95 % CI [.000, 0.166]).

An ANCOVA controlling for instructional design competence,

Table 7
(2a) Proportion of self-generated prompts per session and (2b) Proportion of self-generated words per session.

Measure	Condition	*M*	*Mdn*	*SD*	*Condition effect (ANCOVA)*
(2a) Proportion of self-generated prompts per session	ChatGPT Expert bot	0.71 0.82	0.71 0.88	0.27 0.15	$F(1, 41) = 0.55$, $p = .461$, $\eta^2 = 0.013$, 95 % CI [.000, 0.150]
(2b) Proportion of self-generated words per session	ChatGPT Expert bot	0.67 0.78	0.72 0.82	0.27 0.18	$F(1, 41) = 0.40$, $p = .533$, $\eta^2 = 0.008$, 95 % CI [.000, 0.135]

Table 6
(1a) Total number of prompts written by a participant per session and (1b) total number of words written by a participant per session.

Measure	Condition	*Total*	*M*	*Mdn*	*SD*	*Condition effect (ANCOVA)*
(1a) Total number of prompts written by a participant per session	ChatGPT	537	18.52	16.00	9.42	$F(1, 41) = 8.20$,
	Expert bot	634	24.38	22.50	7.37	$p = 007$, $\eta^2 = 0.126$, 95 % CI [.002, 0.325]
(1b) Total number of words written by a participant per session	ChatGPT	18,215	628.10	520.00	540.70	$F(1, 41) = 0.29$,
	Expert bot	11,996	461.38	352.50	281.48	$p = .595$, $\eta^2 = 0.004$, 95 % CI [.000, 0.114]

familiarity with ChatGPT, and competence in using ChatGPT found no effect of condition on self-generated words per session, $F(1, 41) = 0.40$, $p_{(holm)} = p_{(bonf)} = 0.533$, $\eta^2 = 0.008$, 95 % CI [.000, 0.135]. The covariates were not significant (instructional design competence: $F(1, 41) = 0.01$, $p = .940$, $\eta^2 \approx 0.000$; familiarity with ChatGPT: $F(1, 41) = 2.05$, $p = .160$, $\eta^2 = 0.043$, 95 % CI [.000, 0.213]; competence in using ChatGPT: $F(1, 41) = 3.81$, $p = .058$, $\eta^2 = 0.081$, 95 % CI [.000, 0.269]).

(3) CPS behaviors

The relative frequencies of CPS behaviors (A, B, C, and D) across the two conditions offered valuable insights into the nature of pre-service teachers' interactions with the GenAI bots (Table 8). In the ChatGPT condition, participants most frequently engaged in behavior A (providing information; 0.45) and behavior D (requesting GenAI to perform a task; 0.34). Behavior B (giving feedback) occurred less frequently (0.20), while behavior C (confirming or asking to proceed) was nearly absent (0.02). In contrast, the expert bot condition revealed a different pattern. The most common behaviors were behavior A (0.38) and behavior C (0.32). Behaviors B and D were comparatively rare, with relative frequencies of 0.16 and 0.14, respectively.

A chi-square test of independence revealed a significant association between condition and CPS behavior, $\chi^2(3, N = 1587) = 247.30$, $p < .001$, Cramér's $V = 0.39$. An analysis of standardized residuals indicated that this effect was primarily driven by behaviors C and D (Table 9). Utterances classified as behavior C were significantly less frequent in the ChatGPT condition (standardized residual = −14.88) and more frequent in the expert bot condition (+14.88). In contrast, utterances classified as behavior D were significantly more frequent in the ChatGPT condition (+8.46) and less frequent in the expert bot condition (−8.46). Residuals for categories A and B remained below ±2.5, suggesting they contributed less to the overall association. Because multiple utterances came from the same participants, we conducted robustness checks using a participant-level permutation test ($p = .0002$) and cluster-robust logistic contrasts, which confirmed lower odds of C vs A in ChatGPT (OR = 0.06, $p < .001$) and higher odds of D vs A (OR = 1.99, $p = .005$); results were unchanged.

4.1. (RQ2) what issues concerning pre-service teachers' agency emerged in the open-ended responses?

Overall, across both conditions, participants were very positive about their interactions with ChatGPT and the expert bot. The main benefits were that the tools served as sources of inspiration, providing ideas participants might not have generated on their own, saved time by speeding up the design process, and supported a more structured approach to lesson design. Nevertheless, several issues related to agency were identified.

4.2. ChatGPT condition

The most frequently mentioned issue was that designing with ChatGPT might add work and make the process longer. While participants were advised to give feedback to ChatGPT during the brief training before the experiment, some shared that providing feedback increased their workload: "*Giving ChatGPT feedback seems like it would take much longer to get a good enough lesson plan*"; "*I now have the feeling that it would take me more time to get a proper answer from ChatGPT than when I would just do it myself*"; "*Sometimes there is a lot of added work just to get the right/useful answer*". A related issue was that ChatGPT sometimes provided too much information: "*Besides that, it's also risky to be swamped by all the information it gives*". Additionally, participants noted that ChatGPT lacked knowledge of the teaching context: "*A drawback to ChatGPT is that it is difficult to get ChatGPT familiar with the real context in which you are operating*". Another concern was that the final product might not align with the participant's own teaching: "*So while I think the tool is useful, it can sometimes force you into creating an instruction that is not entirely your way of teaching*". Moreover, participants shared that using ChatGPT in the design process might lead to overreliance: "*It's a good library of all the different aspects of designing a lesson, but it's also risky to become 'lazy'. Like now, I wasn't too critical about the lesson design at first, and after filling in all the questions above, more questions like time management and so on arose*". Lastly, one participant noted that designing with ChatGPT made the process less enjoyable: "*I do miss the fun of creating my own lesson somehow as well*".

Table 8
Relative frequency of CPS behaviors per condition.

CPS behavior	*ChatGPT Mean (SD)*	*Expert bot Mean (SD)*
A	0.45 (0.24)	0.38 (0.14)
B	0.20 (0.15)	0.16 (0.11)
C	0.02 (0.03)	0.32 (0.17)
D	0.34 (0.22)	0.14 (0.12)

4.3. Expert bot condition

The themes in participants' responses ranged from issues related to the bot's user-friendliness to concerns about their own role as teachers. One issue was that the step-by-step structure of the bot was perceived by participants as too fixed and strict: *"The bot is incredibly valuable, but it felt a bit cumbersome to use it according to its fixed step-by-step plan"; "The only risk I see is that the structure is very predefined. You lose freedom as a teacher, and adding creativity is also more difficult now"*. Another theme was trust, with participants noting the need to approach the bot's ideas critically rather than trusting them blindly: *"I do notice that I quickly trust the bot and don't question its answers. I almost automatically assume that everything it suggests is a good idea. I notice that my own input is minimal"; "It is dangerous to become less critical and blindly trust the bot, which can also make mistakes"; "You have to be critical about it"*. Moreover, participants shared that the bot provided too much information, and they sometimes did not know what to do with it: *"It adds interesting ideas that I sometimes don't come up with myself, but it's also hard to break away from the chatbot's ideas, which might also be a limitation"; "I feel like I'm a bit stuck in the ideas that the bot gives"*. Additionally, participants, similar to those in the ChatGPT condition, noted that the expert bot does not account for their real teaching contexts: *"However, it takes everything human out of the designing process. The bot doesn't know your students, your own teaching style, the school context… The lesson is theoretically 'perfect,' but it stands far away from the actual practice in real life"; "It can be very helpful to create lessons, but the social aspect and the estimation of students' mental state and interests 'in the field' is something AI will not so easily take over"*. One participant explained that she contradicted the bot's suggestions because she wanted to preserve the creative aspect of the learning task: *"Had I followed the bot's suggestion to have students complete the exercise entirely with digital writing support, the creative aspect of typography would have been lost"*. Lastly, another participant emphasized that the roles of humans and AI were not designed optimally with the bot: "*The result of following these steps makes it so that either you let the bot do all the work, or the process becomes so stretched out that the guidance the AI gives you is overkill. I do not realistically see myself going through the process for all my lessons, when the result isn't even a completed lesson and just a blueprint*".

5. Discussion

In this study, we examined how different types of GenAI support (non-customized GenAI vs. expert bot) influence pre-service teachers' agency when designing for learning. Our research questions were (RQ1) "What is the impact of interacting with an expert GenAI bot versus a non-customized GenAI bot on pre-service teachers' agency during

Table 9
Chi-square residual analysis of CPS behaviors per condition.

		A	B	C	D	χ^2 test
ChatGPT	Count	334	167	16	240	
	Expected count	312.912	147.870	126.405	169.812	Pearson's $\chi^2(3, N = 1587) = 247.30$, $p < .001$, Cramér's $V = 0.395$ (moderate–large).
	Standardized residuals	2.152	2.452	−14.877	8.456	
Expert bot	Count	322	143	249	116	
	Expected count	343.088	162.130	138.595	186.188	
	Standardized residuals	−2.152	−2.452	14.877	−8.456	

design for learning, as indicated by agency-related indicators?"; (RQ2) "What issues concerning pre-service teachers' agency emerged in the open-ended responses?"

For RQ1, three indicators were selected to provide insights into pre-service teachers' agency during teacher-GenAI interactions: (1) length of interaction, (2) ownership, and (3) CPS behaviors. Table 10 summarizes the key findings across conditions for RQ1.

(1) Length of interaction

The first indicator, (1) length of interaction, showed that interactions with the expert bot were significantly longer than those with ChatGPT in terms of the total number of prompts (1a). This finding confirms H1a and suggests higher teacher agency in the expert bot condition regarding length of interaction. Within the Eteläpelto et al. [37] framework of professional agency, the dimension emphasizing the role of material contexts appears particularly relevant for understanding this difference. According to this framework [37], agency is always shaped by material contexts. Therefore, one potential explanation of the difference lies in the specific design of the expert bot: by prompting teachers to respond to proposed instructional design (ID) components, it was intended to encourage more extended interactions, thereby fostering a more agentic approach regarding interaction length. In contrast, the material context in the ChatGPT condition was different, as the interaction was participant-driven. While this context was less constrained, it appeared to provide less encouragement for sustaining longer interactions. However, this indicator should be interpreted with caution and cannot serve as the sole indicator of agency, as some prompts consisted of only one or two words (e.g., "ok," "next").

The second operationalization of this indicator, the total number of words written by a participant per session (1b), showed no significant difference between conditions, thereby rejecting H1b. This may be explained, first, by the high standard deviation in the ChatGPT condition (Table 6), indicating substantial within-group variability that could have reduced statistical power. Second, the correlation analysis suggests that participants with higher word counts often copied text from the materials. The absence of significant differences in (1b) therefore suggests that copying behavior occurred in both conditions. Interestingly, descriptive statistics (Table 6) showed non-significant differences between conditions, which may indicate potential effects that could be detected in future studies. Focusing specifically on the number of self-generated words might help reveal differences that are not confounded by copying behavior. Overall, the absence of significant differences in the total number of words (1b) suggests that the material contexts of ChatGPT and the expert bot were not fundamentally different in this respect, as both allowed for substantial copying behavior. Moreover, as the framework [37] highlights that individuals' backgrounds serve as affordances and constraints to exercising agency, participants' prior knowledge, experience, and skills may explain why some chose to write more or fewer words. The ANCOVA results showed that the covariates instructional design competence, familiarity with ChatGPT, and competence in using ChatGPT were all significantly associated with the total number of words. This underscores the role of individual-level factors in shaping interaction length, and future research is necessary to examine these factors in a more systematic manner.

Table 10
Key findings for RQ1.

Indicator	Sub-indicator	ChatGPT	Expert bot
(1) Length of interaction	(1a) Total number of prompts written by a participant per session	Shorter interactions	Longer interactions
	(1b) Total number of words written by a participant per session	-	-
(2) Ownership	(2a) Proportion of self-generated prompts per session	Lower ownership[1]	Higher ownership[1]
	(2b) Proportion of self-generated words per session	-	-
(3) CPS behaviors	Distribution of CPS behavior categories (A, B, C, D)	More task delegation (behavior D)	More confirming, agreeing (behavior C)

[1] Only based on the ANOVA results; the ANCOVA analysis showed no significant differences.

(2) Ownership

The second indicator, (2) ownership, initially revealed that participants in the expert bot condition produced a significantly larger proportion of self-generated prompts than those in the ChatGPT condition (2a), supporting H2a and suggesting higher agency in the expert bot condition. However, this effect disappeared in the ANCOVA when instructional design competence, familiarity with ChatGPT, and competence in using ChatGPT were included as covariates. A potential explanation is missing data on some covariates, which reduced the analytic sample and, consequently, statistical power. None of the covariates were significant predictors of ownership, suggesting that individual differences in prior knowledge or ChatGPT experience did not account for the results. Again, these findings can be interpreted through the lens of agency being shaped by material contexts [37]. Because the expert bot asked targeted questions and guided the design process step by step, the task may have become cognitively less demanding for participants than when they interacted with ChatGPT and had to regulate the design process themselves. As the expert bot assumed the regulatory role, participants' cognitive load [64] may have been reduced, enabling them to produce more self-generated prompts. However, this cognitive load explanation remains speculative and should be examined in future research. The absence of significant differences in the proportion of self-generated words per session (2b) suggests that the material contexts of ChatGPT and the expert bot did not differ substantially in terms of promoting or constraining participants' agency regarding the number of self-generated words, as both allowed for copying behavior.

(3) CPS behaviors

The third indicator, (3) CPS behaviors, revealed unexpected findings in this study. While the distribution of behaviors A (providing information) and B (giving feedback) did not differ significantly between conditions, behavior C (agreeing, confirming) was significantly more

frequent in the expert bot condition. In contrast, behavior D (requesting GenAI to perform a task) was significantly more frequent in the ChatGPT condition. Therefore, H3 was rejected. These findings were surprising, as we expected participants in the expert bot condition to exhibit more behavior B (giving feedback), but instead they demonstrated more behavior C (agreeing, confirming).

As the framework of Eteläpelto et al. [37] states that agency is shaped by material contexts, the differences between the observed behavioral patterns can be explained by the differences in the settings of the expert bot and ChatGPT conditions. A likely reason why participants in the expert bot condition tended to agree with the bot's suggestions rather than engage with them critically is that the expert bot presented ready-made instructional design (ID) components and asked for participant's feedback. This design choice aligns with the conceptualization of expert systems, which aim to guide users through a process. Moreover, because this study focuses on performance support rather than learning support, it is rational and efficient for the bot to generate the initial ID outcomes. However, once an ID component is provided, it might become too overwhelming for participants to contribute meaningfully. Critically evaluating an ID component still requires a certain level of ID competence; this study shows that, without it, participants are more likely to agree with the expert bot's suggestions than to provide meaningful feedback. As for the significantly higher occurrence of the lower-agency behavior D (requesting GenAI to perform a task) in the ChatGPT condition, this is also likely because requesting a task is cognitively less demanding.

5.1. *(RQ2) what issues concerning pre-service teachers' agency emerged in the open-ended responses?*

The open-ended responses revealed themes that were shared across conditions as well as themes unique to the ChatGPT and expert bot conditions.

Regarding the common themes, one recurring issue in both conditions was that participants did not find it efficient to discuss each ID component with GenAI, as it increased the time required to design a lesson. This observation highlights that participants tend to perceive GenAI primarily as an automatic task executor (and, indeed, GenAI was built to execute tasks for the user, but it can also be used in many other ways). Engaging in extended discussions with it appears to contradict this perception. This suggests that participants expect GenAI to perform tasks that optimize the design process. Another common theme was that GenAI lacks knowledge of the actual teaching context and therefore provides suggestions that may be "theoretically perfect" but not adapted to real classroom conditions. This underscores the need for teachers, when co-designing learning environments with GenAI, to assume the role of "context holders". Currently, this is something that AI cannot (yet) do, and it remains part of the teacher's role in this human-AI interaction. Finally, participants in both conditions raised concerns about the risk of becoming less critical when relying on AI to guide the design process. Several participants noted that they quickly began to trust the bot and showed limited critical reflection on its suggestions. Only one participant shared, *"Had I followed the bot's suggestion to have students complete the exercise entirely with digital writing support, the creative aspect of typography would have been lost",* which indicates that she possessed enough knowledge to critically engage with the bot's suggestion. This highlights that the role of individual backgrounds in exercising agency is also very important.

Regarding the issues unique to each condition, participants in the expert bot condition reported a diminished sense of agency due to the fixed step-by-step interaction. This suggests that the usability of the expert bot should be improved: while it currently operates within a single chat window, it would ideally be distributed across multiple chats, allowing teachers to decide when to move between them.

6. Limitations

This study has several limitations that warrant consideration. As noted earlier, the indicators selected for this study should be interpreted with caution. Agency is a complex and multidimensional construct [3], and no single measure can fully capture whether it has been enhanced or diminished in teacher-GenAI interactions. This complexity poses a methodological challenge, which we sought to address by employing multiple indicators of agency. However, important limitations remain in the interpretation of these indicators. First, the (1) length of interaction, although related to agency, may reflect verbosity rather than meaningful engagement. Second, (2) the ownership indicator captures ownership at the level of task execution but may not align with participants' subjective perceptions of ownership. Third, (3) the classification of CPS behaviors as "proactive" or "reactive" reflects our own assumptions and was not originally proposed in [58]. We have attempted to be transparent about this and to clarify the potential pitfalls associated with each indicator. Future research would benefit from moving beyond this approach to develop a more holistic view of agency enactment, for which conceptual clarity on the "professional agency" construct and indicators derived from it will be essential. Another limitation pertains to the study design. Because the results reflect a one-time interaction with GenAI bots, repeated or longitudinal measures might provide richer data on teachers' agency and be sufficient for detecting patterns. A further limitation relates to the sampling procedure: participants were recruited from a single university, which may limit the diversity of the sample. Finally, some uncertainty arose in the coding process, as certain utterances were difficult to categorize in terms of CPS behavior. Nevertheless, we achieved substantial inter-rater reliability ($\kappa = 0.70$), which helps mitigate concerns about divergent interpretations.

7. Conclusion

In this study, we aimed to identify which type of GenAI support (expert vs. non-customized bot) is associated with higher or lower levels of pre-service teachers' agency in the context of instructional design; qualitative data from participants' comments complemented the analysis of directly observed interaction indicators. The findings were mixed. On the one hand, the expert bot condition showed higher agency in (1) length of interaction (1a – total number of prompts written by a participant per session) and (2) ownership (2a – proportion of self-generated prompts per session[1]). On the other hand, the indicator (3) CPS behaviors revealed reductions in agency in both conditions, with participants in the ChatGPT condition tending to ask ChatGPT to perform a task and participants in the expert bot condition tending to agree with the bot's suggestions. The qualitative data also revealed agency-related tensions in both conditions, with the expert bot condition perceived as more limiting due to its step-by-step interaction. Thus, at this point, we are not ready to draw a clear conclusion about which condition is more constraining for pre-service teachers' agency, as both conditions displayed their own unique constraints and affordances. While the expert bot condition showed higher agency in the first two indicators, the third indicator and the qualitative data suggested it was more limiting.

It is interesting to note that the specific ways in which the expert bot and ChatGPT impacted participants' agency differed, supporting the assertion in the framework of Eteläpelto et al. [37] that agency is shaped by material contexts. In future studies, however, a more holistic approach would be valuable, employing a broader set of agency indicators, as findings depend on the choice of indicators. However, identifying multiple indicators derived from the conceptualization of agency presents a methodological challenge. For instance, in this study,

[1] Only based on the ANOVA results; the ANCOVA analysis showed no significant differences.

the framework of Eteläpelto et al. [37] was valuable and functional, but it was insufficient for deriving specific indicators, which is why we had to propose our own indicators, inspired by other studies. Therefore, we urge researchers to contribute to the conceptual work on human agency in human-GenAI interactions and to propose a broader set of agency indicators, both direct (observable and process-oriented) and indirect (teacher perceptions).

Lastly, this study also aimed to provide insights for the future development of GenAI-enabled educational technology. Given the observed reductions in agency across both conditions, the question arises whether it is desirable to use GenAI to support pre-service teachers in instructional design. We argue that it still holds considerable potential. Indeed, our previous research [36] showed that using an expert bot in instructional design was associated with higher outcome quality, whereas such improvements were not observed for the ChatGPT condition. Nevertheless, developers should be transparent about the bot's goals and the implications for teacher agency. The objectives of ensuring effective outcomes through the customization of GenAI with instructional design theory and of keeping humans in the loop may not always align, and one of these priorities may ultimately need to take precedence. This raises the question of the extent to which reductions in agency can be considered acceptable and the conditions under which humans can most effectively benefit from technology in this context.

Our findings do not directly validate or extend the Intelligent-TPACK framework [17]; however, the insights from this study may inform future refinements of the framework, particularly by highlighting the need for additional agency-related knowledge indicators.

8. Implications for teacher training

One important aspect highlighted by this study is that, in order to benefit from powerful GenAI-enabled educational technology in the context of instructional design, teachers still need to possess certain ID competences. These competences might help them critically evaluate GenAI's suggestions. This is why we emphasize that the expert bot presented in this study cannot serve as a standalone solution for transforming teachers into instructional designers; rather, it becomes useful after they receive training in ID and first complete some ID tasks without AI support, after which GenAI can enhance their performance in this context. In addition, teachers also need guidance on how to operate the GenAI system and how to understand and reflect on the expert system's output.

9. Recommendations for future research

This study highlights several directions for future research. First and foremost, more conceptual work is needed to improve the clarity of the "professional agency" construct, ideally to the point where well-defined indicators of agency can be derived. Second, empirical research on human agency in teacher-GenAI interactions remains limited, and the field would benefit from further investigation in this context. We emphasize the need for cumulative research and for explicitly defining agency in each study. A key limitation of current studies is the lack of continuity with prior research on agency and educational technology. Third, given the complexity of the agency construct, we advocate for the use of multiple, combined indicators to provide a more nuanced understanding of agency. Moreover, a more holistic approach that examines patterns in agency, including how they may evolve during teacher-GenAI interactions (e.g. through sequential or temporal analysis of interactions), could provide additional valuable insights. Fourth, the relationship between participants' cognitive load, their level of agency, and the type of GenAI interaction requires further exploration. Fifth, more research is needed to examine how participants' prior knowledge influences their ability to exercise agency in GenAI-supported environments; this could potentially be addressed through additional analytical methods such as regression analysis. Future work might also expand the target group to in-service teachers to explore potential differences between pre-service and in-service contexts. Sixth, while this study focused on pre-service teachers' agency in the context of performance support, future research may also examine agency in the context of learning support, as these two contexts differ substantially and may imply different meanings of agency. Finally, studies may investigate how specific features of expert GenAI bots shape human agency. While expert bots show promise in supporting pre-service teachers during the instructional design process, the implications of their design features need to be more clearly understood.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the primary author used ChatGPT Plus to correct language issues (spelling, grammar, finding synonyms) and to occasionally improve readability. After using the tool, the author reviewed and edited the content as needed and took full responsibility for the final publication.

CRediT authorship contribution statement

Kristina Krushinskaia: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jan Elen:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Annelies Raes:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Not applicable.

Appendix 1. Script with guiding questions

1. **Learning goal**
   What is the overarching aim or purpose of this instruction?
2. **Learning objectives**
   What specific outcomes do your students have to achieve?
3. **Learner information and needs**
   Who are the learners, and what do we know about their characteristics, backgrounds, and learning needs?
4. **Prerequisite requirements**
   What prior knowledge or skills are necessary for learners to be successful in this instruction?
5. **Content and activities**
   What kind of activities are needed to reach the learning objectives? What activities will your students engage in?
6. **Instructional methods**
   By what instructional methods will you support learners?

7. **Assessment**

   How will you measure and evaluate the learners' progress and achievement of the learning objectives?

8. **Feedback**

   How will you provide feedback to learners on their performance and progress?

9. **Media / tools**

   What multimedia tools, materials, and resources will be used to support the instruction?

References

[1] Giddens A. The constitution of society: outline of the theory of structuration. University of California Press; 1984.

[2] Prunkl C. Human autonomy in the age of artificial intelligence. Nat Mach Intell 2022;4(2):99–101. https://doi.org/10.1038/s42256-022-00449-9.

[3] Brod G, Kucirkova N, Shepherd J, et al. Agency in educational technology: interdisciplinary perspectives and implications for learning design. Educ Psychol Rev 2023;35:25. https://doi.org/10.1007/s10648-023-09749-x.

[4] Bandura A. Toward a psychology of human agency. Perspect Psychol Sci 2006;1(2):164–80. https://doi.org/10.1111/j.1745-6916.2006.00011.x.

[5] Schunk DH, Zimmerman BJ, editors. Motivation and self-regulated learning: theory, research, and applications. 1st ed. Routledge; 2008. https://doi.org/10.4324/9780203831076.

[6] Fan Y, Tang L, Le H, Shen K, Tan S, Zhao Y, Shen Y, Li X, Gašević D. Beware of metacognitive laziness: effects of generative artificial intelligence on learning motivation, processes, and performance. Br J Educ Technol 2025;56:489–530. https://doi.org/10.1111/bjet.13544.

[7] Lodge JM, Yang S, Furze L, Dawson P. It's not like a calculator, so what is the relationship between learners and generative artificial intelligence? Learn Res Pract 2023;9(2):117–24. https://doi.org/10.1080/23735082.2023.2261106.

[8] Perrotta C. Pedagogical communication and AI. In: Williamson B, Komljenovic J, Gulson K, editors. World yearbook of education 2024. Taylor & Francis; 2023. p. 54–65. https://doi.org/10.4324/9781003359722-5.

[9] Creely E, Blannin J. Creative partnerships with generative AI: possibilities for education and beyond. Think Skills Creat 2025;56:101727. https://doi.org/10.1016/j.tsc.2024.101727.

[10] Bozkurt A, Xiao J, Farrow R, Bai JYH, Nerantzi C, Moore S, Asino TI. The manifesto for teaching and learning in a time of generative AI: a critical collective stance to better navigate the future. Open Praxis 2024;16(4):487–513. https://doi.org/10.55982/openpraxis.16.4.777.

[11] Holmes W, Tuomi I. State of the art and practice in AI in education. Eur J Educ 2022;57:542–70. https://doi.org/10.1111/ejed.12533.

[12] Lepage A, Collin S. Preserving teacher and student Agency: insights from a literature review. In: Urmeneta A, Romero M, editors. Creative applications of artificial intelligence in education. Palgrave studies in creativity and culture. Cham: Palgrave Macmillan; 2024. https://doi.org/10.1007/978-3-031-55272-4_2.

[13] Luckin R. Nurturing human intelligence in the age of AI: rethinking education for the future. Dev Learn Organiz Int J 2024;39(1):1–4. https://doi.org/10.1108/dlo-04-2024-0108.

[14] Ortega-Arranz A, Topali P, Molenaar I. Configuring and monitoring students' interactions with generative AI tools: supporting teacher autonomy. In: Proceedings of the 15th international learning analytics and knowledge conference (LAK '25). ACM; 2025. p. 895–902. https://doi.org/10.1145/3706468.3706533.

[15] Sethuraman KR. Artificial intelligence and education: preserving human agency in a world of automation. SBV J Basic Clin Appl Health Sci 2025;8(2):41–2. https://doi.org/10.4103/SBVJ.SBVJ_19_25.

[16] UNESCO. AI competency framework for teachers. United Nations Educational, Scientific and Cultural Organization; 2024. https://doi.org/10.54675/ZJTE2084.

[17] Celik I. Towards intelligent-TPACK: an empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. Comput Human Behav 2023;138:107468. https://doi.org/10.1016/j.chb.2022.107468.

[18] Mishra P, Koehler MJ. Technological pedagogical content knowledge: a framework for teacher knowledge. Teach Coll Rec 2006;108(6):1017–54. https://doi.org/10.1111/j.1467-9620.2006.00684.x.

[19] Goodyear P, Dimitriadis Y. In medias res: reframing design for learning. Res Learn Technol 2013;21. https://doi.org/10.3402/rlt.v21i0.19909.

[20] Elen J, Depaepe F. Education and technology: elements of a relevant, comprehensive, and cumulative research agenda. Educ Technol Res Dev 2024. https://doi.org/10.1007/s11423-024-10420-7.

[21] Asensio-Pérez, J.I., Dimitriadis, Y., Pozzi, F., Hernández-Leo, D., Prieto, L.P., Persico, D., & Villagrá-Sobrino, S.L. (2017). Towards teaching as design: exploring the interplay between full-lifecycle learning design tooling and teacher professional development. "Computers and Education, 114″, 92–116. 10.1016/j.compedu.2017.06.011.

[22] Goodyear P. Teaching as design. HERDSA Rev High Educ 2015;2:27–50. https://herdsa.org.au/herdsa-review-higher-education-vol-2/27-50.

[23] Jordan ME. Teaching as designing: preparing pre-service teachers for adaptive teaching. Theory Pract 2016;55(3):197–206. https://doi.org/10.1080/00405841.2016.1176812.

[24] Laurillard D. Teaching as a design science. Routledge; 2012. https://doi.org/10.4324/9780203125083.

[25] Bennett S, Agostinho S, Lockyer L. The process of designing for learning: understanding university teachers' design work. Educ Technol Res Dev 2017;65(1):125–45. https://doi.org/10.1007/s11423-016-9469-y.

[26] Lee D. The process of designing for [online] learning: response to Bennett et al. Educ Technol Res Dev 2021;69(1):289–93. https://doi.org/10.1007/s11423-020-09915-w.

[27] Prieto LP, Dimitriadis Y, Craft B, Derntl M, Émin V, Katsamani M, Laurillard D, Masterman E, Retalis S, Villasclaras E. Learning design Rashomon II: exploring one lesson through multiple tools. Res Learn Technol 2013;21. https://doi.org/10.3402/rlt.v21i0.20057.

[28] Seel NM, Lehmann T, Blumschein P, Podolskiy OA. Instructional design for learning: theoretical foundations. SensePublishers; 2017. https://doi.org/10.1007/978-94-6300-941-6.

[29] Cline RW, Merrill MD. Automated instructional design via instructional transactions. In: Tennyson RD, Barron AE, editors. Automating instructional design: computer-based development and delivery tools. Automating instructional design: computer-based development and delivery tools, 140. Springer; 1995. p. 317–53. https://doi.org/10.1007/978-3-642-57821-2_13. NATO ASI Series.

[30] Spector JM, Song D. Automated instructional design advising. In: Tennyson RD, Barron AE, editors. Automating instructional design: computer-based development and delivery tools. Springer; 1995. p. 377–402. https://doi.org/10.1007/978-3-642-57821-2_15.

[31] Seel NM, Eichenwald LD, Penterman NF. Automating decision support in instructional system development: the case of delivery systems. In: Tennyson RD, Barron AE, editors. Automating instructional design: computer-based development and delivery tools. Automating instructional design: computer-based development and delivery tools, 140. Springer; 1995. p. 121–34. https://doi.org/10.1007/978-3-642-57821-2_8.

[32] Salomon G, Perkins DN, Globerson T. Partners in cognition: extending human intelligence with intelligent technologies. Educ Res 1991;20:2–9. https://doi.org/10.3102/0013189X020003002.

[33] Chanowitz B, Langer EJ. Knowing more (or less) than you can show: understanding control through the mindlessness-mindfulness distinction. In: Garber J, Seligman MEP, editors. Human helplessness: theory and application. Academic Press; 1980. p. 97–129.

[34] Choi, G.W., Kim, S.H., Lee, D., & Moon, J. (2024). Utilizing generative AI for instructional design: exploring strengths, weaknesses, opportunities, and threats. TechTrends. 10.1007/s11528-024-00967-w.

[35] Luo, T., Muljana, P.S., Ren, X., Kim, Y., & Xie, K. (2024). Exploring instructional designers' utilization and perspectives on generative AI tools: a mixed methods study. Educational technology research and development. 10.1007/s11423-024-10437-y.

[36] Krushinskaia, K., Elen, J., Raes, A. (under review). Revisiting expert systems for instructional design: An experimental study on the potential of generative AI for novices in instructional design. Manuscript under review.

[37] Eteläpelto A, Vähäsantanen K, Hökkä P, Paloniemi S. What is agency? Conceptualizing professional agency at work. Educ Res Rev 2013;10:45–65. https://doi.org/10.1016/j.edurev.2013.05.001.

[38] Lan Y-J, Chen N-S. Teachers' agency in the era of LLM and generative AI: designing pedagogical AI agents. Educ Technol Soc 2024;27(1). https://doi.org/10.30191/ETS.202401_27(1).PP01.I–XVIII.

[39] Zhai X. Transforming teachers' roles and agencies in the era of generative AI: perceptions, acceptance, knowledge, and practices. J Sci Educ Technol 2024. https://doi.org/10.1007/s10956-024-10174-0.

[40] United Nations Educational, Scientific and Cultural Organization (UNESCO). (2023). Guidance for generative AI in education and research. 10.54675/EWZM9535.

[41] Mishra P, McCaleb L, Oster N. Crafting a teaching compass for the generative AI age: nurturing educator agency, wellbeing, and competencies (OECD future of education and skills 2030 global forum report). OECD; 2024. https://www.oecd.org/content/dam/oecd/en/about/projects/edu/education-2040/global-forum/7th-global-forum/OECD%20Teaching%20Compass%20and%20Generative%20AI.pdf.

[42] Radtke-Bederode I, Meireles-Ribeiro LO. Educational platformization with Generative AI: impacts on teacher autonomy. Alteridad 2025;20(2):173–83. https://doi.org/10.17163/alt.v20n2.2025.02.

[43] Krakowski S. Human-AI agency in the age of generative AI. Inf Organiz 2025;35(1):100560. https://doi.org/10.1016/j.infoandorg.2025.100560.

[44] Frøsig, T., & Romero, M. (2024). Teacher agency in the age of generative AI: towards a framework of hybrid intelligence for learning design. arXiv. 10.48550/arXiv.2407.06655.

[45] Roe, J., & Perkins, M. (2024). Generative AI and agency in education: a critical scoping review and thematic analysis. arXiv. 10.48550/arXiv.2411.00631.

[46] Zhu, G., Sudarshan, V., Kow, J.F., & Ong, Y.S. (2024). Human-generative AI collaborative problem solving: who leads and how students perceive the interactions. arXiv. https://arxiv.org/abs/2405.13048.

[47] Didion JK, Wolski K, Wittchen D, Coyle D, Leimkühler T, Strohmeier P. Who did it? How user agency is influenced by visual properties of generated images. In: Proceedings of the 37th annual ACM symposium on user interface software and technology (UIST '24). Association for Computing Machinery; 2024. p. 1–17. https://doi.org/10.1145/3654777.3676335. Article 94.

[48] Darvishi A, Khosravi H, Sadiq S, Gašević D, Siemens G. Impact of AI assistance on student agency. Comput Educ 2024;210:104967. https://doi.org/10.1016/j.compedu.2023.104967.

[49] Hong H-Y, Chen M-J, Chang C-H, Tseng L-T, Chai CS. AI-supported idea-developing discourse to foster professional agency within teacher communities for STEAM lesson design in a knowledge building environment. Comput Educ 2025;229:105241. https://doi.org/10.1016/j.compedu.2025.105241.
[50] OpenAI. (2024). Prompt engineering best practices for ChatGPT. OpenAI Help Center. https://help.openai.com/en/articles/10032626-prompt-engineering-best-practices-for-chatgpt.
[51] Hu L, Chen G. Exploring turn-taking patterns during dialogic collaborative problem solving. Instr Sci 2022;50(1):63–88. https://doi.org/10.1007/s11251-021-09565-2.
[52] D'Angelo CM, Rajarathinam RJ. Speech analysis of teaching assistant interventions in small group collaborative problem solving with undergraduate engineering students. Br J Educ Technol 2024;55(4):1583–601. https://doi.org/10.1111/bjet.13449.
[53] Csapó B, Funke J. The nature of problem solving: using research to inspire 21st century learning. Educational research and innovation. Paris: OECD Publishing; 2017. https://doi.org/10.1787/9789264273955-en.
[54] Brandl, L., Richters, C., Kolb, N.& Stadler, M. (2025). Can generative artificial intelligence ever be a true collaborator? Rethinking the nature of collaborative problem-solving.
[55] Howard C, Jordan P, Di Eugenio B, Katz S. Shifting the load: a peer dialogue agent that encourages its human collaborator to contribute more to problem solving. Int J Artif Intell Educ 2017;27(1):101–29. https://doi.org/10.1007/s40593-015-0071-y.
[56] Graesser AC, Cai Z, Morgan B, Wang L. Assessment with computer agents that engage in conversational dialogues and trialogues with learners. Comput Human Behav 2017;76:607–16. https://doi.org/10.1016/J.CHB.2017.03.041.
[57] OECD. (2017). PISA 2015 collaborative problem-solving framework. https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf.
[58] Sun C, Shute VJ, Stewart AEB, Beck-White Q, Reinhardt CR, Zhou G, Duran N, D'Mello SK. The relationship between collaborative problem solving behaviors and solution outcomes in a game-based learning environment. Comput Human Behav 2022;128. https://doi.org/10.1016/j.chb.2021.107120.
[59] Buseyne S, Rajagopal K, Danquigny T, Depaepe F, Heutte J, Raes A. Assessing verbal interaction of adult learners in computer-supported collaborative problem solving. Br J Educ Technol 2024;55:1465–85. https://doi.org/10.1111/bjet.13391.
[60] Ravitch SM, Carl NM. Qualitative research: bridging the conceptual, theoretical, and methodological. 2nd ed. SAGE Publications; 2016.
[61] Dick W, Carey L, Carey JO. The systematic design of instruction. 6th ed. Pearson; 2015.
[62] OpenAI Six strategies for getting better results. OpenAI 2025. https://platform.openai.com/docs/guides/prompt-engineering/six-strategies-for-getting-better-results.
[63] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.
[64] Sweller J. Cognitive load theory. In: Mestre JP, Ross BH, editors. Psychology of learning and motivation. Psychology of learning and motivation, 55. Academic Press; 2011. p. 37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8.