Educational Psychology Review (2026) 38:45
https://doi.org/10.1007/s10648-026-10118-7

INTERVENTION STUDY

Think First, ChatGPT Later: Guiding Human–AI Collaboration for Learning Gains in Independent Human Creativity

Sarah Shi Hui Wong[1] · **Sophia Xuefei Qiu**[2]

Received: 1 August 2025 / Accepted: 8 January 2026

Abstract

Generative artificial intelligence (AI) tools such as ChatGPT can boost creative performance, but do these boosts translate into learning gains? This study examined whether the benefits of ChatGPT for creativity persist even when its assistance is removed, and how people can effectively use ChatGPT to enhance their learning and independent creativity. University students ($N = 196$) solved a creative product improvement task either independently (human-only group) or using ChatGPT freely (general-AI group) or using ChatGPT in a guided way (regulated-AI group). Specifically, the regulated-AI group used a novel "think first, ChatGPT later" approach—they first generated their own ideas, then collaborated with ChatGPT to improve, develop, and evaluate them. Thereafter, all groups independently solved a creative product invention task. On the first task, the general-AI group produced more creative solutions than the human-only and regulated-AI groups. But without ChatGPT assistance on the second task, the general-AI group's creativity declined to levels comparable to the human-only group. In striking contrast, despite a lack of performance gains on the first task, the regulated-AI group outperformed both the human-only and general-AI groups in independent creativity on the second task. Process analyses revealed that the general-AI group most often simply dictated ChatGPT to directly generate the solutions. Conversely, the regulated-AI group more frequently collaborated with ChatGPT to improve their self-generated ideas, in turn mediating their later advantage over the general-AI group in independent originality. Thinking of one's own ideas first, then collaborating with ChatGPT to improve them, promotes learning gains in independent human creativity.

Keywords Human–AI interaction · Generative AI · ChatGPT · Creativity · Creative problem-solving · Co-creation · Metacognition

Extended author information available on the last page of the article

Published online: 15 April 2026

Introduction

The advance of generative artificial intelligence (AI) technologies, tools, and chatbots such as ChatGPT has sparked significant interest worldwide in their impact on human cognition and learning. Developed by OpenAI for conversational usage, ChatGPT is a generative AI-based large language model (LLM) that has been trained with vast amounts of data through deep learning techniques to mimic human language processing. Indeed, LLMs such as ChatGPT have been found to pass the Turing test in generating human-like text and content that are statistically indistinguishable from those by humans (Köbis & Mossink, 2021; Mei et al., 2024).

According to Bauer et al.'s (2025) ISAR model, AI can exert four types of effects on learning: *inversion* (i.e., when AI leads to reduced cognitive processing and learning), *substitution* (i.e., when AI provides instructional equivalence to non-AI alternatives without changing cognitive processing or learning), *augmentation* (i.e., when AI enhances instruction by providing additional learning support without necessarily deepening cognitive processing), and *redefinition* (i.e., when AI fosters deep constructive or interactive processes to enhance learning). Thus, redefinition effects may be key for harnessing the transformative potential of AI in education, where AI-enhanced learning activities encourage deep rather than shallow processing in learners (Bauer et al., 2025).

A growing body of research has found that using ChatGPT can promote educationally valued, complex cognitive outcomes such as critical thinking (Deng et al., 2025; Suriano et al., 2025) and, of particular interest to this study, creativity (Lee & Chung, 2024; Urban et al., 2024). Although these findings offer promising opportunities to boost human performance, they also raise important questions: To what extent do people learn deeply when spontaneously using ChatGPT to solve creative tasks? Whereas ChatGPT assistance can increase creative *performance* on the task at hand, do such benefits persist for human *learning* and independent creativity when ChatGPT assistance is subsequently removed? If not, how can people be guided to interact more effectively with ChatGPT for sustained gains in their learning and creativity? The present study seeks to answer these pressing questions.

Creativity and Creative Problem-Solving

What is "creativity"? Based on the "Four Ps" model that has gained wide consensus in creativity research (Plucker et al., 2004; Rhodes, 1961; Runco, 2004), creativity is the interaction among the creative person, creative process, and creative press (or environment) that results in a tangible creative product that is original (or novel) and useful (or effective). Thus, originality and usefulness are both central to creativity (Runco & Jaeger, 2012; see also Kaufman & Glăveanu, 2019), although the rise of AI technologies has stimulated discourse on recalibrating definitions of creativity to also account for AI's "artificial creativity" that lacks the authenticity and intentionality in human creativity (Runco, 2025).

Creativity tends to occur when solving ill-defined problems, particularly those that cannot be solved through rote application of past experience and that may have multiple potentially viable paths to multiple different end states or solutions (Sawyer &

Henriksen, 2023). The production of creative products or solutions is then facilitated by the creative process (Lubart, 2001), which has been viewed as a dynamic interplay between divergent and convergent thinking (Guilford, 1967b; Rawlings et al., 2025). Whereas divergent thinking involves exploring and generating multiple ideas (Acar & Runco, 2019; Runco & Acar, 2012), convergent thinking involves evaluating those ideas and selecting a single best solution (Cropley, 2006). When solving ill-defined problems, people may thus iteratively engage in both divergent and convergent thinking during the various stages of the creative process in generating, developing, evaluating, and selecting ideas that are original and useful (Mumford et al., 1991, 2002; Vernon et al., 2016).

In line with the view of creativity as a process, people can be taught to think more creatively (Plucker et al., 2004; Rhodes, 1961). Much research has shown that creative thinking can be developed, particularly through interventions that train the requisite cognitive skills and heuristics for skill application (Scott et al., 2004). For instance, a single brief training session that guides people to apply creative thinking skills (e.g., brainstorming and developing ideas, making connections among ideas) can improve post-training creative performance (Clapham, 1997; Ritter & Mostert, 2017).

Generative AI and Creativity

Although creativity has long been revered as the hallmark of uniquely human cognition and expression, generative AI has challenged this bastion (Garcia, 2025). Studies comparing ideas generated by humans versus generative AI have found few differences in their originality and usefulness (Gilhooly, 2024; Haase & Hanel, 2023) or even more creative ideas from generative AI (Hubert et al., 2024), suggesting that generative AI has risen to at least human-level creativity or even surpassed it. Moreover, recent studies have found that solving creative tasks with ChatGPT improves creative performance (e.g., Lee & Chung, 2024; Li, 2025; Urban et al., 2024). For instance, Urban et al. (2024) had university students solve a product improvement task—a complex, ill-defined creative problem that involved improving an ordinary product such as a stuffed bunny—either independently (human-only group) or by freely using ChatGPT for assistance (AI group). The AI group produced solutions that were more original, useful, and elaborate than those of the human-only group, demonstrating that ChatGPT can enhance creative performance.

Likewise, Lee and Chung (2024) had participants generate creative ideas for everyday objects (e.g., paperclip, garden hose) or invent product ideas (e.g., a dining table that does not exist in the market). Participants generated more creative ideas when they used ChatGPT for assistance, relative to solving the task independently or using Google for assistance. Presumably, ChatGPT goes beyond traditional web search engines such as Google to synthesize remotely related knowledge into coherent and articulate responses rather than simply listing multiple relevant pages for users to navigate and synthesize, thereby supporting users to combine disparate ideas for more creative solutions (Lee & Chung, 2024).

Unresolved Questions

Notwithstanding the allure of using ChatGPT for better creativity, at least three unresolved yet critical issues emerge. We discuss each of these issues in turn below.

Does Using ChatGPT in Creative Tasks Improve Learning or Mere Performance?

First, does using ChatGPT boost one's deep learning and creative thinking skills, or does it simply inflate one's creative performance when assisted by the tool? Here, it is vital to differentiate between *performance* (i.e., temporary fluctuations in behavior or knowledge) and *learning* (i.e., longer-term changes in behavior or knowledge; Soderstrom & Bjork, 2015). Although this distinction is fundamental, it has often been neglected in extant research on generative AI in education (Yan et al., 2025). For instance, a recent meta-analysis of 62 studies (Deng et al., 2025) aimed to examine whether ChatGPT enhances "student learning", yet reported that ChatGPT improves "academic performance". Reflecting the common conflation of performance versus learning, the meta-analysis included studies that examined students' performance on the ChatGPT-assisted task (e.g., Stadler et al., 2024; Urban et al., 2024) or even students' self-reported perceptions of their thinking skills or dispositions (e.g., Essel et al., 2024; Yilmaz & Karaoglan Yilmaz, 2023). Unsurprisingly, there was significant heterogeneity among the meta-analyzed studies.

Indeed, whereas extant research has focused on the benefits of ChatGPT assistance for learners' immediate creative performance on the *assisted* task (e.g., Lee & Chung, 2024; Li, 2025; Urban et al., 2024), few—if any—studies have examined whether these creative benefits persist even on subsequent *unassisted* tasks when access to ChatGPT is removed. Yet, learners' unassisted creative task performance could reveal whether their prior interactions with ChatGPT have yielded deep learning that supports independent human creativity. Arguably, developing such creative thinking in learners is a common, fundamental goal of many educational systems around the world (Strom & Strom, 2002). If learners excel on creative tasks only when assisted by ChatGPT but not on subsequent unassisted creative tasks, this could potentially reflect appropriation of ChatGPT's responses rather than learning (Schwartz & Lin, 2001).

This issue aligns with broader concerns about the potential pitfalls of overreliance on generative AI (Nah et al., 2023), although similar concerns have loomed even before its advent. Notably, Bainbridge (1983) cautioned about the "ironies of automation" in which overreliance on automated systems may expand rather than eliminate problems with the human operator, particularly in their development of cognitive skills such as long-term knowledge and monitoring (see also Wong & Lim, 2023). Such cautionary note has been echoed in the context of generative AI use (Huber et al., 2024; Kasneci et al., 2023), which has been viewed as potentially harmful when it eliminates the productive struggle that lies at the heart of learning (Rus & Kendeou, 2025; see also Bjork, 1994 for a discussion of "desirable difficulties").

Indeed, such concerns are not unfounded. Some evidence suggests that learners tend to rely on, rather than learn from, generative AI tools such as ChatGPT (Darvishi et al., 2024; Fan et al., 2025). For instance, Fan et al. (2025) found that students who completed an essay-writing task with ChatGPT showed greater improvement in their essay scores on the assisted task, relative to students who completed the task independently or with assistance from a human expert or with the support of writing analytics tools. However, the ChatGPT group did not display better knowledge or transfer of information related to the essay topic than the other groups, suggesting that the students had simply appropriated the content generated by ChatGPT in the essay-writing task without developing a deeper understanding. In another study, Darvishi et al. (2024) provided students with AI prompts for four weeks during peer review tasks. In the next four weeks, these prompts were removed for some students but not others. Students for whom AI assistance was removed then produced lower-quality peer reviews than those who continued to receive AI assistance, suggesting that students tended to rely on AI assistance to maintain the quality of their work instead of learning from it.

How Do Learners Use ChatGPT To Solve Creative Tasks?

This brings us to the second critical issue: Whereas extant research has explored the effects of ChatGPT on creative performance, fewer studies have examined *how* exactly learners interact with ChatGPT to assist them with the creative task, with some exceptions (Boers et al., 2025; Urban et al., 2025). For instance, Boers et al. (2025) examined transcripts of students' conversations with ChatGPT when solving creative tasks such as generating different uses for a common object (e.g., brick) and generating ideas for a societal problem (e.g., improving the use of public trains). Process analyses revealed that most of the students' expressed ideas were produced by literally repeating (or copying) ChatGPT's ideas; seldom did students combine their own self-generated ideas with ChatGPT's ideas. Thus, students often do not spontaneously interact with ChatGPT in constructive and interactive ways—generating ideas beyond the provided material, and collaboratively or mutually generating ideas with frequent turn-taking (Chi & Wylie, 2014)—that would enhance their deep learning for redefinition effects of AI (Bauer et al., 2025).

Instead, students tend to interact with ChatGPT in merely active ways by overtly repeating or manipulating its output, which tends to yield a shallow understanding (Chi & Wylie, 2014). Yet, most students (93%) in Boers et al.'s (2025) study reported that they were most frequently inspired by ChatGPT's ideas to come up with new ideas, revealing a lack of metacognitive awareness about their actual creative problem-solving processes.

How Can Learners Be Guided To Collaborate with ChatGPT in Creative Tasks for Learning Gains?

Taken together, extant research suggests that learners tend to rely on ChatGPT instead of using it in ways that enhance their learning. Consequently, although using ChatGPT can boost creative performance on the assisted task (Lee & Chung, 2024; Li, 2025; Urban et al., 2024), such performance boosts may not translate to learning gains for independent human creativity on subsequent unassisted tasks. Thus, the third critical issue is: How can we guide learners to interact more collaboratively with ChatGPT during creative problem-solving to support their deep learning?

Recent discussions and proposed models of human–AI collaboration for creativity have emphasized *co-creation* in which humans and AI are partners or collaborators in the creative process (Beghetto, 2023; Wu et al., 2021), such that the resulting creative product is synergistic in going beyond what would be possible by humans or AI alone (Vinchon et al., 2023). Thus, under this view, AI does not replace humans. Neither do humans simply dictate or edit AI's output. Rather, humans co-create with AI.

What does "co-creation" involve? To state the obvious, the term implies that humans must (also) create. Indeed, McGuire et al. (2024) found that people were less creative when they received an AI-generated poem then freely edited it, as compared to writing a poem independently. Crucially, this deficit dissipated when people co-created with generative AI, taking turns to write one line of poetry akin to a conversation. Specifically, participants first determined the theme of the poem and wrote the first line of poetry. The generative AI system then responded with another line based on the theme, which participants could freely edit, doing so iteratively until eight lines of poetry had been collaboratively written. Thus, human–AI collaboration is enhanced when people take on the role of a co-creator in setting the creative direction, generating their own creative ideas, and developing the final creative product in tandem with AI, rather than merely reacting to AI's output as an editor.

This idea resonates with a parallel literature on how using the internet to "google" or search for information impacts human learning. Giebl et al. (2023) found that students better recalled the answers to trivia questions when they had first attempted to think of and generate the answers before consulting the internet ("thinking-before-googling"), rather than immediately searching the internet without attempting to think about the answers ("googling-right-away"). Presumably, googling-right-away bypassed the need for students to retrieve and generate the answers on their own, thus robbing them of a valuable learning opportunity since such cognitive processes benefit learning (Bjork & Bjork, 2011; Kornell et al., 2009). Yet, 81% of students in Giebl et al.'s (2023) study reported that they tended to immediately search the internet instead of thinking first, suggesting that explicit guidance may be needed to redirect students from such suboptimal tendencies.

Besides generating one's own creative ideas, co-creation involves improving, developing, and evaluating ideas in tandem with AI. In a process-mining study, Urban et al. (2025) found that on a ChatGPT-assisted creative problem-solving task, high performers iteratively dialogued with ChatGPT in monitoring, regulating, and improving the generation of ideas, whereas low performers tended to use ChatGPT merely as an information resource similar to traditional web search engines such as Google. Thus, effective human–AI collaboration involves interacting with ChatGPT as a brainstorming and feedback partner, rather than as an answer provider only.

For learners to enact such collaborative processes, guiding their metacognition and self-regulated learning is crucial. Although having access to ChatGPT in a creative task can boost learners' *creative self-efficacy* (i.e., belief in one's capacity to produce creative outcomes; Tierney & Farmer, 2002), this increased confidence in one's potential to act creatively does not always correspond to actual creative performance (Haase et al., 2018; Urban et al., 2024; cf. Puente-Díaz & Cavazos-Arroyo, 2017).

When freely interacting with ChatGPT, learners tend to default to a state of "metacognitive laziness"—less frequently engaging in metacognitive and self-regulated learning processes such as planning their strategies, and monitoring and evaluating their progress based on the task goals (Fan et al., 2025). However, such processes are vital for creative problem-solving (Greene et al., 2019; Kaufman & Beghetto, 2013; Lebuda & Benedek, 2025; Puryear, 2016). For instance, during divergent thinking, learners must devise, change, and implement their strategies to achieve the task goals (Benedek & Lebuda, 2025; Lebuda & Benedek, 2025). Moreover, during convergent thinking, learners must metacognitively evaluate their generated ideas to select the most promising ones for further development, while screening out less creative ideas. Metacognitive experiences such as mental effort and perceived task difficulty could provide cues that inform not only learners' creative problem-solving processes (Puente-Díaz, Cavazos-Arroyo, & Vargas-Barrera, 2021), but also their self-monitoring in judging their learning and self-regulation in deciding how to proceed in their problem-solving (van Gog et al., 2020). Indeed, people with higher levels of metacognitive strategies have been found to leverage ChatGPT assistance more effectively for better creative performance (Sun et al., 2025).

The Present Study: Guiding Human–AI Collaboration in Creative Problem-Solving

Based on co-creation principles for effective human–AI collaboration, the present study developed and tested a novel *"think first, ChatGPT later"* approach that guided people to interact collaboratively with ChatGPT during creative problem-solving. The approach comprised three phases. In the first phase, learners were asked to independently think of and generate their own initial ideas to solve a given creative problem. In the second phase, learners were guided to interact collaboratively with ChatGPT to improve, develop, and evaluate their ideas. Specifically, learners were presented with sample prompts intended to scaffold their creative and metacognitive processes when interacting with ChatGPT, although they could also freely use their own prompts. Finally, in the third phase, learners independently proposed a single best solution based on their initial ideas and earlier interactions with ChatGPT.

In this study, university students were first tasked to solve a creative product improvement task either using the guided approach to interact collaboratively with ChatGPT (regulated-AI group) or freely using ChatGPT for assistance (general-AI group) or independently (human-only control group). All three groups then independently solved a second, relatively more complex creative product invention task. Based on findings that ChatGPT assistance improves creative performance (e.g., Lee & Chung, 2024; Li, 2025; Urban et al., 2024): We expected that on the first creative task, the general-AI and regulated-AI groups would produce more creative (i.e., more original and useful) solutions than the human-only group. Conversely, on the second creative task when ChatGPT assistance was removed, the general-AI group's creativity would decline to levels comparable to the human-only group, since people tend to rely on rather than learn from ChatGPT when using it freely (Boers et al., 2025). Crucially, to the extent that the regulated-AI group learned deeply from collaboratively interacting with ChatGPT, they would sustain their advantage on the second creative task even when ChatGPT assistance was removed, outperforming both the human-only and general-AI groups.

To examine how the general-AI and regulated-AI groups used ChatGPT during the first creative task, we coded transcripts of their conversations with ChatGPT for the frequency of collaborative prompts (e.g., prompts aimed at improving one's initial ideas, developing selected ideas, and evaluating ideas with ChatGPT) and non-collaborative prompts (e.g., prompts directly asking ChatGPT for solutions without providing one's own ideas, and using ChatGPT as an information resource). We expected that the regulated-AI group would use more collaborative prompts than the general-AI group, in turn mediating their superior independent creativity on the second creative task.

For exploratory purposes, we further measured whether the three learning groups differed in their subjective creative problem-solving experience (e.g., mental effort, perceived task difficulty, motivation) during both creative tasks. Because creative self-efficacy and metacognitive processing (e.g., monitoring and regulating one's thinking and performance) have been linked to creative performance (Lebuda & Benedek, 2025; Puente-Díaz & Cavazos-Arroyo, 2017), learners were also asked to report their creative self-efficacy and metacognition, including their self-assessments of their performance on both creative tasks.

Method

Participants

The participants were 197 students (125 female, 67 male, 5 undisclosed) between the ages of 18 and 29 ($M = 21.62$, $SD = 2.09$) from a large public university in Singapore, who received either course credit or monetary reimbursement for their participation. The students came from diverse majors and were at different stages in their college education. Outcomes reported below are based on data from 196 participants; one participant who failed to adhere to the experimental instructions was excluded from analyses. The target sample size was determined based on the estimated effect size of $d = 0.55$ reported in Urban et al. (2024) for the benefit of freely using ChatGPT on originality in a product improvement task similar to the one in this study. An a priori power analysis (G*Power; Faul et al., 2007) indicated that at least 53 participants per condition would afford 80% power for two-tailed between-subjects pairwise comparisons at $\alpha = .05$. All participants provided their written informed consent. This study was conducted with ethics approval from the university's institutional review board.

Design

Participants were randomly assigned to one of three learning groups: *human-only* (control condition; $n = 64$), *general-AI* ($n = 67$), or *regulated-AI* ($n = 65$). The main outcomes of interest were participants' creativity on two creative problem-solving tasks: (a) a product improvement task, which participants completed independently in the control condition versus with ChatGPT assistance in both AI groups, and (b) a subsequent product invention task, which all participants completed independently.

Materials

Baseline Creativity Measure

The Alternate Uses Task (AUT; Guilford, 1967a) served as a baseline measure of participants' creative performance at the start of the experiment. The AUT is the prototypical and most commonly used divergent thinking task (Saretzki et al., 2024), although it has been viewed to require a mixture of both divergent and convergent thinking that characterize the creative process (Cortes et al., 2019). In the AUT, participants must generate creative uses for a familiar everyday object (e.g., car tire, paperclip), going beyond its common and mundane functions (e.g., using a car tire as a car part) to develop more unusual ideas (e.g., using a car tire as a chandelier). All participants were given 2 min to write down as many unusual and creative uses as possible for a car tire. Thus, the task instructions emphasized not only quantity of ideas, but also quality and creativity (Acar et al., 2020).

Intervention Creative Task

During the intervention, all participants were given 12 min to solve a product improvement task (Urban et al., 2024; see also Torrance, 2008), in which they proposed their best creative idea to improve an ordinary stuffed bunny to make it more fun to play with and achieve higher sales. Participants were explicitly told that a creative idea must be both original and useful (see detailed instructions for each learning group in Appendix A).

Sample Prompts for Regulated-AI Group

To guide the regulated-AI group's collaborative interactions with ChatGPT, seven sample prompts were created and presented (see Table 1). Based on the core processes in creative problem-solving (e.g., idea generation, development, evaluation, and selection; Vernon et al., 2016), the sample prompts illustrated how participants could collaboratively interact with ChatGPT to improve, develop, and evaluate their ideas. Specifically, the sample prompts scaffolded creative and metacognitive processes such as seeking refinements and feedback for improving ideas, elaborating on promising ideas, connecting and combining ideas, identifying weaknesses in ideas, refining selected ideas, and comparing and evaluating ideas (e.g., Lebuda & Benedek, 2025; Mumford et al., 1991, 2002). Participants were told that the sample prompts were purely for illustration purposes—if so desired, participants could use their own prompts that followed the principles of the given samples, even if their prompts were not worded identically.

Table 1 Sample prompts for regulated-AI group

Creative Problem-Solving Process	Sample Prompt
Seek initial refinements and feedback	"I need to come up with an original and useful idea to improve an ordinary stuffed bunny. My ideas are XXX and XXX. What can I change or add to make the ideas better?"
Seek more ideas and refinements	"How else can we further improve the ideas to make them even more original and useful?"
Elaborate on promising ideas	"I like your suggestion of XXX—can you expand on it?"
Compare and evaluate ideas	"Which of my ideas is/are more creative, and why?"
Connect and combine ideas	"How can I combine these ideas to make a single best creative concept?"
Identify weaknesses in ideas	"What's missing from my bunny idea?"
Refine selected ideas and iterate based on feedback	"Here's my updated idea: [describe changes]. What additional tweaks could we make?"

Post-Intervention Questionnaire

After the intervention, all participants completed a questionnaire in which they self-assessed their creative performance on the product improvement task, and rated their subjective creative problem-solving experience, creative self-efficacy, and metacognition. The general-AI and regulated-AI groups further rated their subjective experience interacting with ChatGPT and their reliance on ChatGPT during the intervention.

Self-Assessment of Creative Task Performance All participants evaluated the originality and usefulness of their solution on the product improvement task (Urban & Urban, 2021). Specifically, they rated how original their idea was for improving the stuffed bunny (1 = *not original at all*; 7 = *extremely original*) and how useful their idea was for increasing sales (1 = *not useful at all*; 7 = *extremely useful*).

Subjective Creative Problem-Solving Experience Participants rated their subjective experience solving the product improvement task on five items (adapted from Lee & Chung, 2024; Urban et al., 2024): (a) "I found solving this task interesting", (b) "I felt motivated to complete this task", (c) "I invested a lot of mental effort in this task", (d) "I feel mentally exhausted after solving this task", and (e) "This task was difficult". All items were rated on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*).

Creative Self-Efficacy To measure participants' creative self-efficacy in solving subsequent similar tasks (Beghetto & Karwowski, 2017; Karwowski et al., 2019), they expressed the extent that they agreed with the following statement: "I am confident that I can come up with original and useful ideas in future similar tasks" (1 = *strongly disagree*; 7 = *strongly agree*). Both the general-AI and regulated-AI groups additionally rated their self-efficacy in creatively solving subsequent similar tasks without ChatGPT (Urban et al., 2024): “Even without ChatGPT, I am confident that I can come up with original and useful ideas in future similar tasks” (1 = *strongly disagree*; 7 = *strongly agree*).

Metacognition in Creative Problem-Solving To measure participants’ metacognition during the intervention, the 11-item Metacognition in Creative Problem-Solving (MCPS) scale (Urban & Urban, 2023) was used. The wording of the scale items was adapted for the present study context (e.g., “When solving the task, …” instead of “When writing an essay or working on a project, …”). The MCPS measured self-assessed frequency of engagement in four metacognitive skills: (a) *planning* (e.g., “Before I started solving the task, I considered how to approach the problem in an original way”), (b) *monitoring* (e.g. “When working on the task, I asked myself if what I am doing is leading to the fulfillment of my goal”), (c) *regulation* (e.g., “When solving the task, I tried to think innovatively about the topic instead of just doing the same thing over and over again”), and (d) *evaluation* (e.g., “When I finished working on the task, I asked myself if my solution has met my goal”). All items were rated from 1 (*strongly disagree*) to 7 (*strongly agree*). Participants’ mean metacognition was computed as their mean rating across all items. The post-intervention MCPS scale displayed high internal consistency, Cronbach’s $\alpha = .85$.

Subjective Experience Interacting with ChatGPT The general-AI and regulated-AI groups further rated their subjective experience interacting collaboratively with ChatGPT on three items adapted from Lee and Chung (2024): (a) “I felt like I was having a conversation with ChatGPT”, (b) “I felt like I was brainstorming together with ChatGPT”, and (c) “I felt like I was having a social interaction with ChatGPT to come up with an idea”. All items were rated on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Participants’ mean perceived collaboration with ChatGPT was computed as their mean rating across all three items. The measure had high internal consistency in this study, Cronbach’s $\alpha = .86$.

In addition, both AI groups reported their reliance on ChatGPT in solving the product improvement task (adapted from Lee & Chung, 2024): “To what extent was your bunny idea generated by you versus external sources?” (1 = *the idea was generated entirely by external sources*; 7 = *the idea was generated entirely by me*). Participants’ ratings were reverse-scored such that higher ratings reflected greater self-reported reliance on ChatGPT.

Following Urban et al. (2024), both the general-AI and regulated-AI groups also rated their perceived usefulness of ChatGPT: “ChatGPT was useful for me when solving this task” (1 = *strongly disagree*; 7 = *strongly agree*).

Post-Intervention Creative Task

After completing the post-intervention questionnaire, all participants were given 12 min to independently solve a product invention task (adapted from McMahon et al., 2016). Specifically, participants invented a new game that helps adults and university students learn basic foreign language vocabulary for everyday life. Participants were asked to propose their best creative idea that is both original and useful, making the game fun and effective for learning. They were also told to develop their idea into a full game design, including specific details such as the game title, required materials, and gameplay description (see detailed instructions in Appendix B). Thus, whereas the product improvement task during the intervention required improving an existing product, the post-intervention product invention task was relatively more complex in that it required creating and developing a new product.

Post-Experiment Questionnaire

At the end of the experiment, all participants completed a questionnaire in which they self-assessed their creative performance on the product invention task, rated their subjective creative problem-solving experience and metacognition during the task, and indicated their prior experience using ChatGPT.

Self-Assessment of Creative Task Performance Participants evaluated how original their game was for learning a foreign language (1 = *not original at all*; 7 = *extremely original*) and how useful their game was for learning a foreign language (1 = *not useful at all*; 7 = *extremely useful*).

Subjective Creative Problem-Solving Experience Participants rated their subjective experience solving the product invention task on the same five items used in the post-intervention questionnaire. They also rated how much they agreed with the following statement: “I used what I learned from the previous task to complete this task”. All six items were rated on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*).

Metacognition in Creative Problem-Solving Participants’ metacognition during the product invention task was assessed via the same 11-item MCPS scale that had been used in the post-intervention questionnaire. The post-experiment MCPS scale displayed high internal consistency, Cronbach’s $\alpha = .88$.

Prior Experience Using ChatGPT Finally, all participants reported their prior experience using ChatGPT by selecting one of seven given options (adapted from Urban et al., 2024): (a) “I do not know ChatGPT”, (b) “I heard about ChatGPT, but I have not used it yet”, (c) “I have only used ChatGPT once”, (d) “I have used ChatGPT several times, but I do not use it regularly”, (e) “I use ChatGPT several times a month”, (f) “I use ChatGPT several times a week”, or (g) “I use ChatGPT daily”.

Procedure

Participants were run in groups of up to five per session in a lab setting. Upon arriving for the experiment, participants were each seated individually and provided with a personal computer. All computers had the same hardware and software settings. The experimental tasks and questionnaires were completed in a single open tab in the web browser window. The human-only control group had no other open tabs or windows, whereas both AI groups had a separate window open with ChatGPT-4 during the intervention. On all computers, memory/model training in ChatGPT-4 was switched off.

The experiment comprised three stages—pre-intervention, intervention, and post-intervention—with a total duration of approximately 30 min. Figure 1 depicts a flowchart of the procedure. To ensure that all participants understood what was required of them, the experimenter verbally reinforced all task instructions.

Pre-Intervention Stage

At the start of the experiment, participants' baseline creative performance was assessed via the AUT. Specifically, all participants were given 2 min to write down as many unusual and creative uses as possible for a car tire.

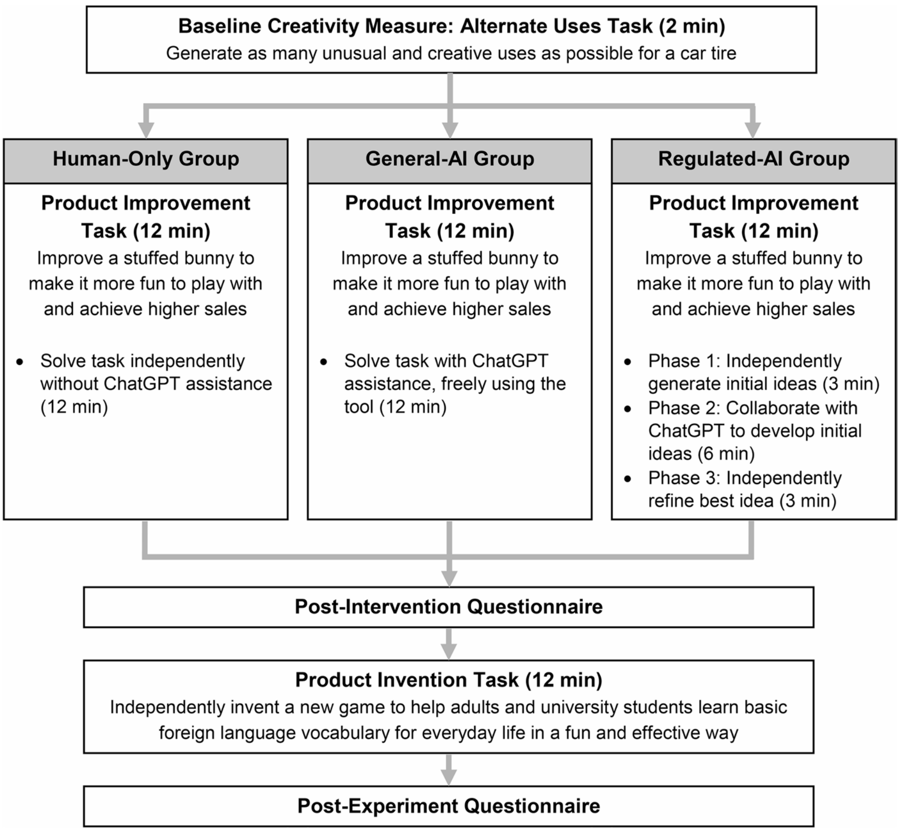

Fig. 1 Flowchart of experimental procedure

Intervention Stage

Next, the intervention began. All participants were given 12 min to solve a product improvement task, in which they proposed their best creative idea to improve an ordinary stuffed bunny to make it more fun to play with and achieve higher sales (see detailed instructions for each learning group in Appendix A). The *human-only* group solved the task independently, whereas the general-AI and regulated-AI groups solved the task with ChatGPT assistance.

The *general-AI* group was instructed to freely use ChatGPT to assist them with the task in any way they liked. To facilitate this, their personal computer was set up in a split-screen layout during the intervention with the ChatGPT-4 window open on one side and the web browser open on the other side for participants to submit their final solutions. Participants were told to give ChatGPT clear, detailed prompts and context to get the best responses from it.

The *regulated-AI* group was told that they would be guided through three phases to complete the task. In Phase 1, participants were given 3 min to independently generate their own initial creative ideas to improve the stuffed bunny. Participants were told that their initial ideas did not have to be perfect, and that they would later be given time to develop and select their best creative idea. They were also encouraged to think freely while focusing on originality and usefulness. In Phase 2, participants were given 6 min to collaboratively interact with ChatGPT to develop their initial creative ideas. During this phase, participants' personal computer was set up in a split-screen layout with the ChatGPT-4 window open on one side and the web browser open on the other side for participants to access their initial creative ideas and take notes. Participants were told to give ChatGPT clear, detailed prompts and context to get the best responses from it. In addition, participants were told to interact with ChatGPT as a brainstorming and feedback partner rather than as a provider of final answers, and to think critically about and evaluate ChatGPT's suggestions. To guide their collaborative interactions with ChatGPT, participants received sample prompts presented in their web browser window for illustration, although they were also allowed to use their own prompts. Finally, in Phase 3, participants were given 3 min to review their notes from Phase 2, and to independently refine and submit their best creative idea to improve the stuffed bunny.

Post-Intervention Stage

After the product improvement task, all participants completed the post-intervention questionnaire. The general-AI and regulated-AI groups were also asked to submit the URL link of their ChatGPT conversation for further analysis.

Next, participants were given 12 min to solve a product invention task—they proposed their best creative idea for a new game that helps adults and university students learn basic foreign language vocabulary for everyday life in a fun and effective way. All participants solved this task independently (i.e., without any access to ChatGPT or the sample prompts that had been presented to the regulated-AI group during the intervention). Finally, they completed the post-experiment questionnaire, reported their demographic information, and were debriefed and thanked.

Scoring Procedure

Two trained experts independently scored all (100%) of the 196 scripts blind to experimental condition. Both experts were well-experienced in scoring diverse creative tasks.

Baseline Creativity Measure

Participants' AUT responses were scored on fluency and flexibility (Guilford, 1967a), where fluency referred to the number of uses that participants proposed for a car tire, and flexibility referred to the number of unique conceptual categories of proposed uses. For the scoring of flexibility, we adapted George and Wiley's (2020) 16 broad categories of uses for a car tire (e.g., "clothes/fashion/jewelry", "exercise/training/sports", "vehicle-related"; available in Appendix C). Participants' flexibility score was computed as the number of unique categories that their ideas spanned, with a maximum possible score of 16. Interrater reliability was high for both fluency and flexibility, absolute agreement intraclass correlations (ICCs) = .996 and .987, 95% CI [.994, .997] and [.983, .990], respectively, based on a two-way random-effects model. Discrepancies between both raters were subsequently reviewed and resolved to yield 100% agreement.

Intervention and Post-Intervention Creative Tasks

Following extant scoring procedures (Lee & Chung, 2024; McMahon et al., 2016; Urban et al., 2024), participants' solutions on the product improvement task and product invention task were scored on three dimensions: originality, usefulness, and elaboration. *Originality* refers to the uniqueness of an idea; *usefulness* refers to how well an idea fulfils the task goals (i.e., in the product improvement task, to make the stuffed bunny more fun to play with and increase sales; in the product invention task, to create a fun and effective game for learning basic foreign language vocabulary for everyday life); *elaboration* refers to how detailed an idea is. Each dimension was scored on a scale from 1 (*not original/useful/elaborate at all*) to 7 (*extremely original/useful/elaborate*).

Originality and usefulness collectively served as measures of participants' creative task performance, in line with the standard bipartite definition of creativity (Runco & Jaeger, 2012) as requiring both originality (or novelty) and usefulness (or effectiveness). On the other hand, some degree of elaboration is likely needed to elucidate an idea's originality and/or usefulness (Dumas et al., 2021; Gonthier & Besançon, 2024), although a more elaborate idea is not necessarily more creative. Hence, participants' responses were scored for elaboration to ascertain that any differences in creativity—originality and usefulness—across groups were not unduly confounded by more detailed responses in any particular group (Forthmann et al., 2019).

Based on Amabile's (1982) Consensual Assessment Technique (CAT), a creative product can be evaluated based on consensual criteria in relation to a pool of products. That is, experts with domain experience can reliably recognize creativity and independently arrive at a reasonable level of agreement in their subjective assessments of creativity, even when explicit criteria are not imposed (Amabile, 1982). Indeed, the CAT has been considered the "gold standard" of creativity assessment (Baer & Kaufman, 2019; Cseh & Jeffries, 2019; Jeffries, 2024). Following the CAT, the expert raters viewed all participants' responses in a different randomized order, then independently rated each response relative to the pool of responses rather than against some absolute standard. The experts were encouraged to make use of the full range of the 7-point scale for each rating dimension. Sample responses for the product improvement task and product invention task are presented in Appendices D and E, respectively.

On the product improvement task, the experts demonstrated excellent internal consistency in their ratings of originality, usefulness, and elaboration, consistency ICCs = .958, .968, and .991, 95% CI [.945, .969], [.958, .976], and [.989, .994], respectively, based on a two-way random-effects model. Likewise, on the product invention task, internal consistency was excellent across the experts' ratings of originality, usefulness, and elaboration, consistency ICCs = .965, .953, and .984, 95% CI [.953, .973], [.938, .964], and [.979, .988], respectively. Hence, on both creative tasks, mean originality, usefulness, and elaboration ratings were computed by averaging both experts' ratings. These mean ratings were then used in subsequent analyses.

ChatGPT Interactions

To examine how the general-AI and regulated-AI groups used ChatGPT during the intervention, their conversation transcripts were first coded for the total number of *prompts* (i.e., the number of times that participants prompted ChatGPT in a back-and-forth exchange), with more prompts reflecting more turn-taking with ChatGPT. Participants' prompts were then classified as either collaborative or non-collaborative (see samples in Table 2). Discrepancies between both raters were reviewed and resolved to yield 100% agreement.

Collaborative prompts were prompts that improved, developed, and evaluated ideas with ChatGPT, in line with the sample prompts in the regulated-AI condition. Participants used three main subtypes of collaborative prompts: (a) *"improve own initial ideas"* (i.e., prompts providing one's initial ideas, then seeking refinements and feedback from ChatGPT), (b) *"develop selected ideas"* (i.e., prompts developing selected ideas with ChatGPT, such as elaborating on ideas, connecting and combining ideas, identifying weaknesses, and refining ideas based on feedback), and (c) *"evaluate ideas"* (i.e., prompts comparing and evaluating ideas with ChatGPT). Interrater reliability was high in scoring the number of each collaborative prompt subtype that participants used, absolute agreement ICCs = .971, .901, and .861, 95% CI [.960, .980], [.864, .929], and [.809, .900], respectively.

All other prompts were classified as *non-collaborative prompts*, comprising three main subtypes: (a) *"request ideas"* (i.e., prompts directly asking ChatGPT for the problem solutions without providing one's own ideas), (b) *"gather information"* (i.e., prompts using ChatGPT as an information resource, similar to Google and Wikipedia), and (c) *"miscellaneous"* (i.e., other non-collaborative prompts such as asking ChatGPT to rephrase its output). Interrater reliability was high in scoring the number of each non-collaborative prompt subtype that participants used, absolute

Table 2 Prompt types and sample prompts during intervention

Prompt Type	Description	Sample Prompts from Participants
Collaborative prompts		
Improve own initial ideas	Seek initial refinements and feedback	"Hi, I need help to come up with an original and useful idea to improve an ordinary stuffed bunny. Some of my initial ideas are adding a LLM into the bunny that can remember and develop a relationship with the bunny, stuffing a flashlight into the bunny, using it as some sort of sleeping aid, making it turn into a hat with good UV protection. Please help me to refine these ideas, pick the best one, or even suggest something different! Thank you." "I need to come up with an original and useful idea to improve an ordinary stuffed bunny. My ideas are to launch DIY outfits for the bunny and that people can submit their original outfits to a competition and whoever wins will have a new line of bunnies named after them. What can I change or add to make the ideas better?"
	Seek more ideas and refinements	"How can we further improve the idea of magnetic accessories and changeable expressions to make them even more original and useful?" "How else can my idea on voice activation controls for the bunny be improved?"
Develop selected ideas	Elaborate on promising ideas	"I like the idea of incorporating AR. Could you explain more on this?" "I like your idea of temperature responsive material! Do you think it can be useful as well in monitoring the child's health, like when they get a fever and can subsequently notify the parents by getting lighted up in warning colors or via mobile app? Can you expand on this?"
	Connect and combine ideas	"Is there any way we could combine the idea of extendable ears with being able to flatten the body of the bunny? I don't want it to just be bouncy." "I like the story integration idea. Some of my other ideas were also to have the bunny be weighted, scented, have a Siri-like voice assistant, etc. How can I combine all of these ideas into a single creative concept"
	Identify weaknesses in ideas	"What's missing from my bunny idea? How can it be improved to be more original and useful?" "Is it viable for mass production"
	Refine selected ideas and iterate based on feedback	"Here's the updated changes: increased security for the chair + customizable settings and adjustable from the app, anything to add on for surveillance bunny?" "How about we focus on something unique which is the voice modulation, the rabbit will have emotive responses which includes image recognition and touch response, voice response etc., and then the main focus will be the voice modulation which is animal-like voice, (human language will not be focused as it might scare the audience due to fear of AI/robots). Voice modulation will be set up in an application (allow customization features of what the bunny will sound 'Nyu' or 'Myu' with different intonations). How about this?"
Evaluate ideas	Compare and evaluate ideas	"Which of these ideas do you think is the most creative" "Which idea is better in the future world, can earn high profit and become a new trend"
Non-collaborative prompts		

Table 2 (continued)

Prompt Type	Description	Sample Prompts from Participants
Request ideas	Directly ask ChatGPT for solutions without providing one's own ideas	"Object: Stuffed bunny, 30 cm size I want you to give me 10 very unique ways I can improve this object such that it becomes more fun to play with" "Give me more unique ideas which haven't been done before"
Gather information	Use ChatGPT as an information resource	"Why do people buy things" "Example of stretchy materials as stuffing"
Miscellaneous	Other non-collaborative prompts	"Write in simple and short English" "Nono write as an idea with all the different elements"

agreement ICCs = .956, .878, and .776, 95% CI [.939, .969], [.832, .912], and [.698, .836], respectively.[1]

Results

Preliminary analyses were first conducted to ascertain the equivalence of the three learning groups in their: (a) baseline creative performance on the AUT, and (b) prior experience using ChatGPT.

Next, the main confirmatory analyses focused on the primary outcomes of: (a) participants' creativity on both creative tasks (product improvement vs. product invention), as assessed via the originality and usefulness of their solutions, (b) the elaboration of participants' solutions in both creative tasks to ascertain that any differences in creativity were not merely due to more detailed responses in any particular group, and (c) the general-AI and regulated-AI groups' use of collaborative versus non-collaborative prompts when interacting with ChatGPT during the intervention.

Finally, exploratory analyses focused on participants' ratings on the post-intervention and post-experiment questionnaires relating to the exploratory outcomes of: (a) self-assessed creative task performance, (b) subjective creative problem-solving experience, (c) creative self-efficacy, and (d) metacognition in creative problem-solving.

Preliminary Analyses

Baseline Creativity Measure

One-way between-subjects analyses of variance (ANOVAs) were conducted to ascertain that the three learning groups did not differ in their baseline creative performance on the AUT. Indeed, there were no significant differences in the learning

[1] The subsequent analyses of participants' frequency of using collaborative versus non-collaborative prompt subtypes were based on the total number of instances that they used each prompt subtype. Correspondingly, interrater reliability was assessed using ICC based directly on these total frequency scores (as opposed to the classification of each prompt under a particular subtype).

groups' fluency scores, $F(2, 193)=0.30$, $p=.74$, $\eta^2 = 0.003$, and flexibility scores, $F(2, 193)=0.32$, $p=.73$, $\eta^2 = 0.003$. Means and standard deviations are presented in Table 3.

Prior Experience Using ChatGPT

To ascertain that the three learning groups did not differ in their prior experience using ChatGPT, their self-reported experience on the post-experiment questionnaire was analyzed. A chi-square test revealed that there was no significant difference in the distribution of frequencies for participants' self-reported prior experience using ChatGPT, $\chi^2(10, N=196)=8.35$, $p=.60$. Table 4 displays the frequency counts and percentage values for each learning group.

Table 3 Means and standard deviations for baseline creativity measure, post-intervention questionnaire, and post-experiment questionnaire

Variables	Human-Only		General-AI		Regulated-AI	
	M	SD	M	SD	M	SD
Baseline creativity measure						
Fluency	5.48	2.68	5.52	2.48	5.78	1.99
Flexibility	4.11	1.44	4.07	1.61	4.28	1.57
Post-intervention questionnaire						
Self-assessed originality	4.73	1.28	4.15	1.18	4.45	1.16
Self-assessed usefulness	5.30	1.05	5.33	0.98	5.31	1.06
Interesting	5.30	1.31	5.45	1.15	5.72	1.07
Motivated	5.17	1.22	5.37	1.19	5.60	1.09
Mental effort	4.91	1.57	4.31	1.44	5.03	1.22
Mentally exhausted	3.41	1.66	3.03	1.54	3.38	1.54
Task difficulty	3.89	1.74	3.30	1.46	3.68	1.59
Creative self-efficacy	4.59	1.19	4.51	1.47	4.63	1.13
Creative self-efficacy without ChatGPT			4.31	1.84	4.31	1.55
Metacognition	5.14	0.93	4.88	0.96	5.15	0.69
Perceived collaboration with ChatGPT			4.13	1.57	4.67	1.38
Reliance on ChatGPT			4.30	1.41	3.57	1.20
Usefulness of ChatGPT			5.81	1.21	6.08	0.92
Post-experiment questionnaire						
Self-assessed originality	5.11	1.25	5.09	1.28	4.71	1.47
Self-assessed usefulness	5.48	1.14	5.37	0.92	5.40	1.07
Interesting	5.67	1.24	5.96	1.08	5.43	1.20
Motivated	5.44	1.25	5.82	1.11	5.34	1.20
Mental effort	5.33	1.40	5.43	1.20	5.48	1.11
Mentally exhausted	4.02	1.99	3.75	1.69	4.15	1.62
Task difficulty	4.39	1.69	4.15	1.57	4.45	1.59
Applied learning from product improvement task	4.14	1.88	3.79	1.70	4.14	1.51
Metacognition	5.30	0.94	5.30	0.92	5.23	0.75

N = 196. Ratings on the post-intervention questionnaire pertained to the product improvement task during the intervention, whereas ratings on the post-experiment questionnaire pertained to the product invention task after the intervention. The following variables on the post-intervention questionnaire were not applicable to the human-only group: creative self-efficacy without ChatGPT, perceived collaboration with ChatGPT, reliance on ChatGPT, and usefulness of ChatGPT

Table 4 Frequency counts and percentages of participants' prior ChatGPT experience

Prior ChatGPT Experience	Human-Only (n=64)	General-AI (n=67)	Regulated-AI (n=65)
"I do not know ChatGPT"	0 (0%)	0 (0%)	0 (0%)
"I heard about ChatGPT, but I have not used it yet"	0 (0%)	1 (2%)	0 (0%)
"I have only used ChatGPT once"	1 (2%)	2 (3%)	1 (2%)
"I have used ChatGPT several times, but I do not use it regularly"	8 (13%)	10 (15%)	7 (11%)
"I use ChatGPT several times a month"	8 (13%)	12 (18%)	18 (28%)
"I use ChatGPT several times a week"	23 (36%)	23 (34%)	22 (34%)
"I use ChatGPT daily"	24 (38%)	19 (28%)	17 (26%)

Main Analyses

Creativity

To analyze participants' creativity during and after the intervention, we compared the originality and usefulness of their solutions on the product improvement task versus product invention task. The Bonferroni correction was applied to control for multiple pairwise comparisons.

Originality on Creative Tasks A 3 (Learning Group) × 2 (Creative Task) mixed ANOVA with participants' originality scores as the dependent variable revealed a significant interaction, $F(2, 193) = 13.80$, $p < .001$, $\eta_p^2 = 0.13$, which we decomposed by examining the simple effects (see Fig. 2A). On the product improvement task during the intervention, the general-AI ($M = 4.08$, $SD = 1.22$) group produced more original solutions than both the human-only ($M = 3.06$, $SD = 1.37$) and regulated-AI ($M = 3.30$, $SD = 1.30$) groups, $p < .001$ and $p = .002$, $d = 0.79$ and -0.62, respectively. However, the regulated-AI group did not differ from the human-only group in their originality during the intervention, $p = .90$, $d = 0.18$.

Strikingly, this pattern reversed on the product invention task after the intervention, whereby the regulated-AI ($M = 3.42$, $SD = 1.63$) group produced more original solutions than both the human-only ($M = 2.63$, $SD = 1.45$) and general-AI ($M = 2.69$, $SD = 1.44$) groups, $p = .009$ and .017, $d = 0.52$ and 0.48, respectively. In contrast, the general-AI group did not differ from the human-only group in their originality after the intervention, $p = 1.00$, $d = 0.04$.

At the same time, both the human-only and general-AI groups declined in their originality from the product improvement task to the relatively more complex product invention task, $p = .039$ and $p < .001$, $d = -0.31$ and -1.05, respectively. Conversely, the regulated-AI group sustained their originality across both tasks, $p = .56$, $d = 0.08$.

Usefulness on Creative Tasks Likewise, a 3 (Learning Group) × 2 (Creative Task) mixed ANOVA with participants' usefulness scores as the dependent variable revealed a significant interaction, $F(2, 193) = 11.08$, $p < .001$, $\eta_p^2 = 0.10$ (see Fig. 2B). On the product improvement task during the intervention, the general-AI ($M = 4.29$, $SD = 1.05$) group produced more useful solutions than both the human-only ($M = 3.79$,

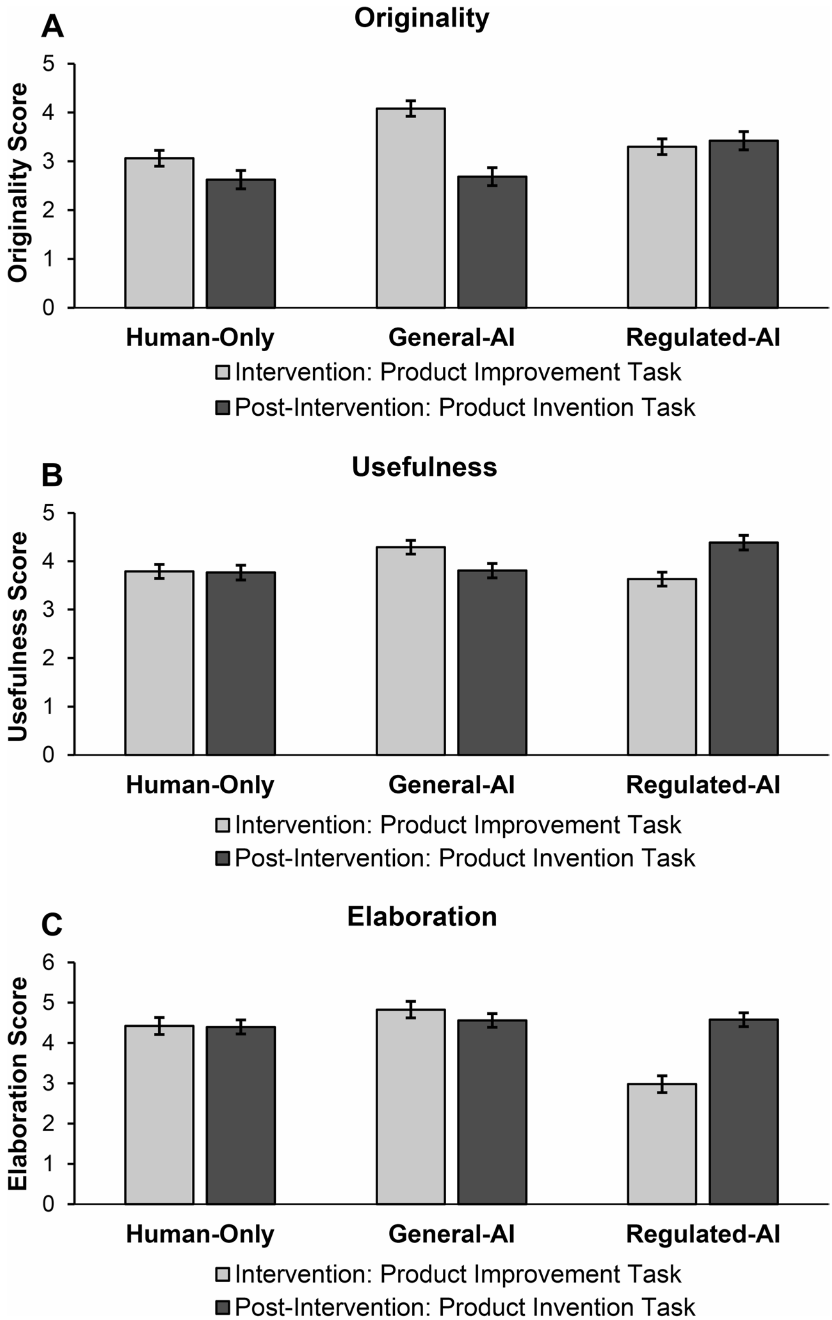

Fig. 2 Originality, usefulness, and elaboration scores on creative problem-solving tasks. *Note*. (**A**), (**B**), and (**C**) show the mean originality, usefulness, and elaboration scores, respectively. The general-AI group produced more creative—original and useful—solutions than the human-only and regulated-AI groups when assisted by ChatGPT during the intervention, but failed to sustain this creative advantage after the intervention when unassisted by ChatGPT. In contrast, the regulated-AI group showed learning gains in outperforming both the human-only and general-AI groups on independent creativity after the intervention. Error bars indicate standard errors

SD = 1.27) and regulated-AI (*M* = 3.63, *SD* = 1.16) groups, $p = .042$ and .004, $d = 0.43$ and −0.60, respectively. Conversely, the regulated-AI group did not significantly differ from the human-only group in their usefulness scores during the intervention, $p = 1.00$, $d = -0.13$.

Yet, the critical finding once again was that this pattern reversed on the product invention task after the intervention. Specifically, the regulated-AI (*M* = 4.39, *SD* = 1.22) group now produced more useful solutions than both the human-only (*M* = 3.77, *SD* = 1.23) and general-AI (*M* = 3.81, *SD* = 1.20) groups, $p = .013$ and .021, $d = 0.50$ and 0.48, respectively. In contrast, the general-AI group did not differ from the human-only group in their usefulness scores after the intervention, $p = 1.00$, $d = 0.03$.

Comparisons of each group's usefulness scores on both creative tasks showed that the human-only group maintained their performance during and after the intervention, $p = .90$, $d = -0.02$, and the general-AI group's usefulness scores declined on the product invention task, $p = .01$, $d = -0.43$, whereas the regulated-AI group showed gains, $p < .001$, $d = 0.63$.

Elaboration in Creative Tasks

To ascertain that the observed differences in the learning groups' creativity—originality and usefulness—were not simply due to more elaborate responses in any particular group, we analyzed participants' elaboration scores across both creative tasks. A 3 (Learning Group) × 2 (Creative Task) mixed ANOVA with participants' elaboration scores as the dependent variable revealed a significant interaction, $F(2, 193) = 21.82$, $p < .001$, $\eta_p^2 = 0.18$, albeit with a different pattern from their originality and usefulness scores (see Fig. 2C). On the product improvement task during the intervention, the regulated-AI (*M* = 2.98, *SD* = 1.79) group produced less elaborate solutions than both the human-only (*M* = 4.42, *SD* = 1.77) and general-AI (*M* = 4.83, *SD* = 1.48) groups, both *p*s < .001, $d = -0.81$ and −1.13, respectively. The latter two groups did not differ, $p = .51$, $d = 0.25$.

In contrast, on the product invention task after the intervention, the regulated-AI (*M* = 4.58, *SD* = 1.24) group's solutions did not differ in elaboration from those of the human-only (*M* = 4.40, *SD* = 1.58) and general-AI (*M* = 4.56, *SD* = 1.33) groups, both *p*s = 1.00, $d = 0.13$ and 0.01, respectively. Neither did the latter two groups differ, $p = 1.00$, $d = 0.11$. Thus, the regulated-AI group's superior creativity could not have been due to any differences in elaboration (e.g., producing more detailed responses).

Comparisons of each group's elaboration scores on both creative tasks showed that the human-only and general-AI groups maintained their performance across tasks, $p = .92$ and .21, $d = -0.01$ and −0.19, respectively. Conversely, the regulated-AI group produced more elaborate solutions after than during the intervention, $p < .001$, $d = 1.04$.

ChatGPT Interactions

To compare how the general-AI and regulated-AI groups interacted with ChatGPT during the intervention, independent samples *t*-tests were conducted to analyze the frequency of prompt types that they used. In instances where Levene's test for equality of variances was violated, the adjusted degrees of freedom are reported below; in instances where data were found to be skewed, nonparametric tests were applied. Table 5 displays the means and standard deviations for each prompt type, as well as the frequency that participants used each prompt type as a percentage of their total prompts.

Total Prompts The general-AI and regulated-AI groups did not significantly differ in the total number of prompts that they used, $t(102.55) = -1.08$, $p = .28$, $d = -0.18$. Thus, it is unlikely that the differences in both groups' creative task performance were due to more or less frequent turn-taking with ChatGPT during the intervention.

Collaborative Prompts As predicted, the regulated-AI group used significantly more collaborative prompts than the general-AI group, $t(130) = 8.28$, $p < .001$, $d = 1.44$. Whereas only 29.1% of the general-AI group's prompts were collaborative in nature, collaborative prompts formed the majority (88.6%) of the regulated-AI group's total prompts, with most belonging to the "improve own initial ideas" (40.5%) and "develop selected ideas" (41.4%) subtypes.

Further analyses of the three collaborative prompt subtypes showed that: Relative to the general-AI group, the regulated-AI group used more "improve own initial ideas" prompts, $t(130) = 11.90$, $p < .001$, $d = 2.06$, and "develop selected ideas" prompts, $t(130) = 2.69$, $p = .008$, $d = 0.47$. Both groups used relatively few "evaluate ideas" prompts, with none in the general-AI condition; "evaluate ideas" prompt scores were nonnormally distributed with skewness of 2.74 ($SE = 0.21$) and kurtosis of 7.09 ($SE = 0.42$). Thus, nonparametric bootstrapping with 10,000 samples was applied to robustly estimate the standard error and 95% bias-corrected and accelerated confidence interval (BCa CI) for the mean difference between both groups. The regulated-AI group used more "evaluate ideas" prompts than the general-AI group, $M_{\text{difference}} = 0.26$, bootstrap $SE = 0.06$, 95% BCa CI [0.16, 0.38].

Table 5 Means, percentages, and standard deviations for prompt types in general-AI and regulated-AI conditions

Prompt Type	General-AI		Regulated-AI	
	M (Percentage)	SD	M (Percentage)	SD
Total prompts	4.10 (100%)	2.64	3.71 (100%)	1.43
Collaborative prompts	1.30 (29.1%)	1.31	3.20 (88.6%)	1.33
Improve own initial ideas	0.24 (7.9%)	0.50	1.29 (40.5%)	0.52
Develop selected ideas	1.06 (21.2%)	1.29	1.65 (41.4%)	1.22
Evaluate ideas	0.00 (0%)	0.00	0.26 (6.7%)	0.48
Non-collaborative prompts	2.81 (70.9%)	1.94	0.51 (11.4%)	0.92
Request ideas	2.21 (59.6%)	1.49	0.20 (4.8%)	0.51
Gather information	0.18 (5.0%)	0.52	0.15 (3.6%)	0.62
Miscellaneous	0.42 (6.3%)	0.86	0.15 (3.0%)	0.44

The general-AI group did not use any "evaluate ideas" prompts. Percentage (%) values represent the frequency that participants used each prompt type as a percentage of their total prompts

Non-Collaborative Prompts In contrast, the general-AI group used more non-collaborative prompts than the regulated-AI group, $t(94.92) = -8.74$, $p<.001$, $d = -1.51$. Whereas non-collaborative prompts formed only 11.4% of the regulated-AI group's total prompts, 70.9% of the general-AI group's prompts were non-collaborative in nature, with most (59.6%) belonging to the "request ideas" subtype.

Further analyses of the three non-collaborative prompt subtypes revealed that the general-AI group used more "request ideas" prompts than the regulated-AI group, $t(81.39) = -10.42$, $p<.001$, $d = -1.80$. "Gather information" and "miscellaneous" prompt scores were nonnormally distributed with skewness of 4.31 and 2.89 ($SE=0.21$), and kurtosis of 21.32 and 9.16 ($SE=0.42$), respectively. Hence, nonparametric bootstrapping with 10,000 samples was applied, indicating that both groups did not differ in their number of "gather information" prompts used, $M_{\text{difference}} = -0.03$, bootstrap $SE=0.10$, 95% BCa CI [−0.21, 0.18], but that the general-AI group used more "miscellaneous" prompts than the regulated-AI group, $M_{\text{difference}} = -0.26$, bootstrap $SE=0.12$, 95% BCa CI [−0.52, −0.04].

ChatGPT Interactions and Post-Intervention Creativity

Correlations Correlational analyses examined whether the frequencies of each prompt (sub)type that both AI groups used during the intervention were associated with their subsequent independent creativity (i.e., originality and usefulness scores) on the product invention task. Table 6 displays the correlation matrix.

Post-intervention originality scores significantly and positively correlated with the number of collaborative prompts that participants used when interacting with ChatGPT during the intervention, $r(130)=.22$, $p=.01$, although this relationship was driven mainly by the "improve own initial ideas" prompt subtype, $r(130)=.30$, $p=.001$. The "develop selected ideas" and "evaluate ideas" prompt subtypes were not significantly associated with post-intervention originality scores, $r(130)=.12$ and $-.05$, $p=.16$ and .61, respectively.

In addition, post-intervention originality scores significantly and negatively correlated with the number of non-collaborative prompts that participants used during the intervention, $r(130)=-.17$, $p=.049$. This relationship was driven mainly by the "gather information" prompt subtype, $r(130)=-.17$, $p=.049$; the "request ideas" and "miscellaneous" prompt subtypes were not significantly associated with post-intervention originality scores, $r(130)=-.10$ and $-.12$, $p=.26$ and .18, respectively.

Post-intervention usefulness scores significantly and positively correlated only with the number of "improve own initial ideas" prompts that participants used during the intervention, $r(130)=.21$, $p=.019$. None of the other prompt subtypes were significantly associated with usefulness scores on the product invention task, all ps$>.05$.

Mediation Analyses Overall, the preceding results showed that the regulated-AI group outperformed the general-AI group in independent human creativity after the intervention. Furthermore, process analyses of both groups' interactions with ChatGPT during the intervention showed that the regulated-AI group used more collaborative prompts than the general-AI group. Crucially, only the "improve own initial ideas" prompt subtype was significantly associated with creativity—both originality

Table 6 Correlations for ChatGPT interactions during intervention and post-intervention independent creativity (general-AI and regulated-AI groups)

Variables		1	2	3	4	5	6	7	8	9	10
1.	Independent originality on product invention task	—									
2.	Independent usefulness on product invention task	.47***	—								
3.	Collaborative prompts	.22*	.09	—							
4.	"Improve own initial ideas" prompts	.30**	.21*	.56***	—						
5.	"Develop selected ideas" prompts	.12	.01	.83***	.04	—					
6.	"Evaluate ideas" prompts	−.05	−.05	.43***	.32***	.08	—				
7.	Non-collaborative prompts	−.17*	−.08	−.28**	−.57***	.04	−.26**	—			
8.	"Request ideas" prompts	−.10	−.08	−.36***	−.60***	−.04	−.27**	.88***	—		
9.	"Gather information" prompts	−.17*	.03	−.17	−.20*	−.08	−.07	.28**	−.06	—	
10.	"Miscellaneous" prompts	−.12	−.09	.14	−.12	.26**	−.09	.62***	.29**	.09	—

*p<.05. **p<.01. ***p< .001

and usefulness scores—on the product invention task. This pattern did not occur for all other prompt subtypes. Indeed, the "improve own initial ideas" prompt subtype formed 40.5% of the regulated-AI group's total prompts, but only 7.9% of the general-AI group's total prompts.

Hence, regression analyses were conducted to test whether the regulated-AI group's greater use of "improve own initial ideas" prompts during the intervention mediated their creative advantage over the general-AI group after the intervention. A percentile bootstrap estimation approach with 10,000 samples was used in Model 4 of Hayes's (2013) PROCESS macro. Figure 3 depicts the results of the mediation analyses. With the number of "improve own initial ideas" prompts as the mediator, the indirect effect of learning method was significant for originality on the product invention task, 0.61, 95% CI [0.12, 1.19] (see Fig. 3A), and non-significant for usefulness on the product invention task, 0.13, 95% CI [−0.27, 0.62] (see Fig. 3B). Thus, the regulated-AI group engaged in more collaborative interactions with ChatGPT

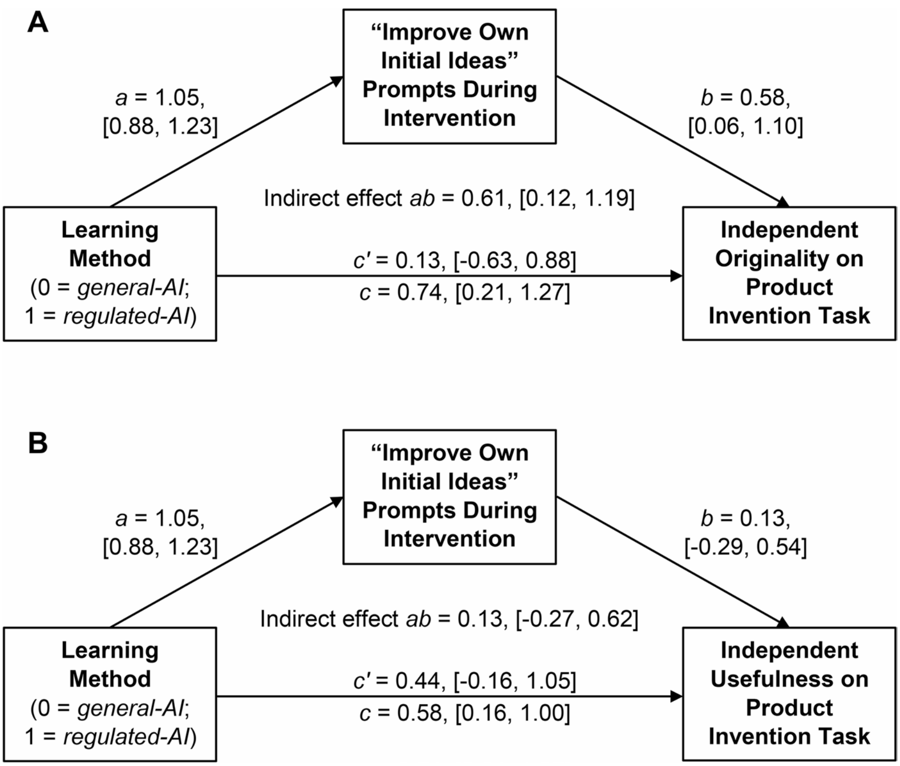

Fig. 3 Results of mediation analyses. *Note*. (**A**) and (**B**) depict the mediation models with post-intervention originality and usefulness scores as the dependent variable, respectively. Unstandardized regression coefficients are presented for each path. The *c′* path coefficients represent the direct effects, whereas the *c* path coefficients represent the total effects. Values in brackets represent the 95% confidence intervals for the regression coefficients using a percentile bootstrap estimation approach with 10,000 samples

aimed specifically at improving their own initial ideas, in turn mediating their superior independent originality over the general-AI group after the intervention.

Exploratory Analyses

For exploratory purposes, participants' ratings on the post-intervention and post-experiment questionnaires were analyzed to examine their self-assessed creative task performance, subjective creative problem-solving experience, creative self-efficacy, and metacognition. The Bonferroni correction was applied to control for multiple pairwise comparisons. Means and standard deviations are presented in Table 3.

Self-Assessment of Creative Task Performance

On the product improvement task during the intervention, the three learning groups significantly differed in their self-assessed originality, $F(2, 193)=3.85$, $p=.023$, $\eta^2 = 0.04$, but not their self-assessed usefulness, $F(2, 193)=0.02$, $p=.98$, $\eta^2 < 0.001$. In contrast to the expert raters' evaluations, the general-AI group perceived their solutions as less original than the human-only group, $p=.018$, $d = -0.47$, and just as original as the regulated-AI group, $p=.48$, $d=0.26$. The latter two groups did not significantly differ in their self-assessed originality on the product improvement task, $p=.53$, $d=-0.23$.

On the product invention task after the intervention, the three learning groups neither differed in their self-assessed originality, $F(2, 193)=1.88$, $p=.16$, $\eta^2 = 0.02$, nor self-assessed usefulness, $F(2, 193)=0.20$, $p=.82$, $\eta^2 = 0.002$. Thus, participants' self-evaluations of their creative task performance were at odds with the expert raters' evaluations on overall.

Subjective Creative Problem-Solving Experience

On the product improvement task, there were no significant differences in how interesting the three learning groups found the task, $F(2, 193)=2.18$, $p=.12$, $\eta^2 = 0.02$, how motivated they felt to complete the task, $F(2, 193)=2.18$, $p=.12$, $\eta^2 = 0.02$, how mentally exhausted they felt after solving the task, $F(2, 193)=1.19$, $p=.31$, $\eta^2 = 0.01$, and how difficult they found the task, $F(2, 193)=2.32$, $p=.10$, $\eta^2 = 0.02$. However, the groups differed in the amount of mental effort that they reported investing in the task, $F(2, 193)=4.84$, $p=.01$, $\eta^2 = 0.05$. Specifically, the general-AI group reported less mental effort than the regulated-AI group, $p=.012$, $d=0.54$. The human-only group did not significantly differ from the general-AI and regulated-AI groups in their self-reported mental effort on the product improvement task, $p=.053$ and 1.00, $d=-0.40$ and 0.09, respectively.

On the product invention task, there were no significant differences in how motivated the three learning groups felt to complete the task, $F(2, 193)=3.05$, $p=.050$, $\eta^2 = 0.03$, how much mental effort they reported investing in the task, $F(2, 193)=0.25$, $p=.78$, $\eta^2 = 0.003$, how mentally exhausted they felt after solving the task, $F(2, 193)=0.91$, $p=.41$, $\eta^2 = 0.01$, how difficult they found the task, $F(2, 193)=0.63$, $p=.53$, $\eta^2 = 0.01$, and how much they perceived using what they had learned from the earlier intervention to complete the task, $F(2, 193) = 0.93$, $p = .40$, $\eta^2 = 0.01$. However, the groups differed in how interesting they found the product invention task, $F(2, 193) = 3.32$, $p = .038$, $\eta^2 = 0.03$. Specifically, the general-AI group perceived the task as more interesting than the regulated-AI group, $p = .033$, $d = -0.46$, but not the human-only group, $p = .50$, $d = 0.25$. The latter two groups did not significantly differ in how interesting they found the product invention task, $p = .73$, $d = -0.20$.

Creative Self-Efficacy

After the intervention, the three learning groups did not significantly differ in their creative self-efficacy in solving future similar tasks, $F(2, 193) = 0.16$, $p = .85$, $\eta^2 = 0.002$. Neither did the general-AI and regulated-AI groups differ in their creative self-efficacy in solving future similar tasks without ChatGPT, $t(130) = -0.02$, $p = .99$, $d = 0$.

Metacognition in Creative Problem-Solving

There were no significant differences in the three learning groups' self-reported metacognition when solving the product improvement task during the intervention, $F(2, 193) = 1.98$, $p = .14$, $\eta^2 = 0.02$, and when solving the product invention task after the intervention, $F(2, 193) = 0.14$, $p = .87$, $\eta^2 = 0.001$.

Subjective Experience Interacting with ChatGPT

The general-AI and regulated-AI groups did not differ in their perceived usefulness of ChatGPT for solving the product improvement task during the intervention, $t(123.29) = 1.45$, $p = .15$, $d = 0.25$. Nonetheless, the regulated-AI group perceived significantly higher collaboration with ChatGPT than the general-AI group, $t(130) = 2.10$, $p = .037$, $d = 0.37$. In addition, the regulated-AI group reported lower reliance on ChatGPT during the intervention than the general-AI group, $t(130) = -3.19$, $p = .002$, $d = -0.56$. Thus, in line with the guided approach's design to foster co-creation with ChatGPT, the regulated-AI group expressed less agreement than the general-AI group that their solution on the product improvement task had been generated entirely by external sources rather than themselves.

Discussion

The fundamental distinction between performance versus learning is crucial (Soderstrom & Bjork, 2015), but has often been neglected in extant research on generative AI in education (Yan et al., 2025). Although using generative AI tools such as ChatGPT can benefit creative performance on the assisted task (Lee & Chung, 2024; Urban et al., 2024), do these benefits persist when ChatGPT assistance is later removed? The present findings reveal that the answer to this question hinges critically on *how* people use ChatGPT.

When learners freely used ChatGPT to solve a creative product improvement task, they produced more creative—original and useful—solutions than learners who solved the task without ChatGPT assistance. Yet, this advantage was not sustained for independent human creativity—when ChatGPT assistance was later removed on a more complex, creative product invention task, the general-AI group's creativity declined to levels comparable to that of the human-only group. Thus, freely using ChatGPT produced only transient performance boosts but not deeper learning gains.

Yet, a striking reversal occurred when learners used a guided approach to interact with ChatGPT. Specifically, we developed a novel "think first, ChatGPT later" approach that guided learners to first independently think of and generate their own creative solutions, then collaborate with ChatGPT to improve, develop, and evaluate their ideas. Despite a lack of immediate performance gains on the assisted task, the regulated-AI group outperformed both the human-only and general-AI groups in independent creativity when ChatGPT assistance was later removed. Moreover, whereas the human-only and general-AI groups' originality declined on the more complex product invention task, the regulated-AI group sustained their originality and even improved in the usefulness of their solutions.

These findings align with the notion of "desirable difficulties", whereby some effective techniques induce difficulties that may degrade performance during initial training or acquisition, but in fact yield deep learning from the beneficial cognitive processes fostered when confronting and resolving difficulties (Bjork, 1994; Bjork & Bjork, 2011). Thus, substantial learning can occur in the absence of any discernible changes in performance (Soderstrom & Bjork, 2015).

Theoretical Implications

What beneficial processes did the regulated-AI group engage in that supported their learning? Based on the present data, some potential explanations can be ruled out. For one, it is unlikely that elaboration (i.e., producing more detailed responses) accounted for the observed differences in creativity. The regulated-AI group produced less elaborate solutions during the intervention, likely because they had relatively less time to write their best creative idea after having to first generate their own ideas then interact with ChatGPT, whereas the human-only and general-AI groups could freely write their best creative idea throughout the intervention. In addition, the three learning groups did not differ in their elaboration on the subsequent product invention task, yet the regulated-AI group's solutions were more original and useful. At the same time, the exploratory analyses revealed that the three groups did not differ in their self-reported creative self-efficacy, motivation, mental exhaustion, and perceived task difficulty during both creative tasks. Moreover, the general-AI group perceived the product invention task as more interesting than the regulated-AI group, yet performed less creatively. Thus, all these variables cannot explain the regulated-AI group's superior independent creativity, although some caution is warranted when interpreting participants' ratings of their subjective learning experience and processes, given that these variables were measured via self-report and for exploratory purposes only.

Rather, process analyses of learners' conversations with ChatGPT during the intervention revealed that the majority (88.6%) of the regulated-AI group's prompts were collaborative in improving, developing, and evaluating ideas with ChatGPT, whereas the majority (70.9%) of the general-AI group's prompts were non-collaborative. Notably, the regulated-AI group used more collaborative prompts aimed specifically at improving their initial self-generated ideas, relative to the general-AI group. In turn, greater engagement in improving one's initial ideas with ChatGPT mediated the advantage of the "think first, ChatGPT later" approach over freely interacting with ChatGPT for learners' independent originality later on.

Why did greater use of "improve own initial ideas" prompts with ChatGPT boost independent human originality? Divergent thinking in exploring and generating multiple ideas is central to the creative process (Acar & Runco, 2019; Runco & Acar, 2012). Yet, the general-AI group often bypassed this divergent thinking process, outsourcing it to ChatGPT. Consistent with findings that people often default to a state of "metacognitive laziness" when freely using ChatGPT (Fan et al., 2025), 59.6% of the general-AI group's prompts directly asked ChatGPT to generate ideas to solve the task without providing their own ideas, whereas only 7.9% of their prompts were aimed at improving their own ideas. Unsurprisingly, the general-AI group reported less mental effort during the intervention than the regulated-AI group. In all, the general-AI group tended to immediately dictate ChatGPT to produce the solutions without first attempting to think about them, thus robbing themselves of a valuable learning opportunity.

In contrast, the regulated-AI group first generated their own ideas then collaborated with ChatGPT to improve and develop them. This approach could have enhanced their deep learning by orienting them toward the cognitive processes needed specifically for producing more original ideas. For instance, the regulated-AI group could have honed their divergent thinking skills when first brainstorming their own ideas. Further, when improving their self-generated ideas with ChatGPT's feedback, the regulated-AI group could have learned how their initial ideas could be transformed into the next evolutionary form (e.g., from a stuffed bunny to a virtual bunny in a fully immersive digital environment) by "digging deeper", and learning how to make new associations between seemingly unrelated ideas by "digging elsewhere" (Ritter & Mostert, 2017). It is also possible that the regulated-AI group learned how to apply and/or self-regulate more deliberate, strategic approaches to solve creative problems (Gilhooly et al., 2007; Rubenstein et al., 2018) through the scaffolding provided via the sample prompts. In turn, learners' independent originality could have been enhanced when they transferred these learned skills and strategies to solve the subsequent product invention task.

Although the regulated-AI group also independently produced more useful solutions than the general-AI group, this advantage was not significantly mediated by greater use of "improve own initial ideas" prompts during the intervention. Whereas developing an original solution draws on divergent thinking, ensuring that this solution is useful also requires convergent thinking to assess its effectiveness (Cropley, 2006). Thus, to the extent that using more "improve own initial ideas" prompts reflected more or deeper divergent thinking (e.g., when generating one's own ideas before interacting with ChatGPT), this could then have mediated the regulated-AI group's advantage for independent originality but not necessarily usefulness. Given that none of the other prompt subtypes were significantly associated with post-intervention usefulness scores, further work is needed to probe the mechanisms beneath the regulated-AI group's more useful independent solutions. For instance, learners may have benefited from a more productive interplay between divergent and convergent thinking (e.g., when iteratively generating, developing, and evaluating their ideas with ChatGPT) that may not have been fully captured by the separate prompt subtypes alone.

At the same time, it should be noted that simply generating one's own ideas during the intervention did not suffice for learning gains. The human-only group independently generated and developed all their ideas without any ChatGPT assistance, yet performed more poorly than the regulated-AI group after the intervention when required to invent a new product. Indeed, the human-only group declined in their originality on this more complex task than when they simply improved an existing product during the intervention. This suggests that collaborating with ChatGPT can confer additional benefits for human creativity and learning, yielding redefinition effects that surpass what humans can achieve alone without generative AI (Bauer et al., 2025; Vinchon et al., 2023).

Practical Implications

Foreseeably, digitally mediated creative behavior will increase rather than recede in the future (Ceh et al., 2024). Thus, it is vital to guide learners to interact effectively—collaboratively—with generative AI tools such as ChatGPT for sustained learning gains rather than transient performance boosts. Applied in educational contexts, the present data offer support for a "think first, ChatGPT later" approach in enhancing human learning and creativity. Specifically, learners should be guided to first independently think of and generate their own solutions to a given creative problem, then collaborate with ChatGPT to improve, develop, and evaluate their ideas. This approach circumvents learners' suboptimal tendencies to assume the role of an editor rather than a co-creator when spontaneously using ChatGPT (Boers et al., 2025; Fan et al., 2025), while leveraging ChatGPT for learning gains beyond what learners could otherwise achieve alone.

However, as this and other studies have revealed (Urban et al., 2024; Urban & Urban, 2021), learners may have low metacognitive awareness in assessing their creative products. Because inaccurate metacognitive monitoring can dampen creative performance (Urban & Urban, 2025; see also Jaarsveld & van Leeuwen, 2005; Puente-Díaz, Cavazos-Arroyo, & Puerta-Sierra, 2021), it could be valuable to support learners in developing their metacognition through feedback. For instance, recent advances in automated techniques using semantic analysis to assess originality enable real-time feedback to be delivered rapidly and efficiently during creative idea generation, thus facilitating learners' metacognition and creative performance (de Chantal & Organisciak, 2025; see also Marrone et al., 2026).

Limitations and Future Directions

Notwithstanding the initial promising evidence, more room remains to further develop and refine the "think first, ChatGPT later" approach. For instance, the sample prompts in the regulated-AI condition were presented purely for illustration purposes, and were not engineered to produce the "best" output from ChatGPT. In future work, prompt engineering (White et al., 2023; Zamfirescu-Pereira et al., 2023) could potentially boost the effectiveness of learners' prompts when collaborating with ChatGPT. In addition, when independently generating their initial ideas, the regulated-AI group was not explicitly guided to plan the implementation of these ideas (e.g., identifying key causes, restrictions, consequences, implementation strategies, contingencies; Mumford et al., 2001), as with both the human-only and general-AI groups. Because training people's planning skills can improve their creative thinking (Osburn & Mumford, 2006), it could be fruitful to incorporate such elements in future iterations of the "think first, ChatGPT later" approach. Likewise, supplementing this approach with self-regulation prompts (e.g., uncertainty acceptance, emotion regulation and dealing with obstacles, adjusting one's problem-solving approach) may further boost creativity (Zielińska et al., 2025), given that errors and failure are associated with creative insight (Sawyer, 2019) and improved learning outcomes (Wong & Lim, 2022a, 2022b), but may impose emotional costs that prevent effective learning from them (Wong & Lim, 2019b).

An arising question relates to the extent that learners could reap similar learning benefits for their creativity if they received the same guided approach and sample prompts as the regulated-AI group during the intervention but did not interact with ChatGPT. To address this question, future work could compare the regulated-AI group with a "regulated-human" group that independently works through the sample prompts on their own during the intervention. On one hand, working alone to systematically generate, improve, and develop one's own ideas could push learners to produce more ideas for better learning gains in divergent thinking skills than when collaborating with ChatGPT. For instance, human brainstorming groups have often been observed to suffer a productivity loss where each member generates fewer creative ideas on average when working with the group than when working alone, presumably due to coordination losses during turn-taking among group members (Nijstad, 2015; Paulus, 2000; Stroebe et al., 2010). On the other hand, groups can be more creative than individuals in developing higher-quality ideas (McMahon et al., 2016). Presumably, interacting with others could expose individuals to more diverse ideas, which could stimulate knowledge activation and enrich the creative process for better learning (Nijstad, 2015; Paulus, 2000). If so, learners may benefit more from applying the guided approach with ChatGPT as a collaborator than alone. Disentangling these prospects is an interesting avenue for future research.

Beyond creative problem-solving, the "think first, ChatGPT later" approach could potentially be adapted to target other valued learning outcomes too. For instance, learners may benefit from first attempting to generate their own solutions before collaborating with ChatGPT when learning and applying concepts (e.g., Giebl et al., 2021; Wong, 2023; Yap & Wong, 2024), critically thinking and reasoning about information (e.g., Wong, 2025; Wong & Lim, 2019a), and generating creative research questions (e.g., Lim et al., 2024; Wong et al., 2023; Wong & Lim, 2026). Testing the generalizability of the "think first, ChatGPT later" approach across diverse learning goals and with learners of varying levels of expertise (Imundo et al., 2024) would enable insights on when and for whom this approach is effective.

More broadly, future work could examine how the "think first, ChatGPT later" approach can be incorporated in AI literacy programs. Based on Allen and Kendeou's (2024) ED-AI Lit framework, AI literacy includes six components: *knowledge* (i.e., understanding how AI works), *evaluation* (i.e., being able to critically judge AI technologies), *collaboration* (i.e., skills in effectively communicating and collaborating with AI), *contextualization* (i.e., understanding how to use AI as a tool in real-world settings), *autonomy* (i.e., self-determination in actions and decision-making when interacting with AI), and *ethics* (i.e., recognizing and addressing moral issues related to AI technologies). Whereas the "think first, ChatGPT later" approach fosters the collaboration skillset, implementing this approach in tandem with other components of AI literacy could create a more holistic learning experience.

Conclusion

The present research cautions against freely using generative AI tools such as ChatGPT during creative problem-solving, for learners often fail to do so in ways that enhance their learning, instead outsourcing the creative thinking process to the tool. Although such use of ChatGPT increases creative performance on the assisted task at first, this performance boost does not translate into learning gains later on, but rapidly declines once ChatGPT assistance is removed. Thus, to truly reap learning gains that enhance independent human thought and creativity, learners should first conceive their own solutions to the problem, then collaboratively interact with ChatGPT to improve, develop, and evaluate them. In all, generative AI is not to be a substitute for human creativity but, when strategically harnessed, a powerful tool for enhancing it.

Appendix A: Intervention: Product Improvement Task Instructions

Task Overview for All Three Learning Groups.

Mattel is an American toy manufacturer. In terms of sales, it is the second largest toy manufacturer in the world, right after the Lego Group. However, Mattel's goal for this year is to become the largest toy manufacturer in the world. Imagine you have been hired by Mattel as a consultant.

Your task is to come up with your best creative idea to improve an ordinary stuffed bunny, about 30 cm in size, to make it more fun to play with.

A creative idea must meet two criteria—it must be both original AND useful.

How can the bunny be improved to help Mattel achieve higher sales?

Instructions for Human-Only (Control) Group.

Propose your ONE best creative idea to improve an ordinary stuffed bunny, about 30 cm in size, to make it more fun to play with.

Your creative idea should be both original AND useful.

You have 12 min for this task—manage your time wisely.

Write your creative idea below, in as much detail as possible.

Instructions for General-AI Group.

With the assistance of ChatGPT, propose your ONE best creative idea to improve an ordinary stuffed bunny, about 30 cm in size, to make it more fun to play with.

Your creative idea should be both original AND useful.

Use ChatGPT to assist you with this task in any way you like. To get the best responses from ChatGPT, give it clear, detailed prompts and context.

You have 12 min for this task—manage your time wisely.

Write your creative idea below, in as much detail as possible.

Instructions for Regulated-AI Group.

In the next 12 min, propose your ONE best creative idea to improve an ordinary stuffed bunny, about 30 cm in size, to make it more fun to play with.

Your creative idea should be both original AND useful.

You will be guided through 3 phases to complete this task. Please follow all instructions carefully.

Phase 1.

For the next 3 min, generate a list of initial creative ideas to improve an ordinary stuffed bunny, about 30 cm in size, to make it more fun to play with.

These ideas don't need to be perfect—focus on originality AND usefulness. Think freely, and don't worry about refining them just yet. You will later be given time to develop and select your best creative idea.

Write each of your initial creative ideas on a separate line.

Phase 2.

For the next 6 min, interact with ChatGPT to develop your initial creative ideas. To get the best responses from ChatGPT, give it clear, detailed prompts and context.

***Important! ***.

- Use ChatGPT to brainstorm, get feedback, compare, and refine your initial ideas.
- Interact with ChatGPT—treat it like a brainstorming partner rather than a provider of final answers.
- Think critically about and evaluate ChatGPT's suggestions. Ask yourself:

  - Does ChatGPT's feedback make my ideas better?
  - Is it adding something new and/or useful?
  - Do I prefer my initial ideas, or can I combine the merits of both parties' ideas?

Below are some sample prompts for interacting with ChatGPT as a brainstorming and feedback partner. These are just for illustration—feel free to use your own prompts that follow these principles.

- Ask for initial refinements and feedback: "*I need to come up with an original and useful idea to improve an ordinary stuffed bunny. My ideas are XXX and XXX. What can I change or add to make the ideas better?*"
- Ask for more ideas: "*How else can we further improve the ideas to make them even more original and useful?*"

- Elaborate on new ideas that interest you: "*I like your suggestion of XXX—can you expand on it?*"
- Compare your ideas: "*Which of my ideas is/are more creative, and why?*"
- Draw connections between ideas: "*How can I combine these ideas to make a single best creative concept?*"
- Identify and improve weaknesses: "*What's missing from my bunny idea? How can it be improved to be more original and useful?*"
- Refine your idea based on ChatGPT's suggestions, and ask for further feedback: "*Here's my updated idea: [describe changes]. What additional tweaks could we make?*"

You have 6 min for this task—manage your time wisely. Start with the ideas you think have the best potential, but don't fixate on just one!

Here are your initial ideas from Phase 1 earlier: *[XXX]*.

For this phase, you don't need to submit your final response yet, but you may take notes below.

Phase 3.

Finally, you have 3 min to refine and submit your ONE best creative idea to improve an ordinary stuffed bunny, about 30 cm in size, to make it more fun to play with.

Based on your initial ideas and interaction with ChatGPT, decide which aspects you want to keep, change, or add.

Your creative idea should be both original AND useful.

Write your creative idea below, in as much detail as possible.

Appendix B: Post-Intervention: Product Invention Task Instructions

Imagine you have been hired by a global educational company specializing in language-learning games. Your task is to invent a new game that helps adults and university students learn basic foreign language vocabulary for everyday life.

Based on the skills you have learned earlier in the Product Improvement Task:

- On your own, propose your ONE best creative idea that is both original AND useful, making the game fun and effective for learning.
- Develop your idea into a full game design, including specific details (e.g., game title, required materials, gameplay description).

You have 12 min to complete this task—manage your time wisely.

Write your creative idea below, in as much detail as possible.

Appendix C: Alternate Uses Task: 16 Categories for Flexibility Scoring (adapted from George & Wiley, 2020)

- air use.
- arts and crafts.
- boundary/barricade.
- buoyancy.
- clothes/fashion/jewelry.
- containing/storing/holding objects.
- décor.
- exercise/training/sports.
- fun/games/entertainment.
- furniture/house stuff.
- holding/weighing down.
- padding/protection/stabilization.
- recycle/construction material.
- tool.
- vehicle-related.
- weapons/harm/destruction.

Appendix D

Table 7 Sample responses on product improvement task

Originality and Usefulness Ratings	Sample Response
1	"Make it more plump and rounded so that it is more huggable and make its skin smoother so that it feels nicer to touch. Give its eyes eyelashes so that it looks more adorable and put a smile on its face so that users smile when looking at it. Make its skin a light grey or brown so that it looks soft and inviting. Make its ears very nice and long so that users can play with it."
2	"Add a voice button to the bunny so that once the button is pressed, it will generate some voice. The function is customized and programmable, which means that users can record their own sounds. Users can use the bunny to record bedtime stories and educational phrases for their kids, and also wishes and encouragements to their friends."
3	"The chosen creative idea is 'Hoppy Harmony' in which the stuffed bunny has interactive movement where there are different modes like follow mode and dance mode using motion sensors. The bunny also would have educational and musical elements wherein the bunny can sing a variety of nursery rhymes and educational songs for children to learn from."
4	"The bunny can be made to have the ability of understanding and speaking to the player. It can be implemented by using AI. Moreover, make the bunny more 'secret' by adding hidden pockets to their ears or back so that kids can discover hidden features in the toy. In addition, we can use heat-sensitive fabric for the bunny's fur so that when the child holds or pets the bunny, the warmth of their hands will change the color of these patches. Furthermore, we can use reversible design to make it turn into a different animal or character when it is turned inside out, offering more fun and variety in play."
5	"The augmented reality (AR) feature for the stuffed bunny transforms it into a dynamic educational tool, bridging the physical and digital worlds for children. By using a smartphone or tablet, kids can interact with their bunny through an app that brings it to life in various virtual settings. For instance, when pointed at the bunny, the app could overlay an animated environment where the bunny guides children on virtual learning adventures, such as a safari in Africa or a journey through the solar system. This AR experience could teach children about animals, planets, historical events, and more, with interactive quizzes and facts presented by their bunny companion. This immersive, engaging approach not only makes learning more enjoyable but also deepens the child's connection to their toy, encouraging regular educational play sessions."
6	"Add like an online component to the game, so they have a whole online environment for their bunny to play in like Club Penguin. Each virtual bunny will look like their real-world bunny equivalent. In the virtual world they can build like rabbit houses together, and play minigames together, or just explore the rabbit world together. So you can have a social component! So there's a physical component and a virtual component. Physically you can buy lots of different attachments for the bunny. For example, you can buy different hats, and you can buy like crazy things like a car. The car can be maybe magnetically attached to the bunny? Then it can be pushed around like a real toy car! And when you have the car in real life you can also get a virtual version of the car that would make you go faster in the virtual world. You can also buy little houses/dioramas of the locations from the virtual world, that are interactable. Let's say if you have a rabbit jumping obstacle course in the game. You can also buy it in real life and play with it. Maybe we can have different animals as well so it's like an animal world instead of a bunny world. Then can vary the animals bought."

Table 7 (continued)

Originality and Usefulness Ratings	Sample Response
7	"I plan to make a 'scary-cute' Terminator machine bunny. It has interactive learning features like a speaker enhanced with AI so that it functions like Siri and also helps with AR app integration. It has glowing red eyes and detachable robotic arms and mini-tools for the adventure game, where kids can complete missions with their bunny. The theme is 'The Robotic Rebellion': Background: In this storyline, the bunny is a reformed robot from a rogue AI faction. With the child's help, it aims to prevent a rebellion led by other robots who misinterpreted their instructions to take over the home. Missions: 1. Infiltration Adventure • Objective: Sneak into the virtual robot bases (mapped over the house rooms) and gather intel. • AR Feature: Transform rooms into high-tech fortresses using the camera, where kids must avoid virtual lasers and traps. 2. Code Cracking • Objective: Decrypt messages and codes to find out the location of the next rebel meeting. • AR Feature: Interactive code panels appear in AR, and kids use logical thinking to solve them. 3. Alliance Building • Objective: Recruit other toys (through AR) by helping them understand the miscommunication, turning them into allies. • AR Feature: Kids can "scan" other toys in their home, which become characters in the game, offering quests or joining the cause. Each of these plots offers a mix of adventure, technology, and learning, making the AR experience immersive and educational. The overarching theme of cooperation, technology management, and problem-solving not only entertains but also educates in a playful environment."

Appendix E

Table 8 Sample responses on product invention task

Originality and Usefulness Ratings	Sample Response
1	"Game title: Get more food for the animal There are several single-word and A–Z vocabulary cards on the ground, gameplayers would be children who are learning English words. They are separated into two groups. Each group needs to find a word or the combination of letters of a word, as many as possible in a limited time. The word must be a kind of food eaten by an animal. At last, the winner is the team which finds most food words. For every wrong answer, the team members need to eat lemon. Materials: Cards, playground, stopwatch, lemon."
2	"Game title: Language Chef Required materials: Colored pictures of new vocabulary, ChatGPT Gameplay description: In this game, we will guide you to become a master language chef who can cook many different language dishes. First, we will divide all of the players into 2 teams: team A and team B. Then, we will show each of you some vocabulary of this new language with their images in real life. In round 1, each team will have members display the actions that represent the words that we have just shown. In a total of 10 minutes, the team that guesses the most words will receive a reward. In round 2, each team can use an AI chatbot (e.g., ChatGPT) to assist. The organizers will give each team the description of the word in English, and then the task for each team is finding the respective words."
3	"Game Title: Monopoly for a daily-life version written in foreign language Required materials: All the usual materials in the common Monopoly game. In this game, every player is like the person who lives and works in this foreign place … each unit in the monopoly can be some common actions in daily life written in the foreign language, including having meals (breakfast, lunch, dinner), travelling and trips, and renting/buying a house, which are all actions that need to spend money; while working, doing part-time jobs and doing sports are some kinds of actions that we can earn money… Each unit/activity in this map is all written in that foreign language with pronunciation and introduction on it, and the player needs to say it aloud and learn the actual spelling of it, which would be a good way for players to enjoy and learn foreign language vocabulary."
4	"Game title: Pictolingo Required materials: An app with AI features for image recognition Description: This game revolves around the user taking pictures of his/her surroundings and then uploading it for the AI to recognize and generate the translation in the new language, and from there generate more related topics for discussion… For example, take photos of 4 items you use, take photos of 4 places you frequently visit. From there, the user should take photos of whatever items are relevant. This allows the user to have personalized learning as these are words that the individual would most likely need or use, rather than irrelevant words purely generated by the computer. This gives autonomy in learning and might motivate the user as well. Then, after learning these words, it will be put in a game format, where the AI will show pictures of the same objects uploaded by the user and the user has to correctly assign the translation to the image. Then, at higher levels, the AI can ask the user to find 4 pictures of [foreign word; for example, chair in Korean]. This tests the user's ability to apply. The AI can also take relevant small details from the pictures to test users on related things like colors."

Table 8 (continued)

Originality and Usefulness Ratings	Sample Response
5	"Game title: VR simulation of real-life foreign language interactions. Rationale: The biggest challenge in picking up a foreign language and even be close to mastering it for use in one's daily life is the lack of opportunities in real life to practice the language. Required materials: VR headset Gameplay: Upon entering the VR game setting, players will be placed in a virtual environment that closely mimics the city environment with shops, school etc. where the characters can go on with their day-to-day life. Players can interact with the NPCs just like how they would in real life, except the conversations will be in foreign language. Initially, players can verbalize their thoughts and wants in their home language and there will be guidance by the system on the proper phrasing in the selected foreign language, with teaching of proper pronunciation. Players will then repeat the foreign language to the NPC to elicit a response. Real-time feedback will be given on the accuracy of pronunciation and sentence grammar. Through this, players will be able to pick up useful day-to-day sentences that they can use when actually conversing with someone else in that foreign language in their real lives. This rinse and repeat method of learning will allow the language picked up to be ingrained in their minds. The NPCs are also not limited by the type of speech input by the player and can adapt to any situations and sentences that are befitting of their character (e.g., barista, doctor, parents). Players will basically be able to live a virtual life that situates him or her in a setting where conversations are all held in that foreign language."
6	"Game Title: Living in a Foreign Country—Simulator Required materials: Computer, Unreal Engine or something that can feasibly help to program and visualize the game, programing knowledge, access to people from all over the world willing to help contribute information and expertise to the game. Game is an online simulator game in which the player will be playing a character who recently moved to a foreign country (which has the language they are meant to be learning) and has to go about their daily life trying to interact with other nonplayable characters (NPC) on asking for directions, buying food, making new friends etc. Dialogues of the NPCs can be based off real world conversations that actually take place, making the player feel more immersed. At easier levels, player can choose the right option out of a few choices (Multiple Choice Question style). At harder levels, players will have to type out their responses (to be determined by an algorithm on its suitability as its response) and the NPCs will answer based on the typed response. Additional tools that can be used to help support the player include example responses and a dictionary that the player can use to help search up particular words if they are struggling to create their responses. Giving the wrong response will cause the player's popularity and comfort level to decrease and their goal should be to maintain or increase their comfort level or else their character in game gives up and returns back to their home country. In the game, the player spends each day with a set of tasks to complete before the end of the day, and will have to travel around to various places in the game map (supermarket, pharmacy, library, school etc.) to accomplish those tasks to earn money which they can spend on upgrades for their character to increase their comfort level, and reaching a certain level helps them to beat the game. Deepening their connections with their neighbors/new friends/NPCs also helps to increase their comfort level. Occasionally, special events will take place, giving the player more opportunities to earn more comfort/popularity by spending time with their community/immersing themselves in the culture of the place, making the game more exciting and including a variety of activities."

Table 8 (continued)

Originality and Usefulness Ratings	Sample Response
7	"The game title will be 'Alien Invasion of [country name, e.g., Japan]'. The game should be played from the POV of an alien that lands on earth in that country with no money and has to figure out the local language by induction so that they can buy food (they are hungry!) and slowly blend into life there. It will be more engaging if it develops a plot where the supporting characters are helpful and kind and have funny interactions where the user learns an important word or phrase (e.g., 'is there a washroom nearby?' or 'what is your name?'), maybe even eventually having a love interest or found family. There should be in-game currency that users earn from answering (typing on in-app keyboard or saying verbally to the phone e.g., how you talk to Siri) prompts at the correct moments, such that they can save up to buy clothes and other fun daily stuff like a house. It should be played on the phone in portrait mode, somewhat similar to the Sims or Mystic Messenger, such that it really feels like AR integration into reality except it's fully graphics. … A possible plot twist when the user has reached the intermediate/advanced stage of language learning is for another 'alien' to land on the same planet, such that the intermediate/advanced user is effectively teaching the beginner user and drawing upon real-user interactions to solidify language learning takeaways and have real conversations, bringing that human element into the game and making it less programed. There can be fun Easter eggs related to that country such as Godzilla emerging in Shinjuku, Tokyo in an apocalyptic scenario for Japan (in the Japanese learning mode), then they'll have to learn words about monsters or earthquakes and disaster preparedness. Can have some battle modes with Super Smash Bros graphics where users face off against each other and compete to see who can get the most answers correct in a battle to glean 'experience', and winner gets in-game currency. 'Experience' can eventually be exchanged for in-game currency. The game ends with a UFO abduction/alien full-on invasion so that everyone can come to learn the language and the way of the humans."

Acknowledgements The authors thank Zhi Wang for her assistance with data collection.

Author contributions Sarah Shi Hui Wong served as lead for resources, supervision, writing–original draft, and writing–review and editing, contributed equally to conceptualization, methodology, and funding acquisition, and served in a supporting role for data curation and formal analysis.
Sophia Xuefei Qiu served as lead for investigation, data curation, and formal analysis, contributed equally to conceptualization, methodology, and funding acquisition, and served in a supporting role for writing–original draft.

Funding This research was supported by a Singapore Management University grant awarded to Sarah Shi Hui Wong, and a Yale-NUS College Capstone Research grant awarded to Sophia Xuefei Qiu.

Data Availability The materials for this study are available in the Appendices. The data and analysis code that support the findings of this study are available via the Open Science Framework: https://osf.io/t7an8/overview?view_only=fc490583cdc8458e8b895b29e4df2345.

Declarations

Ethics Approval This research was conducted with the appropriate ethics approval from the National University of Singapore's institutional review board.

Consent to Participate Informed consent was obtained from all participants included in the study.

Competing interests 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.

Reproducibility Note During the intervention in this study, the general-AI and regulated-AI groups used the GPT-4 version of ChatGPT via the web interface (https://chatgpt.com/) that was available from OpenAI at the time through a ChatGPT Plus subscription. Default parameters in the interface were used, except that memory/model training was switched off.

References

Acar, S., & Runco, M. A. (2019). Divergent thinking: New methods, recent research, and extended theory. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 153–158. https://doi.org/10.1037/aca0000231

Acar, S., Runco, M. A., & Park, H. (2020). What should people be told when they take a divergent thinking test? A meta-analytic review of explicit instructions for divergent thinking. *Psychology of Aesthetics, Creativity, and the Arts, 14*(1), 39–49. https://doi.org/10.1037/aca0000256

Allen, L. K., & Kendeou, P. (2024). ED-AI lit: An interdisciplinary framework for AI literacy in education. *Policy Insights From the Behavioral and Brain Sciences, 11*(1), 3–10. https://doi.org/10.1177/23727322231220339

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43*(5), 997–1013. https://doi.org/10.1037/0022-3514.43.5.997

Baer, J., & Kaufman, J. C. (2019). Assessing creativity with the consensual assessment technique. In I. Lebuda & V. P. Glăveanu (Eds.), *The Palgrave handbook of social creativity research* (pp. 27–37). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-95498-1_3

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking beyond the hype: Understanding the effects of AI on learning. *Educational Psychology Review, 37*, Article 45. https://doi.org/10.1007/s10648-025-10020-8

Beghetto, R. A. (2023). A new horizon for possibility thinking: A conceptual case study of human × AI collaboration. *Possibility Studies & Society, 1*(3), 324–341. https://doi.org/10.1177/27538699231160136

Beghetto, R. A., & Karwowski, M. (2017). Toward untangling creative self-beliefs. In M. Karwowski & J. C. Kaufman (Eds.), *The creative self: Effect of beliefs, self-efficacy, mindset, and identity* (pp. 3–22). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-809790-8.00001-7

Benedek, M., & Lebuda, I. (2025). Managing your muse: Exploring three levels of metacognitive control in creative ideation. *Creativity Research Journal, 37*(4), 570–581. https://doi.org/10.1080/10400419.2024.2347770

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). The MIT Press.

Boers, J., Etty, T., Baars, M., & van Broekhoven, K. (2025). Exploring cognitive strategies in human-AI interaction: ChatGPT's role in creative tasks. *Journal of Creativity, 35*(1), Article 100095. https://doi.org/10.1016/j.yjoc.2025.100095
Ceh, S. M., Rafner, J., & Benedek, M. (2024). Creativity in digitally mediated times: How digital tools support creativity across domains. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000729
Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823
Clapham, M. M. (1997). Ideational skills training: A key element in creativity training programs. *Creativity Research Journal, 10*(1), 33–44. https://doi.org/10.1207/s15326934crj1001_4
Cortes, R. A., Weinberger, A. B., Daker, R. J., & Green, A. E. (2019). Re-examining prominent measures of divergent and convergent creativity. *Current Opinion in Behavioral Sciences, 27*, 90–93. https://doi.org/10.1016/j.cobeha.2018.09.017
Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal*, *18*(3), 391–404. https://doi.org/10.1207/s15326934crj1803_13
Cseh, G. M., & Jeffries, K. K. (2019). A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 159–166. https://doi.org/10.1037/aca0000220
Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education, 210*, Article 104967. https://doi.org/10.1016/j.compedu.2023.104967
de Chantal, P.- L., & Organisciak, P. (2025). Automated feedback and creativity: On the role of metacognitive monitoring in divergent thinking. *Psychology of Aesthetics, Creativity, and the Arts, 19*(5), 1064–1083. https://doi.org/10.1037/aca0000592
Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224
Dumas, D., Organisciak, P., Maio, S., & Doherty, M. (2021). Four text-mining methods for measuring elaboration. *The Journal of Creative Behavior, 55*(2), 517–531. https://doi.org/10.1002/jocb.471
Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence, 6*, Article 100198. https://doi.org/10.1016/j.caeai.2023.100198
Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146
Forthmann, B., Oyebade, O., Ojo, A., Günther, F., & Holling, H. (2019). Application of latent semantic analysis to divergent thinking is biased by elaboration. *The Journal of Creative Behavior*, *53*(4), 559–575. https://doi.org/10.1002/jocb.240
Garcia, M. B. (2025). The paradox of artificial creativity: Challenges and opportunities of generative AI artistry. *Creativity Research Journal, 37*(4), 755–768. https://doi.org/10.1080/10400419.2024.2354622
George, T., & Wiley, J. (2020). Need something different? Here's what's been done: Effects of examples and task instructions on creative idea generation. *Memory & Cognition, 48*(2), 226–243. https://doi.org/10.3758/s13421-019-01005-4
Giebl, S., Mena, S., Sandberg, R., Bjork, E. L., & Bjork, R. A. (2023). Thinking first versus googling first: Preferences and consequences. *Journal of Applied Research in Memory and Cognition*, *12*(3), 431–442. https://doi.org/10.1037/mac0000072
Giebl, S., Mena, S., Storm, B. C., Bjork, E. L., & Bjork, R. A. (2021). Answer first or Google first? Using the internet in ways that enhance, not impair, one's subsequent retention of needed information. *Psychology Learning & Teaching*, *20*(1), 58–75. https://doi.org/10.1177/1475725720961593
Gilhooly, K. (2024). AI vs humans in the AUT: Simulations to LLMs. *Journal of Creativity, 34*(1), Article 100071. https://doi.org/10.1016/j.yjoc.2023.100071

Gilhooly, K. J., Fioratou, E., Anthony, S. H., & Wynn, V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology, 98*(4), 611–625. https://doi.org/10.1111/j.2044-8295.2007.tb00467.x

Gonthier, C., & Besançon, M. (2024). It is not always better to have more ideas: Serial order and the trade-off between fluency and elaboration in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts, 18*(4), 480–492. https://doi.org/10.1037/aca0000485

Greene, J. A., Freed, R., & Sawyer, R. K. (2019). Fostering creative performance in art and design education via self-regulated learning. *Instructional Science, 47*(2), 127–149. https://doi.org/10.1007/s11251-018-9479-8

Guilford, J. P. (1967a). Creativity: Yesterday, today, and tomorrow. *The Journal of Creative Behavior*, *1*(1), 3–14. https://doi.org/10.1002/j.2162-6057.1967.tb00002.x

Guilford, J. P. (1967b). *The nature of human intelligence*. McGraw-Hill.

Haase, J., & Hanel, P. H. P. (2023). Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *Journal of Creativity, 33*(3), Article 100066. https://doi.org/10.1016/j.yjoc.2023.100066

Haase, J., Hoff, E. V., Hanel, P. H. P., & Innes-Ker, Å. (2018). A meta-analysis of the relation between creative self-efficacy and different creativity measurements. *Creativity Research Journal*, *30*(1), 1–16. https://doi.org/10.1080/10400419.2018.1411436

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Huber, S. E., Kiili, K., Nebel, S., Ryan, R. M., Sailer, M., & Ninaus, M. (2024). Leveraging the potential of large language models in education through playful and game-based learning. *Educational Psychology Review, 36*, Article 25. https://doi.org/10.1007/s10648-024-09868-z

Hubert, K. F., Awa, K. N., & Zabelina, D. L. (2024). The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports, 14*, Article 3440. https://doi.org/10.1038/s41598-024-53303-w

Imundo, M. N., Watanabe, M., Potter, A. H., Gong, J., Arner, T., & McNamara, D. S. (2024). Expert thinking with generative chatbots. *Journal of Applied Research in Memory and Cognition, 13*(4), 465–484. https://doi.org/10.1037/mac0000199

Jaarsveld, S., & van Leeuwen, C. (2005). Sketches from a design process: Creative cognition inferred from intermediate products. *Cognitive Science, 29*(1), 79–101. https://doi.org/10.1207/s15516709cog2901_4

Jeffries, K. K. (2024). Measuring creativity with the Consensual Assessment Technique (CAT). In L. J. Ball & F. Vallée-Tourangeau (Eds.), *The Routledge international handbook of creative cognition* (pp. 17–31). Routledge. https://doi.org/10.4324/9781003009351-3

Karwowski, M., Lebuda, I., & Beghetto, R. A. (2019). Creative self-beliefs. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (2nd ed., pp. 396–417). Cambridge University Press. https://doi.org/10.1017/9781316979839.021

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kaufman, J. C., & Beghetto, R. A. (2013). In praise of Clark Kent: Creative metacognition and the importance of teaching kids when (not) to be creative. *Roeper Review, 35*(3), 155–165. https://doi.org/10.1080/02783193.2013.799413

Kaufman, J. C., & Glăveanu, V. P. (2019). A review of creativity theories: What questions are we trying to answer? In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (2nd ed., pp. 27–43). Cambridge University Press. https://doi.org/10.1017/9781316979839.004

Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior, 114*, Article 106553. https://doi.org/10.1016/j.chb.2020.106553

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989–998. https://doi.org/10.1037/a0015729

Lebuda, I., & Benedek, M. (2025). Contributions of metacognition to creative performance and behavior. *The Journal of Creative Behavior, 59*(1), Article e652. https://doi.org/10.1002/jocb.652

Lee, B. C., & Chung, J. (2024). An empirical investigation of the impact of ChatGPT on creativity. *Nature Human Behaviour, 8*(10), 1906–1914. https://doi.org/10.1038/s41562-024-01953-1
Li, L. (2025). Hybrid ideas of artificial intelligence and human: Growth of creativity in college educational programs. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000750. Advance online publication.
Lim, S. W. H., Wong, S. S. H., & Visessuvanapoom, P. (2024). Durable benefits of learning-by-teaching for research question generation performance: A field experiment. *The Journal of Experimental Education*. https://doi.org/10.1080/00220973.2024.2364625. Advance online publication.
Lubart, T. I. (2001). Models of the creative process: Past, present and future. *Creativity Research Journal, 13*(3–4), 295–308. https://doi.org/10.1207/S15326934CRJ1334_07
Marrone, R., Cropley, D., & Medeiros, K. (2026). How does narrow AI impact human creativity? *Creativity Research Journal, 38*(1), 150–160. https://doi.org/10.1080/10400419.2024.2378264
McGuire, J., De Cremer, D., & Van de Cruys, T. (2024). Establishing the importance of co-creation and self-efficacy in creative collaboration with artificial intelligence. *Scientific Reports, 14*, Article 18525. https://doi.org/10.1038/s41598-024-69423-2
McMahon, K., Ruggeri, A., Kämmer, J. E., & Katsikopoulos, K. V. (2016). Beyond idea generation: The power of groups in developing ideas. *Creativity Research Journal, 28*(3), 247–257. https://doi.org/10.1080/10400419.2016.1195637
Mei, Q., Xie, Y., Yuan, W., & Jackson, M. O. (2024). A Turing test of whether AI chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences of the United States of America, 121*(9), Article e2313925121. https://doi.org/10.1073/pnas.2313925121
Mumford, M. D., Lonergan, D. C., & Scott, G. (2002). Evaluating creative ideas: Processes, standards, and context. *Inquiry: Critical Thinking Across the Disciplines, 22*(1), 21–30. https://doi.org/10.5840/inquiryctnews20022213
Mumford, M. D., Mobley, M. I., Uhlman, C. E., Reiter-Palmon, R., & Doares, L. M. (1991). Process analytic models of creative capacities. *Creativity Research Journal, 4*(2), 91–122. https://doi.org/10.1080/10400419109534380
Mumford, M. D., Schultz, R. A., & Van Doorn, J. R. (2001). Performance in planning: Processes, requirements, and errors. *Review of General Psychology, 5*(3), 213–240. https://doi.org/10.1037/1089-2680.5.3.213
Nah, F. F.-H., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research, 25*(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814
Nijstad, B. A. (2015). Creativity in groups. In M. Mikulincer, P. R. Shaver, J. F. Dovidio, & J. A. Simpson (Eds.), *APA handbook of personality and social psychology, Vol. 2: Group processes* (pp. 35–65). American Psychological Association. https://doi.org/10.1037/14342-002
Osburn, H. K., & Mumford, M. D. (2006). Creativity and planning: Training interventions to develop creative problem-solving skills. *Creativity Research Journal, 18*(2), 173–190. https://doi.org/10.1207/s15326934crj1802_4
Paulus, P. B. (2000). Groups, teams, and creativity: The creative potential of idea-generating groups. *Applied Psychology, 49*(2), 237–262. https://doi.org/10.1111/1464-0597.00013
Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist, 39*(2), 83–96. https://doi.org/10.1207/s15326985ep3902_1
Puente-Díaz, R., & Cavazos-Arroyo, J. (2017). Creative self-efficacy: The influence of affective states and social persuasion as antecedents and imagination and divergent thinking as consequences. *Creativity Research Journal, 29*(3), 304–312. https://doi.org/10.1080/10400419.2017.1360067
Puente-Díaz, R., Cavazos-Arroyo, J., & Puerta-Sierra, L. (2021). Idea generation, selection, and evaluation: A metacognitive approach. *Journal of Creative Behavior, 55*(4), 1015–1027. https://doi.org/10.1002/jocb.505
Puente-Díaz, R., Cavazos-Arroyo, J., & Vargas-Barrera, F. (2021). Metacognitive feelings as a source of information in the evaluation and selection of creative ideas. *Thinking Skills and Creativity, 39*, Article 100767. https://doi.org/10.1016/j.tsc.2020.100767
Puryear, J. S. (2016). Inside the creative sifter: Recognizing metacognition in creativity development. *Journal of Creative Behavior, 50*(4), 321–332. https://doi.org/10.1002/jocb.80
Rawlings, B. S., Chetwynd-Talbot, D., Husband, E., Nuttall, A., Quinn, E., Taggart, R., & Roome, H. E. (2025). Divergent thinking is linked with convergent thinking; implications for models of creativity. *Thinking & Reasoning, 31*(4), 586–608. https://doi.org/10.1080/13546783.2025.2485059

Rhodes, M. (1961). An analysis of creativity. *Phi Delta Kappan, 42*(7), 305–310.
Ritter, S. M., & Mostert, N. (2017). Enhancement of creative thinking skills using a cognitive-based creativity training. *Journal of Cognitive Enhancement, 1*(3), 243–253. https://doi.org/10.1007/s41465-016-0002-3
Rubenstein, L. D., Callan, G. L., & Ridgley, L. M. (2018). Anchoring the creative process within a self-regulated learning framework: Inspiring assessment methods and future research. *Educational Psychology Review, 30*(3), 921–945. https://doi.org/10.1007/s10648-017-9431-5
Runco, M. A. (2004). Creativity. *Annual Review of Psychology, 55*, 657–687. https://doi.org/10.1146/annurev.psych.55.090902.141502
Runco, M. A. (2025). Updating the standard definition of creativity to account for the artificial creativity of AI. *Creativity Research Journal, 37*(1), 1–5. https://doi.org/10.1080/10400419.2023.2257977
Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66–75. https://doi.org/10.1080/10400419.2012.652929
Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092
Rus, V., & Kendeou, P. (2025). Are LLMs actually good for learning? *AI & Society, 40*(7), 5617–5618. https://doi.org/10.1007/s00146-025-02323-9
Saretzki, J., Forthmann, B., & Benedek, M. (2024). A systematic quantitative review of divergent thinking assessments. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. https://doi.org/10.1037/aca0000691
Sawyer, R. K. (2019). The role of failure in learning how to create in art and design. *Thinking Skills and Creativity, 33*, Article 100527. https://doi.org/10.1016/j.tsc.2018.08.002
Sawyer, R. K., & Henriksen, D. (2023). *Explaining creativity: The science of human innovation* (3rd ed). Oxford University Press. https://doi.org/10.1093/oso/9780197747537.001.0001
Schwartz, D. L., & Lin, X. (2001). Computers, productive agency, and the effort after shared meaning. *Journal of Computing in Higher Education, 12*(2), 3–33. https://doi.org/10.1007/BF02940954
Scott, G., Leritz, L. E., & Mumford, M. D. (2004). The effectiveness of creativity training: A quantitative review. *Creativity Research Journal, 16*(4), 361–388. https://doi.org/10.1080/10400410409534549
Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10*(2), 176–199. https://doi.org/10.1177/1745691615569000
Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, Article 108386. https://doi.org/10.1016/j.chb.2024.108386
Stroebe, W., Nijstad, B. A., & Rietzschel, E. F. (2010). Beyond productivity loss in brainstorming groups: The evolution of a question. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology, Vol. 43* (pp. 157–203). Academic Press. https://doi.org/10.1016/S0065-2601(10)43004-X
Strom, R. D., & Strom, P. S. (2002). Changing the rules: Education for creative thinking. *Journal of Creative Behavior, 36*(3), 183–200. https://doi.org/10.1002/j.2162-6057.2002.tb01063.x
Sun, S., Li, Z. A., Foo, M.-D., Zhou, J., & Lu, J. G. (2025). How and for whom using generative AI affects creativity: A field experiment. *Journal of Applied Psychology, 110*(12), 1561–1573. https://doi.org/10.1037/apl0001296
Suriano, R., Plebe, A., Acciai, A., & Fabio, R. A. (2025). Student interaction with ChatGPT can promote complex critical thinking skills. *Learning and Instruction, 95*, Article 102011. https://doi.org/10.1016/j.learninstruc.2024.102011
Tierney, P., & Farmer, S. M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management Journal, 45*(6), 1137–1148. https://doi.org/10.2307/3069429
Torrance, E. P. (2008). *Torrance tests of creative thinking: Norms-technical manual, verbal forms A and B*. Scholastic Testing Service.
Urban, K., & Urban, M. (2023). How can we measure metacognition in creative problem-solving? Standardization of the MCPS scale. *Thinking Skills and Creativity, 49*, Article 101345. https://doi.org/10.1016/j.tsc.2023.101345
Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education, 215*, Article 105031. https://doi.org/10.1016/j.compedu.2024.105031

Urban, M., Lukavský, J., Brom, C., Hein, V., Svacha, F., Děchtěrenko, F., & Urban, K. (2025). Prompting for creative problem-solving: A process-mining study. *Learning and Instruction, 99*, Article 102156. https://doi.org/10.1016/j.learninstruc.2025.102156

Urban, M., & Urban, K. (2021). Unskilled but aware of it? Cluster analysis of creative metacognition from preschool age to early adulthood. *Journal of Creative Behavior, 55*(4), 937–945. https://doi.org/10.1002/jocb.499

Urban, M., & Urban, K. (2025). Do we need metacognition for creativity? A necessary condition analysis of creative metacognition. *Psychology of Aesthetics, Creativity, and the Arts, 19*(6), 1467–1483. https://doi.org/10.1037/aca0000647

van Gog, T., Hoogerheide, V., & van Harsel, M. (2020). The role of mental effort in fostering self-regulated learning with problem-solving tasks. *Educational Psychology Review, 32*(4), 1055–1072. https://doi.org/10.1007/s10648-020-09544-y

Vernon, D., Hocking, I., & Tyler, T. C. (2016). An evidence-based review of creative problem-solving tools: A practitioner's resource. *Human Resource Development Review*, *15*(2), 230–259. https://doi.org/10.1177/1534484316641512

Vinchon, F., Lubart, T., Bartolotta, S., Gironnay, V., Botella, M., Bourgeois-Bougrine, S., Burkhardt, J.-M., Bonnardel, N., Corazza, G. E., Glăveanu, V., Hanchett Hanson, M., Ivcevic, Z., Karwowski, M., Kaufman, J. C., Okada, T., Reiter-Palmon, R., & Gaggioli, A. (2023). Artificial intelligence & creativity: A manifesto for collaboration. *Journal of Creative Behavior, 57*(4), 472–484. https://doi.org/10.1002/jocb.597

White, J., Fu, O., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. In J. Yoder, R. Gabriel, V. Vranić, & K. Brown (Eds.), *PLoP '23: Proceedings of the 30th Conference on Pattern Languages of Programs* (Article 5). The Hillside Group.

Wong, S. S. H. (2023). Deliberate erring improves far transfer of learning more than errorless elaboration and spotting and correcting others' errors. *Educational Psychology Review*, *35*(1), Article 16. https://doi.org/10.1007/s10648-023-09739-z

Wong, S. S. H. (2025). Learning by teaching with deliberate errors promotes argumentative reasoning. *Journal of Educational Psychology, 117*(7), 1013–1038. https://doi.org/10.1037/edu0000934

Wong, S. S. H., Lim, K. Y. L., & Lim, S. W. H. (2023). To ask better questions, teach: Learning-by-teaching enhances research question generation more than retrieval practice and concept-mapping. *Journal of Educational Psychology, 115*(6), 798–812. https://doi.org/10.1037/edu0000802

Wong, S. S. H., & Lim, S. (2026). The powerful teacher: A power hypothesis for the benefits of learning-by-teaching. *Educational Psychology Review*, *38*(1), Article 26. https://doi.org/10.1007/s10648-025-10097-1

Wong, S. S. H., & Lim, S. W. H. (2019a). From JOLs to JOLs+: Directing learners' attention in retrieval practice to boost integrative argumentation. *Journal of Experimental Psychology: Applied, 25*(4), 543–557. https://doi.org/10.1037/xap0000225

Wong, S. S. H., & Lim, S. W. H. (2019b). Prevention–permission–promotion: A review of approaches to errors in learning. *Educational Psychologist, 54*(1), 1–19. https://doi.org/10.1080/00461520.2018.1501693

Wong, S. S. H., & Lim, S. W. H. (2022a). Deliberate errors promote meaningful learning. *Journal of Educational Psychology*, *114*(8), 1817–1831. https://doi.org/10.1037/edu0000720

Wong, S. S. H., & Lim, S. W. H. (2022b). The derring effect: Deliberate errors enhance learning. *Journal of Experimental Psychology: General*, *151*(1), 25–40. https://doi.org/10.1037/xge0001072

Wong, S. S. H., & Lim, S. W. H. (2023). Take notes, not photos: Mind-wandering mediates the impact of note-taking strategies on video-recorded lecture learning performance. *Journal of Experimental Psychology: Applied*, *29*(1), 124–135. https://doi.org/10.1037/xap0000375

Wu, Z., Ji, D., Yu, K., Zeng, X., Wu, D., & Shidujaman, M. (2021). AI creativity and the human–AI co-creation model. In M. Kurosu (Ed.), *Lecture notes in computer science: Vol. 12762. Human-computer interaction. Theory, methods and tools* (pp. 171–190). Springer. https://doi.org/10.1007/978-3-030-78462-1_13

Yan, L., Greiff, S., Lodge, J. M., & Gašević, D. (2025). Distinguishing performance gains from learning when using generative AI. *Nature Reviews Psychology, 4*(7), 435–436. https://doi.org/10.1038/s44159-025-00467-5

Yap, J. B. K., & Wong, S. S. H. (2024). Deliberately making and correcting errors in mathematical problem-solving practice improves procedural transfer to more complex problems. *Journal of Educational Psychology*, *116*(7), 1112–1128. https://doi.org/10.1037/edu0000850

Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence, 4*, Article 100147. https://doi.org/10.1016/j.caeai.2023.100147

Zamfirescu-Pereira, J. D., Wong, R., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In A. Schmidt, K. Väänänen, T. Goval, P. O. Kristensson, A. Peters, S. Mueller, J. R. Williamson, & M. L. Wilson (Eds.), *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Article 437). Association for Computing Machinery. https://doi.org/10.1145/3544548.3581388

Zielińska, A., Lebuda, I., Czerwonka, M., & Karwowski, M. (2025). Self-regulation prompts improve creative performance. *The Journal of Creative Behavior, 59*(1), Article e674. https://doi.org/10.1002/jocb.674

Authors and Affiliations

Sarah Shi Hui Wong[1] · **Sophia Xuefei Qiu**[2]

✉ Sarah Shi Hui Wong
sarahwong@smu.edu.sg

[1] School of Social Sciences, Singapore Management University, 10 Canning Rise, #05-01, Singapore 179873, Singapore

[2] Division of Social Sciences, Yale-NUS College, Singapore, Singapore