Thinking Skills and Creativity 60 (2026) 102116

Contents lists available at ScienceDirect

Thinking Skills and Creativity

journal homepage: www.elsevier.com/locate/tsc

Augmentation or limitation? Generative AI's influence on college students' creative problem-solving

Huanhuan Wang [a,1], Ting Da [a,1], Yuanrong Wang [a], Fenghao Luo [a], Yifan Zhang [a,*], Ahmed Tlili [b], Dong Yang [c], Yarong Wang [b], Jiaxin Xu [b], Xixian Zhu [d], Man Wan [b], Ronghuai Huang [a]

[a] *National Engineering Research Center of Cyberlearning and Intelligent Technology, Beijing Normal University, Jia No. 2 Manjing Road, Changping District, Beijing 102206, China*
[b] *Smart Learning Institute, Beijing Normal University, No. 12 Xueyuan South Road, Haidian District, Beijing 100082, China*
[c] *College of Education for the Future, Beijing Normal University, No.18, Jinfeng Road, Tangjiawan, Zhuhai City, Guangdong Province 519087, China*
[d] *Educational Research Centre, Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hung Hom, Hong Kong, China*

ARTICLE INFO

Keywords:
Generative AI
Creative problem-solving
Mixed-methods study
Human-AI collaboration
Educational equity

ABSTRACT

Generative artificial intelligence (GenAI) has been increasingly used in teaching and learning. However, its effect on the learning outcome of creative problem-solving (CPS) remains inconclusive. To address this problem, this mixed-methods study examines the dual role of GenAI in college students' CPS using a natural experiment and a post-interview. By comparing outcomes of a design competition between non-GenAI student users (Year 2023, $N_{teams} = 10$; $N_{students} = 56$) and GenAI users (Year 2024, $N_{teams} = 18$; $N_{students} = 80$) at three stages ($N_{pre+mid+post} = 82$), supplemented by competition mentor interviews analysis, we found that GenAI is significantly positive-associated with students' CPS performance ($p < 0.001$), with greater gains among students with lower baseline creativity. Additionally, not only did students' CPS performance improve significantly across the three testing stages ($p < 0.001$), but also the trajectory of improvement differed markedly between cohorts: students with ChatGPT support exhibited steeper gains across stages ($p < 0.001$). However, the content analysis of mentor interviews identified risks of idea homogenization that emerged from the submitted work and students' over-reliance on GenAI tools, with students failing to critically validate AI outputs. Mentor insights further highlight the mediating role of AI literacy and team dynamics: Teams with balanced human-AI collaboration achieved more innovation than those overly dependent on GenAI. These findings underscore GenAI's paradoxical impact: a catalyst for stimulating creativity and a potential inhibitor of independent thinking. These results suggest that teachers and instructional designers should adopt active learning-oriented instructional design strategies to balance student-AI collaboration on creativity. The study advances research discourse on human-AI co-creativity while offering actionable strategies for sustainable AI integration in education.

* Corresponding author.
E-mail addresses: zyf6225041@gmail.com, zhangyifan@bnu.edu.cn (Y. Zhang).
[1] Huanhuan Wang and Ting Da made an equal contribution to this study.

https://doi.org/10.1016/j.tsc.2025.102116
Received 9 April 2025; Received in revised form 18 December 2025; Accepted 21 December 2025
Available online 23 December 2025

1. Introduction

The rapid advancement of generative artificial intelligence (GenAI) has caused both enthusiasm and skepticism in educational research. While GenAI tools like ChatGPT demonstrate unprecedented capabilities in idea generation and problem-solving support (Fiialka et al., 2024; Urban et al., 2024), a set of evidence suggests their integration in learning may inadvertently undermine critical cognitive skills (Dwivedi et al., 2023; Zhai et al., 2024). This tension is particularly salient in creative problem-solving (CPS), where AI's role can range from a catalyst for innovation to a crutch that fosters dependency.

The existing literature yields mixed results regarding GenAI's educational impact. Proponents highlight its potential to foster creativity through instant access to diverse perspectives (Hwang & Won, 2021) and to enhance solution quality (Al-Zahrani, 2024). Conversely, critics warn of homogenized outputs (Lim et al., 2023), reduced metacognitive engagement (Habib et al., 2024), and sociocultural biases in AI-generated content (Essel et al., 2024). Crucially, these studies predominantly examine AI's effects in controlled settings, leaving a critical gap in understanding how GenAI shapes CPS processes in authentic educational contexts.

To address this gap, we designed our investigation with a natural experiment component in a realistic design competition activity. We followed Rhodes' (1961) Four Perspectives on Creativity (4Ps Model), which provides a holistic framework for analyzing Person (learner characteristics), Process (application strategies), Press (environmental factors, including AI tools), and Product (creative outcomes). While previous research has primarily focused on AI's impact on creative products, applying the 4Ps Model in this study enables a systemic examination of how GenAI reconfigures the entire creative ecosystem in a realistic setting, from individual features, GenAI application strategy, creative process, and final product.

This study investigates GenAI's dual role in an international design competition, comparing CPS outcomes between GenAI-assisted (year 2024) and non-GenAI (year 2023) student teams. Through mixed methods integrating natural experiment and a qualitative analysis of mentor observation and perceptions, we address three unresolved research questions:

1. How is the GenAI application associated with CPS's quality and trajectory across different design stages?
2. Can GenAI mitigate creativity disparities among students with varying baseline abilities?
3. What individual and team-level factors mediate the effectiveness of human-AI collaboration?

2. Literature Review

2.1. Theoretical lens of creative problem solving

CPS was proposed in the 1950s and has become a universal innovation concept across various fields. It was defined as a structured process that separates idea generation from evaluation, integrates divergent and convergent thinking, aims to solve complex problems without clear answers, and generates novel and practical solutions (Parnes, 1985). CPS can be applied across multiple fields, such as education, technology, business, and engineering (Kovač et al., 2023). CPS provides a procedure-oriented framework for transforming abstract creativity into concrete actions (Mumford et al., 1991).

Creativity has been a subject of extensive research since the early 20th century, with scholars proposing various theories and models. For example, Rhode (1961) proposed the Four Perspectives on Creativity (4Ps Model) as one well-known theoretical framework. The 4Ps Model encompasses the dimensions of Person, Process, Press, and Product. The Person dimension focuses on the intrinsic characteristics of individuals involved in the creative process, including personality, intellect, and other psychological aspects. The Process dimension examines the cognitive and procedural aspects of creativity, such as idea generation and problem-solving strategies. The Press dimension considers both the immediate physical environment and the broader socio-cultural context. Finally, the Product dimension evaluates the originality, usefulness, and aesthetic value of creative outputs (Rhode, 1961).

The 4Ps Model offers a holistic view of the creative ecosystem, integrating all key elements, unlike models that focus only on specific aspects. It also balances personal and external factors, recognizing their interaction in fostering or hindering creativity. The 4Ps Model is widely applied in education to increase students' creativity in various education domains, such as information science education (Couger, 1993), young children's STEM education (Tippett & Gonzalez, 2022), and music education (Wen, 2024).

In this study, guided by the 4P Model's *Person* dimension, we collected data on students' demographic questionnaires and human-AI interaction records during the project development. The *Process* dimension informs us to analyze their project development process. Then, we investigated the differential use of GenAI tools across groups in our study, highlighting their role as an environmental factor corresponding to the *Press* dimension. Last, we focused on the student projects assessed by experts for the *Product* dimension. We also incorporated the Process Model of Creativity (Inclusive Innovation, 2017) to further enhance the granularity of the activity design, enriching the content of each stage of creative problem-solving: clarification, ideation, development, and implementation. This process model serves as a reference for analyzing the interview data to understand how students used GenAI during CPS.

2.2. Related studies

2.2.1. CPS's influencing factors

A series of individual traits, cognitive processes, environmental support, and cross-cultural contexts can influence CPS. First, for individual traits, intellectual capacity, domain-specific knowledge, and intrinsic motivation form the foundation for innovation. Sternberg (2006) emphasizes the effect of IQ and the interplay of synthetic, analytical, and practical intelligence in solution generation. Intrinsic motivation, as highlighted by Amabile et al. (2018), consistently outperforms extrinsic rewards in sustaining exploratory behavior and risk tolerance, while personality traits such as independence, openness, and ambiguity tolerance enable unconventional problem reframing (Sternberg, 2006). Second, cognitive processes require a dynamic balance: divergent thinking (e.g., fluency, flexibility, originality) must be strategically coupled with convergent evaluation (Sternberg, 2006), supported by meta-cognitive strategies like conditional knowledge and self-questioning to optimize ideation (Allwood & Selart, 2010). Third, environmental factors, including autonomy-supportive climates (Amabile, 2018) and inclusive educational strategies, could foster creativity. Last, cross-culturally, Kharkhurin's (2024) research demonstrates that multilingualism and multicultural exposure enhance cognitive flexibility and conceptual blending, while cross-cultural teams leverage individualism-collectivism synergies to overcome functional fixedness. These findings collectively suggest that CPS efficacy emerges from synergistic optimization of individual traits, cognitive and meta-cognitive procedures, and culturally responsive environments.

2.2.2. Comparing the CPS abilities of humans and GenAI

The comparative assessment of CPS capabilities between humans and GenAI remains debated. Proponents of AI's creative potential, such as Raz et al. (2024), compared human and AI question-asking behaviors and found that CPS outcomes were equally original in both groups. However, they also found that AI-generated questions were more creative and complex. Conversely, critics, including Marrone (2022) and Elik (2023), maintain that GenAI has not yet matched human-level creative proficiency despite its technical contributions. Runco (2023) further emphasizes AI's inherent limitations in replicating human traits like intrinsic motivation and authentic reasoning. Methodological considerations, however, complicate these comparisons: the selected problem types in existing studies may favor AI's computational strengths, while participant samples often lack demographic diversity, potentially skewing results. A critical nuance emerged from Jia et al.'s (2024) findings, which demonstrated that GenAI-augmented creativity exhibits skill-dependent efficacy, benefiting experts significantly but having a limited impact on non-specialists. These observations collectively suggest that GenAI's creative enhancement is dynamic, which varies with users' baseline competencies, underscoring the necessity for inclusive design strategies to broaden its applicability.

2.2.3. The perceptions of GenAI in the context of CPS

Current research on perceptions of GenAI in CPS reveals both promising findings and methodological limitations that warrant further investigation. Al-Zahrani (2024) and Marrone et al. (2022) reported that participants perceived AI could enhance CPS outcomes, but their conclusions remain context-bound. Al-Zahrani (2024) found that GenAI can support CPS. However, it is crucial to maintain human judgment and intuition, which calls for harmonious interaction between AI-generated suggestions and human creative thinking. Marrone et al. (2022) were restricted to specific student age groups, underscoring the necessity for lifespan studies encompassing diverse educational stages. Additionally, So et al. (2024) demonstrated that educators developed negative perceptions of GenAI following a professional development workshop. However, the study's short duration (4 h) and narrow geographic focus (Indonesian teachers only) constrain the generalizability of these results, highlighting the need for longitudinal and cross-cultural replications. Similarly, Qawqzeh's (2024) stratified sampling study identified conflicting perceptions on GenAI's influence on problem-solving across populations, suggesting that prior AI experience, educational attainment, and task-specific contexts may mediate attitudes. In sum, these studies emphasize that understanding GenAI's role in CPS requires addressing three critical gaps: (1) extended intervention timelines to assess perception sustainability, (2) demographic and cultural diversification of samples, and (3) task- and domain-specific evaluations to disentangle contextual effects from universal patterns.

2.2.4. The influence of GenAI on CPS

Research on the influence of GenAI on CPS has yielded inconsistent and complex results. Some studies, such as those by Al-Zahrani (2024), Hwang and Won (2021), Urban et al. (2024), and Wei et al. (2025), suggested that GenAI can enhance creativity. For example, Urban et al. (2024) compared students who used ChatGPT with those who did not on a specific task and found that ChatGPT can improve CPS performance. However, the sample size was relatively small ($N_{students} = 77$), which might have reduced the statistical power of the results. Zha et al. (2024) developed an intelligent agent for students and reported positive effects, but the long-term impact and scalability of this approach need further investigation. Hwang and Won (2021) found that participants consistently contributed more and higher-quality ideas when they perceived their teamworking partner as a bot. Hwang and Won (2021) also examined whether the perceived dominance of a partner and the pressure to generate ideas during the task mediated positive outcomes of idea generation, depending on whether the conversational style of the bot partner was robot- or human-like. Lim et al. (2023) found that AI's involvement can enhance the diversification of students' creative paths. In addition, Liu et al. (2023) explored the influence of ChatGPT on teachers' creativity. They found that teachers who used it demonstrated higher abilities. However, the study focused only on a specific group of university teachers and did not consider other educational levels or professions. Wei et al. (2025) also reflected students' concerns about overreliance on AI, associated cognitive offloading, and a lack of emotional support.

In summary, the current research on GenAI's influence on CPS is diverse and inconclusive. Therefore, this study examines the effect of GenAI on college students' CPS from a complementary perspective. While prior research has yielded valuable insights, its focus has often been on controlled settings and quantitative outcomes. To complement this body of work, our study investigates CPS within a naturalistic educational environment, namely a design competition, to capture the authentic process and mediating factors of human-AI collaboration. This approach allows us to address the how and why behind GenAI's influence, providing much-needed contextual depth to the existing, and at times inconclusive, findings from more controlled experiments.

3. Methodology

3.1. Study design

We adopted an explanatory mixed-methods design (Creswell & Clark, 2017) to comprehensively investigate the influence of GenAI on college students' CPS. The combination of a natural experiment and content analysis based on interview transcript data was chosen for several reasons. The natural experiment allows us to compare the CPS performance of students in the years 2024 (using GenAI) and 2023 (not using GenAI) under relatively similar real-world conditions, minimizing the artificiality often associated with controlled experiments. This design helps capture GenAI's actual association with CPS in a practical educational setting. The content analysis (Weber, 1990) of interview transcripts complements the quantitative data regarding the project scores by providing in-depth insights into students' usage patterns and experiences with GenAI during the CPS process, enabling a more comprehensive understanding of the phenomenon.

3.2. Study context

The study was conducted in a design competition focused on the future of education, organized annually in July by a prominent public university (BU) in China specializing in teacher education. This competition was selected because it provides a rich, natural, and complex environment where students are required to identify educational problems and develop innovative solutions, closely aligning with the essence of creative problem-solving. The three-month-long competition included participant recruitment, training, and a 48-hour on-site final stage, which served as the experimental setting. In 2024, the introduction of GenAI tools throughout the process, in contrast to their absence in 2023, created a natural experiment to examine the impact of GenAI on students' CPS performance.

3.3. Study process

The 48-hour competition in 2023 and 2024 consisted of three iterative stages (see Fig. 1). At each stage, students participated in activities such as brainstorming, design, and development. During brainstorming, students selected educational problems based on recommended themes and formulated initial ideas. Then, students designed the solution framework, and during the development phase, they further refined and completed key components, including images, videos, or mobile apps. At the end of each stage, students presented their work to a panel of experts who evaluated and graded it using a pre-defined rubric (see Appendix A). Students used the experts' assessment to improve their work iteratively in the following stages. These three rounds of grading results constituted the CPS performance scores, ensuring a comprehensive assessment of students' creative outcomes throughout the process.

A platform that offers access to multiple GenAI tools, including ChatGPT, DeepSeek, MidJourney, was provided to students. Students were encouraged but not forced to use these tools. Although this platform could record and log data, most students chose to use their preferred GenAI tools directly on their official websites rather than through the research team's platform. Therefore, log data were disregarded in the analysis, as only a few log entries (fewer than 20) were obtained and could not provide meaningful insights.

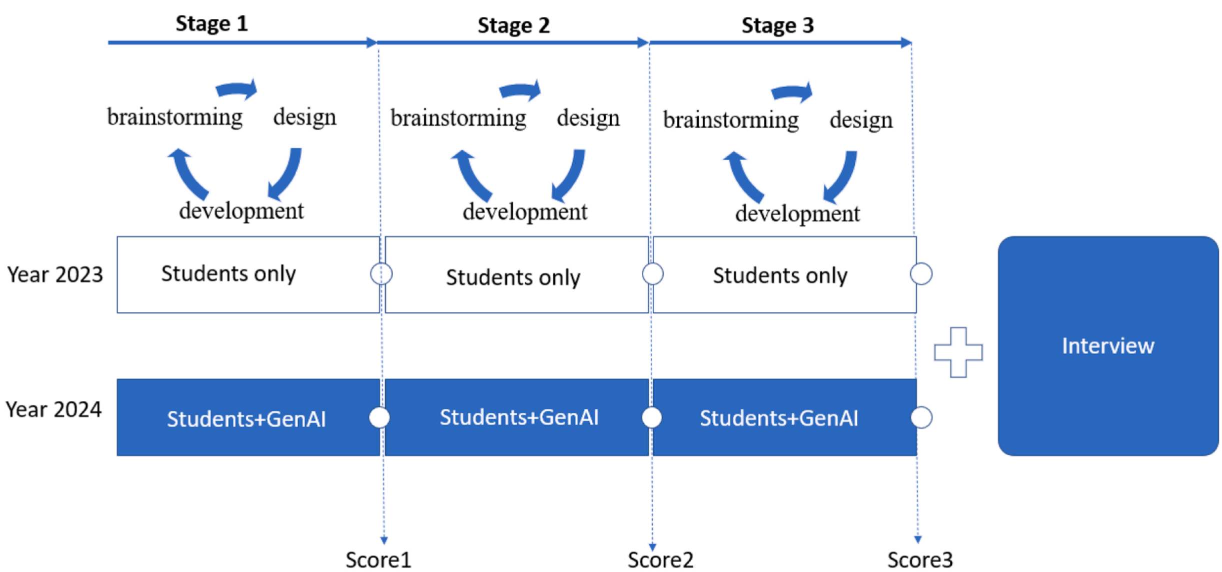

Fig. 1. Study process, consisted of three iterative design stages and post interview.

3.4. Participants

The study included multiple teams of students who had successfully passed the regional selection contests and were admitted to the competition's final stage. In 2023, 10 teams entered the final competition. Among them, two teams were eliminated from the competition, and only 8 teams completed all three stages of presentation and measurement to obtain a final score. As a result, only these 8 teams had complete data sets and were included in the final analysis. In 2024, 26 teams were admitted to the final competition, 8 of which were eliminated in the second round. As a result, 18 teams reached the final scoring stage and had complete datasets for analysis. To ensure data comparability, only teams with a final score were included in the analysis. Tables 1 and 2 provide detailed demographic information on the student participants, including team size, gender distribution, educational background, educational level, and the selected mentor participants who observed students' activity throughout the process. The sample included students from various majors and educational levels, enhancing the study's representativeness to some extent.

3.5. Data collection

Student background data were collected through an online survey during the registration process. These survey data included students' majors, prior academic achievements, and other factors.

The project scores were obtained from evaluations conducted by the same group of 4 experts in 2023 and 2024. The rater experts, with extensive experience in design, education, and technology, independently scored all student projects using the predefined rubric (Appendix A). The rubric was informed by the 4 Ps theoretical work (Rhode, 1961) and focused on the creative "Product" dimension. More specifically, the key aspects to measure product were determined based on the Creative Product Semantic Scale (CPSS, Besemer & O'Quin, 1999), and cover key elements for evaluating a product's creativity, including novelty (composed of originality and surprise), resolution (composed of logical, useful, valuable, and understandable), and elaboration and synthesis (also called style; composed of organic, well-crafted, and elegant). Based on these dimensions, the rubric was further reframed for easy understanding and implementation, which covers five major dimensions. First, Problem Awareness (20 points, associated with resolution, which highlights analyzing phenomena to discover what the problems really are. Second, Innovation Spirit (30 points, associated with novelty), which suggests that the design should be original, with original concepts and methods grounded in scientific and pedagogical theories. Third, Science and Education Integration (15 points, associated with novelty and resolution), which emphasized that the design should integrate innovative technologies with education to enhance solutions. Fourth, Application Prospect (15 points, associated with resolution), which stressed that the designed solutions should be helpful in specific teaching scenarios and be easy to understand and accept. Last, Presentation and Expression (20 points, associated with elaboration and synthesis), which stressed the visual quality of the project, and the final presentation should be complete and in a logical, clear-cut form.

Before the official evaluation, the rater experts participated in two calibration sessions using non-study samples to familiarize themselves with the scoring criteria and to align their interpretation of each dimension. To examine statistical consistency, we calculated the Intraclass Correlation Coefficient (ICC, two-way random model, absolute agreement) across raters. The ICC mean value is 0.59, indicating moderate reliability (Koo & Li, 2016). Rather than pursuing perfect numerical consistency, our rating procedure intentionally followed the Consensual Assessment Technique (CAT) developed by Amabile (1982, 1996), which is widely recognized as a valid method for assessing creativity in open-ended products such as artworks, designs, or writing. The CAT emphasizes the use of multiple independent expert judgments representing diverse disciplinary perspectives, as creativity is inherently multidimensional and interpretive. Accordingly, our panel included raters from design, education, and technology domains to capture a broader understanding of creativity. In this context, moderate interrater agreement is both expected and acceptable, as diversity of expert perspectives enhances validity (Kaufman et al., 2008; Baer & McKool, 2009). Thus, while the ICC statistics provide an index of reliability,

Table 1
Student participants' demographic information.

Items	Year 2023	Year 2024	Total
Team number	8	18	26
Team size (*Mean*)	5.5	4.4	N/A
Students	44	80	124
Female	29 (66 %)	51 (64 %)	80
Male	15 (34 %)	29 (36 %)	44
Education background	Art (9, 21 %)	Art (15, 18 %)	N/A
	Management (2, 5 %)	Management (2, 3 %)	
	Engineering (2, 5 %)	Engineering (7, 8 %)	
	ICT (2, 5 %)	ICT (3, 4 %)	
	Design (7, 16 %)	Design (11, 13 %)	
	Education (9, 20 %)	Education (21, 26 %)	
	Computer science (6, 13 %)	Computer science (10, 12 %)	
	Psychology (4, 11 %)	Psychology (5, 6 %)	
	Management Information System (3, 4 %)	Biology& Medicine (3, 4 %)	
		Finance (2, 2 %)	
		Sociology (1, 1 %)	
Educational Level	Undergraduates 11 (25 %)	Undergraduate 26 (32 %)	Undergraduate 40 (29 %)
	Graduates 33 (75 %)	Graduates 54 (68 %)	Graduates 96 (71 %)

Table 2
Mentor interviewee's background information.

Pseudonym	Sex	Academic Background	Affiliation	Year(s) of Participation
Xi Lin	Female	Computer Science and Technology	University	2024
Lan Zhou	Female	Visual Communication, Education	University	2022–2024
Zhou Wang	Male	Product Design	University	2018–2024
Xing Chen	Male	Visual Communication	Enterprise	2020–2024
Jun Zhang	Male	Sustainable Design	Enterprise	2021–2024
Ran Li	Male	Education	Freelancer	2024

the overall validity of our scoring rests on expert consensus achieved through disciplinary diversity, consistent with the CAT framework. The final score for each team output was the average of all raters' independent scores.

The semi-structured interview data were collected from six highly experienced mentors with expertise in design, education, and technology. The interview protocol was presented in Appendix B. All these mentors observed the 48-hour final stage in 2024, and most of them observed the situation in 2023. The mentors provided detailed insights into students' use of GenAI throughout the competition.

3.6. Data analysis methods

Firstly, descriptive statistical analysis was employed to summarize the essential characteristics of the CPS performance scores for 2023 and 2024, including measures of central tendency and dispersion. Two mid-test (score 2) scores were missing in the 2023 dataset. These missing values were identified as missing at random (MAR), as the absence of data was due to incidental reasons (e.g., student absence) and was not systematically related to their performance or group assignment. This study employed a Linear Mixed Model (LMM) for a 2 (Year) $\times$ 3 (Time) mixed design with a covariate, which estimates parameters based on all available observations using restricted maximum likelihood (REML).

Accordingly, the analysis was conducted without the need for data imputation or case deletion. The assumptions of linearity, independence of residuals, homoscedasticity, and approximate normality were examined through visual inspection of residual and Q–Q plots and the Shapiro–Wilk test. While residuals showed slight deviations from normality, all assumptions were considered adequately met given the robustness of the LMM and the sample size ($N = 82$). The LMM was specified to compare students' final CPS performance (score 3) between the 2023 and 2024 cohorts while controlling for pretest CPS scores to adjust for baseline ability differences. In addition, the model examined changes in CPS scores over time (from score 1 to score 2 to score 3) within each cohort, thereby capturing both overall cohort differences and within-year developmental trajectories. The details of assumption checking are reported in the results section.

We adopted a content analysis approach (Weber, 1990) to analyze the interview transcripts. Two independent researchers conducted the coding process. They read the data thoroughly and developed an initial schema based on the research questions and the data content. After coding a subset of the data, the researchers compared and discussed their codes to ensure consistency and reliability. Discrepancies were resolved through further discussion and reference to the original data. Based on the refined codebook, the remaining data were coded by both coders, and themes were identified by grouping related codes. Through multiple rounds of coder discussion and revision, all inconsistent coding items were addressed, resulting in 100 % intercoder agreement. The identified themes were further analyzed and interpreted to explain the quantitative results and provide a more comprehensive understanding of students' experiences with GenAI in the CPS process. Table 3 shows the final coding scheme.

3.7. Ethical considerations

We followed the World Medical Association's Code of Ethics (Declaration of Helsinki). The study materials and procedures received approval from the ethics committee of B University (IRB ID, blind for review), China. Informed consent was obtained from all participants. All participants were fully informed about the study's goal and were made aware of their right to withdraw at any time without consequences. To ensure participants' privacy, all data collected in this study were de-identified and treated with strict confidentiality.

4. Results

4.1. How is GenAI associated with CPS's quality and trajectory across different stages?

Descriptive statistics were initially conducted to provide an overview of the CPS performance scores in 2023 and 2024. The results of descriptive statistics revealed contrasting trends between the 2023 (non-GenAI) and 2024 (GenAI) cohorts. In 2023, pretest CPS scores ($M_{pre2023} = 86.32$, $SD_{pre2023} = 2.46$) exceeded posttest scores ($M_{post2023} = 80.74$, $SD_{post2023} = 3.35$), while in 2024, GenAI-assisted students showed reversed improvement ($M_{pre2024} = 80.26$, $SD_{pre2024} = 2.50$; $M_{post2024} = 87.69$, $SD_{post2024} = 2.15$). Comparing across years, the initial CPS pretest scores in 2023 were higher than those in 2024, while the posttest scores showed the opposite trend.

Table 3
Coding schema.

Level 1 Themes	Level 2 Grouped Topics	Level 3 Specific Codes
I. Positive Influence of GenAI on CPS	1.1 Information & Problem Identification	• AI-assisted background and market research • Resource acquisition • Clarifying problem scope • Distinguishing real vs. pseudo problems • Surfacing underlying needs
	1.2 Creative Ideation & Solution Formulation	• Generating cross-disciplinary examples • Providing idea inspiration • AI as a brainstorming partner • Promoting divergent thinking • Building structured outlines or frameworks • Producing multi-perspective solutions
	1.3 Solution Evaluation & Optimization	• AI as expert evaluator • High-efficiency feedback • Technical refinement • AI-simulated stakeholder feedback • Enhancing innovation and usability • Iterative prompt optimization
	1.4 Efficiency & Learning Benefits	• Improving completion speed • Improving solution quality • Supporting cross-disciplinary capability • Providing entry-level competence • Expanding cognitive awareness • Enhancing visualization and presentation quality
II. Negative Influence of GenAI on CPS	2.1 Quality & Originality Concerns	• Homogenized ideas and VR/metaverse clichés • Low creative variance • Over-generalized solutions • Loss of uniqueness and aesthetic depth
	2.2 Bias & Factual Issues	• Regional or cultural bias • Data-source misrepresentation • Inaccurate references • Policy citation errors • Superficial detail generation
	2.3 Misuse & Over-reliance	• Skipping human ideation ("zero-to-one" missing) • Blind trust in AI • Reduced critical thinking • Failure to leverage human strengths • Cognitive laziness or passive dependence
	2.4 Underuse & Skill Deficits	• Lack of usage awareness • Treating AI as a simple search • Insufficient depth of use • Lack of practical guidance or training • Limited institutional resource support
III. Individual-Level Factors (Mediators)	3.1 Cognitive & Technical Abilities	• Thinking and expression abilities • Critical thinking and logic • Creativity and problem-solving competence • Self-regulation and reflection
	3.2 AI Literacy & Experience	• Prior GenAI usage experience • Understanding tool functions and limits • Prompting skills and depth of interaction • Expectations and motivation toward AI
	3.3 Personal Background	• Academic major differences • Prior disciplinary knowledge • Cultural/technological exposure • Family or regional access to digital resources
IV. Team-Level & Environmental Factors	4.1 Team Dynamics & Collaboration	• Collaborative atmosphere • Peer modeling ("early adopters" leading others) • Role division • Communication and coordination • Engagement balance among members
	4.2 Guidance & External Support	• Mentor facilitation • Feedback from advisors • Availability of technical or resource support • Institutional training and AI integration
	4.3 Group Composition & Diversity	• Interdisciplinary collaboration • Cultural and regional diversity • Different AI experiences within teams • Impact of collective usage climate ("usage atmosphere")

For assumption checking, as shown in Appendix C, the residual plots indicated that residuals were randomly and evenly distributed around zero, suggesting that the assumptions of linearity, homoscedasticity, and independence were satisfied. The Q–Q plot showed that the residuals were approximately normally distributed, though with slight deviations at the tails. Both the Shapiro–Wilk ($W = 0.868$, $p < 0.001$) and Kolmogorov–Smirnov ($D = 0.198$, $p < 0.001$) tests indicated a deviation from normality. However, given the moderate sample size ($N = 82$) and the robustness of the LMM to violations of normality, this deviation was not considered problematic.

Model Overview. An LMM was conducted to examine changes in students' CPS performance across three time points (pretest, mid-test, and post-test) in two cohorts (2023 vs. 2024), while controlling for pretest CPS as a covariate. The model included fixed effects of year, time, pretest, year × time, and year × pretest, with a random intercept for subjects (subject id). The repeated factor was modeled with an unstructured covariance structure to allow for freely estimated correlations among time points. Model fit statistics indicated a satisfactory model fit (−2 Log Likelihood = 98.23, AIC = 112.23, BIC = 128.36), see Table 4. The linear mixed-effects model accounted for a substantial proportion of variance in CPS scores. The marginal R^2 indicated that fixed effects explained 69.7 % of the variance, while the conditional R^2 suggested a comparable level of variance explained when random effects were included ($R^2c = 0.708$). These values should be interpreted in light of the study's context and model specification. In addition, to address our research focus on the unique contribution of individual predictors, we also reported partial eta squared (ηp^2) for fixed effects based on Type III tests. These effect sizes provide a more comprehensive and transparent quantification of the model's explanatory power and the relative contribution of key fixed effects.

The results revealed significant main effects of Year, $F(1, 30.78) = 27.522$, $p < 0.001$, $\eta p^2 = 0.472$, and Time, $F(2, 24.73) = 5.075$, $p = .014$, $\eta p^2 = 0.291$, as well as a highly significant Year × Time interaction, $F(2, 24.73) = 61.203$, $p < 0.001$, $\eta p^2 = 0.832$. As shown in Table 5, significant fixed effects were observed for Year (Estimate = −13.25, SE = 1.38, $t(91.14) = -9.61$, $p < 0.001$) and time ($time_1$: Estimate = −7.43, SE = 0.83, $t(90.87) = -9.00$, $p < 0.001$; $time_2$: Estimate = −5.99, SE = 0.83, $t(19.47) = -7.24$, $p < 0.001$), indicating that CPS scores significantly varied across both year and time.

Notably, the Year × Time interaction was also significant ($time_1 \times year_1$: Estimate = 13.00, SE = 1.38, $t(90.87) = 9.42$, $p < .001$; $time_2 \times year_1$: Estimate = 14.51, SE = 1.46, $t(21.20) = 9.94$, $p < 0.001$), suggesting that the trajectories of CPS development differed between the 2023 and 2024 cohorts.

The covariate pretest was highly significant (Estimate = 0.997, SE = 0.000, $t(47.51) = 2419.80$, $p < 0.001$), confirming that baseline CPS performance was a strong predictor of later scores. Additionally, the Year × Pretest interaction was significant (Estimate = 0.003, SE = 0.001, $t(47.51) = 4.25$, $p < 0.001$), indicating that the strength of the relationship between pretest CPS and subsequent performance varied slightly between the two cohorts.

Taken together, these findings suggest that while both cohorts showed improvement over time, the patterns of CPS change differed between 2023 and 2024, and the predictive influence of pretest performance was somewhat stronger in one cohort than the other.

Parameter estimates for the fixed effects are presented in Table 5. The intercept (Estimate = 7.67, SE = 0.82, $t = 9.31$, $p < 0.001$) represents the estimated CPS score for the reference group (the 2024 cohort at the posttest phase). The main effect of Year was negative and significant (Estimate = −13.25, SE = 1.38, $t(91.14) = -9.61$, $p < 0.001$), indicating that, after adjusting for pretest scores, the 2023 cohort's overall CPS performance was significantly lower than that of the 2024 cohort. This suggests that the 2024 group outperformed the 2023 group in overall creative problem-solving ability.

The main effect of Time was also significant, with both the pretest (Estimate = −7.43, $p < 0.001$) and midtest (Estimate = −5.99, $p < 0.001$) scores being significantly lower than the posttest (reference) scores. This pattern reflects consistent CPS improvement over time across participants. The Year × Time interaction effects were significant, with positive coefficients for the 2023 cohort at earlier time points (pretest: Estimate = 13.00, SE = 1.38, $t(90.87) = 9.42$, $p < 0.001$; midtest: Estimate = 14.51, SE = 1.46, $t(21.20) = 9.94$, $p < 0.001$). These findings indicate that the CPS growth trajectories differed between the two years.

Specifically, students in the 2024 cohort demonstrated greater gains from pretest to posttest than those in the 2023 cohort. The trend in non-GenAI teams' CPS scores in 2023 showed a wave followed by a decline in performance across stages (Stage 1 to 3: $M = 86.32 \rightarrow 88.58 \rightarrow 80.74$), whereas GenAI teams in 2024 maintained steadily increased scores ($M = 80.26 \rightarrow 81.70 \rightarrow 87.69$). See Fig. 2. This suggests that GenAI may have a supportive role in sustaining CPS momentum during project development.

Regarding the covariate, Pretest had a strong positive effect (Estimate = 0.997, SE = 0.000, $t(47.51) = 2419.80$, $p < 0.001$), confirming that students with higher baseline CPS tended to achieve higher overall performance. The Year × Pretest interaction was

Table 4
Type III tests of fixed effects in the linear mixed model predicting CPS scores.

Source	Numerator df	Denominator df	F	Sig.	ηp^2
Intercept	1	30.775	8.843	<0.006	
Year	1	30.775	27.522	<0.001	0.472
Time	2	24.733	5.075	0.014	0.291
Pretest	1	47.509	7938537.347	<0.001	0.999
Year*Time	2	24.733	61.203	<0.001	0.832
Year*Pretest	1	47.509	18.017	<0.001	0.275

Note. Partial eta squared (ηp^2) was calculated based on F values and degrees of freedom from the Type III tests of fixed effects. The calculation method is listed here. $\eta_p^2 = \frac{F \times df_{effect}}{F \times df_{effect} + df_{error}}$

Table 5
Estimates of fixed effects.

Parameter	Estimate	Std.Error	df	t	Sig.	95 % CI (Lower, Upper)
Intercept	7.665	0.823	91.053	9.310	< 0.001	[6.029, 9.300]
Year	−13.245	1.378	91.141	−9.612	< 0.001	[−15.982, −10.508]
Time (Pretest)	−7.425	0.825	90.866	−9.004	< 0.001	[−9.063, −5.787]
Time (Midtest)	−5.993	0.828	19.474	−7.237	< 0.001	[−7.723, −4.262]
Pretest (Covariate)	0.997	0.000	47.508	2419.797	< 0.001	[0.996, 0.998]
Year (2023) × Time (Pretest)	13.003	1.380	90.866	9.423	< 0.001	[10.262, 15.744]
Year (2023) × Time (Midtest)	14.506	1.460	21.199	9.938	< 0.001	[11.473, 17.540]
Year * Pretest	0.003	0.001	47.509	4.245	<0.001	[0.002, 0.004]

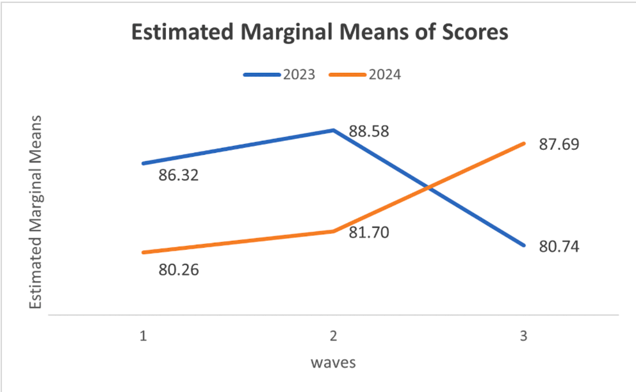

Fig. 2. Three waves of estimated marginal means of scores in the years 2023 and 2024.

small but statistically significant (Estimate = 0.003, SE = 0.001, t (47.51) = 4.25, $p < 0.001$), suggesting a slightly stronger predictive relationship between baseline and subsequent CPS in one cohort.

In summary, the linear mixed model revealed several key effects. First, a significant main effect of Year indicated that, after controlling for baseline CPS, students in 2024 achieved higher overall CPS scores than those in 2023. Second, a significant main effect of Time reflected meaningful improvement in CPS across the three assessment points. Third, a significant Year × Time interaction showed that the trajectories of CPS growth differed between cohorts, with the 2024 group showing a steeper increase. Fourth, Pretest emerged as a strong positive covariate, confirming that baseline ability predicted overall performance. Finally, although the Year × Pretest interaction reached significance, its small magnitude suggests that the predictive relationship between baseline and later CPS was largely consistent across years.

The mentor interviews can further explain the identified positive association between GenAI and CPS. Mentors unanimously attested to the significant role of GenAI in empowering all stages of the CPS process ($N_{reference} = 33$). GenAI's ability to efficiently integrate and screen information was highly regarded during the initial data collection and investigation stage ($N_{reference}=4$). For instance, Jun Zhang's observation that *"AI tools can conduct thorough background research on product information to help students analyze competing products"* was a common sentiment among the mentors. This GenAI function saves students time and broadens their information sources, lowering the threshold for information gathering and helping them focus on idea brainstorming ($N_{reference}=3$). During the problem analysis phase, GenAI's functions of rapid comparison, induction, and insight were frequently mentioned ($N_{reference}=1$). Ran Li's comment, *"AI can help our students better position their problems, determine whether it is a real issue or a false proposition, and even assist them in uncovering the real needs within the problem,"* highlights how GenAI enables students to delve deeper into the problem's essence. By providing a wealth of data and different perspectives, GenAI catalyzes students to think more critically about the problem at hand. In the idea generation stage, GenAI functions as a virtual companion ($N_{reference} = 5$). Zhou Wang's belief that *"by asking questions, students can get inspiration from the answers generated by GenAI and then conduct further research based on this inspiration and train of thought"* was echoed by many other mentors. This interactive brainstorming process, facilitated by GenAI, encouraged students to explore a broader range of creative ideas, breaking the limitations of traditional thinking. GenAI's convenience in constructing solution frameworks and organizing logic was emphasized during solution construction ($N_{reference} = 5$). Jun Zhang's statement, *"using AI is equivalent to helping students search for and generate a more professional and mature outline or solution framework,"* along with Lan Zhou's mention of providing *"a variety of perspectives for solutions,"* showcases how GenAI assists students in organizing their thoughts and developing more comprehensive solutions. Moreover, GenAI's role in evaluating and improving solutions was also noted ($N_{reference}=7$). Xi Lin commented, *"Artificial intelligence can play an effective role not only in the stage of understanding the problem but also in the process of creating problem solutions and finally evaluating their solutions. Moreover, for the provided solutions, we can ask the large model to expand each solution, etc.,"* indicates that GenAI provides valuable feedback and suggestions, enhancing the innovation and practical value of the solutions. The significant improvement in the visual presentation of the works in 2024, as pointed out by Zhou

Wang, further attests to the positive impact of GenAI on the overall quality of the solutions ($N_{reference}$=3). In addition, GenAI was found to have a positive effect on students' task completion ($N_{reference} = 5$). Lan Zhou's view that *"GenAI helps students break disciplinary boundaries and promotes interdisciplinary thinking"* was supported by examples of students expanding their cognitive horizons and proposing new topics after using GenAI. This indicates that GenAI catalyzes students' continuous growth and development in the CPS process.

In contrast to the identified supportive effects, interview data also revealed several critical negative impacts of GenAI on CPS ($N_{reference} = 8$). A major issue identified in the solution quality area was the homogenization of solutions ($N_{reference} = 2$). Xi Lin's observation that "*Most solutions from participants remain based on traditional VR and metaverse technologies, struggling to break free from conventional frameworks to achieve meaningful innovation*" highlighted the lack of originality and innovation when students relied too heavily on GenAI without tailoring it to specific requirements. This homogenization also extended to the visual presentation ($N_{reference} = 1$). Lan Zhou explained that when students lack foundational design knowledge and rely solely on large text-to-image models, the generated images tend to be similar. Another significant problem is the potential bias and unfairness introduced by GenAI ($N_{reference} = 1$). Lan Zhou noted that "*Large language models, shaped by inherent biases in their training data, may favor specific regions or cultures. For example, some students used AI-generated data from Indian regions to represent the Asian cultural context, leading to inaccurate or neglectful portrayals in their solutions.*" This issue affects the quality of the solutions and raises concerns about the reliability and ethics of using GenAI in educational settings. GenAI also struggles with addressing detailed or specialized problems ($N_{reference} = 1$). Lan Zhou's comment, "*For detailed or domain-specific issues, such as those in special education, the performance of GenAI is frequently inadequate*," indicates that GenAI's capabilities are limited in handling complex, specialized problems that require in-depth domain knowledge. Furthermore, the generation of inaccurate information, such as statistics, references, and citations, is a common problem ($N_{reference} = 1$). Lan Zhou's example of students encountering difficulties verifying the accuracy of AI-generated references when preparing supporting materials demonstrates the potential pitfalls of relying on GenAI for information. GenAI-generated images often lack focus and detail in the visual aesthetics domain ($N_{reference} = 2$), as Jun Zhang and Xing Chen noted. This shortcoming limits GenAI's effectiveness in creative design tasks and underscores the need for further improvement in its visual generation capabilities.

As the interview reflected, the negative effect mentioned above can be caused by the inappropriate use of GenAI in CPS. Mentors identified several issues with students' use of GenAI during the CPS process ($N_{reference} = 25$). The lack of in-depth understanding and application of GenAI was a top-one common problem ($N_{reference} = 8$). Xi Lin's comment that students view GenAI as a traditional search engine reflects their superficial understanding of the tool's capabilities. Xing Chen's observation that students rely on mentors for advice because they lack AI experience suggests a lack of independent problem-solving ability. The other prominent problem is dependence on AI tools ($N_{reference} = 5$). Zhou Wang's observation that *"A significant number of students overly rely on the tool, expecting*

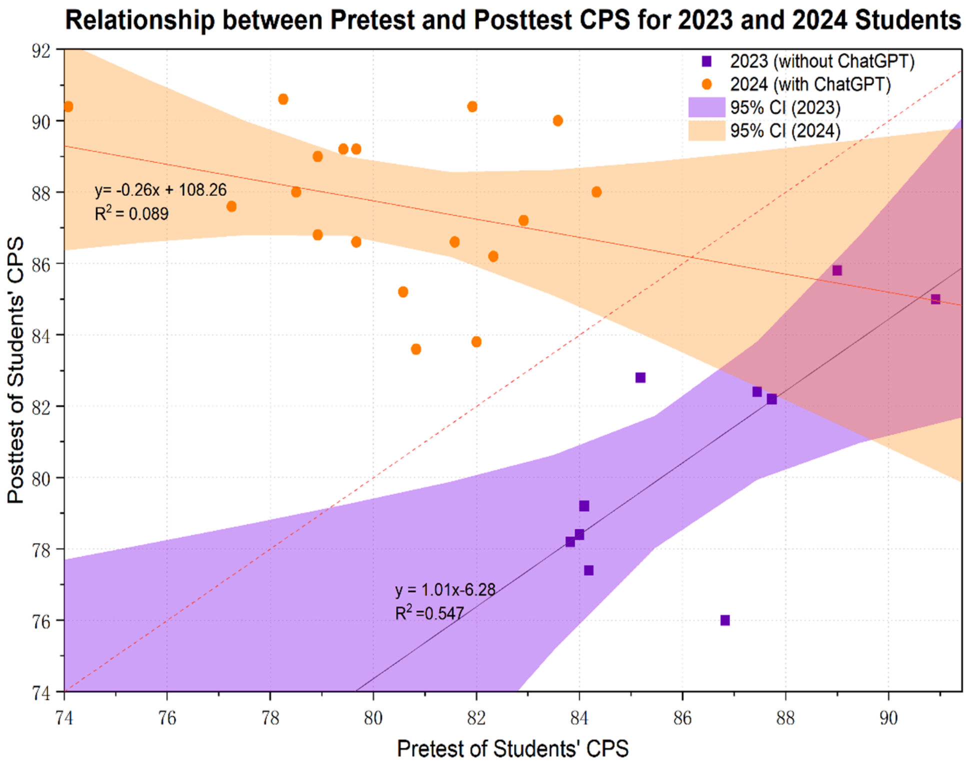

Fig. 3. Scatter plot for pre-post scores in two years.
Note: Purple: year 2023, Yellow: year 2024; point above 45-degree dotted line indicates increased pre-post; point below 45-degree dotted line indicates decreased pre-post.

GenAI to directly provide a perfect answer or solution, without thinking by themselves" and Lan Zhou's comment about students' lack of validation of AI-generated content indicate that students often fail to engage in independent critical thinking and validation. This prevents them from fully utilizing the potential of GenAI and undermines their own critical thinking and CPS abilities. Regarding insufficient use, several mentors noted a lack of awareness of usage ($N_{reference} = 3$). Zhou Wang explained that students need guidance in using GenAI for tasks such as literature research and reading. Some students abandoned the tool due to a lack of skills or the belief that they can perform better without it. Teamwork problems were also noted ($N_{reference} = 3$), with Xi Lin observing that students do not collaborate effectively, resulting in wasted time and inefficient teamwork.

Another aspect of improper use is the failure to leverage students' professional advantages ($N_{reference} = 2$). Xi Lin's example of psychology students not showcasing their expertise when using GenAI shows that students often overlook the combination of their professional knowledge and GenAI's capabilities, leading to suboptimal use of the tool. Students' weak problem awareness among students was identified ($N_{reference} = 2$). Xing Chen's example of students focusing on finding solutions to specific problems rather than designing solutions tailored to rural needs demonstrates that students often fail to grasp the essence of the problem. Xi Lin's explanation that students' incomplete understanding of the topic and technology leads to the formulation of inappropriate problems further emphasizes this issue. In addition, insufficient external support exacerbates these issues ($N_{reference} = 1$). As Xi Lin pointed out, most schools do not offer formal courses on the use of GenAI, forcing students to rely on informal channels such as social media and videos for information. However, these sources often lack systematic and scientific content, limiting students' understanding and application of GenAI. Mentors also pointed out that students' limited thinking skills, influenced by exam-oriented education, restrict their ability to explore diverse solutions and inhibit their CPS skills ($N_{reference} = 1$).

4.2. Can GenAI mitigate creativity disparities among students with varying baseline abilities?

The significant interaction between year and pretest, F (47.51) = 4.25, $p < 0.001$. in Table 4, suggests that GenAI's effectiveness in improving students' CPS performance was contingent upon their initial CPS levels. To further clarify this finding, the estimated marginal means plot of scores was calculated (Fig. 3).

As illustrated in the scatterplot (Fig. 3), in 2024, students with initially lower CPS levels demonstrated greater improvements when using GenAI than their 2024 peers with higher initial CPS levels. In contrast, in 2023, the posttest score is positively related to the pretest score. Students with initially higher CPS levels demonstrated greater improvement when they did not use GenAI than their 2023 peers with lower initial CPS levels. The figure's visual representation provides early evidence from a global competition held in China, indicating that GenAI played a crucial role in shaping students' performance across varying baseline levels during the CPS process.

In addition, Fig. 3 shows the regression lines for 2023 and 2024, which illustrate different trends in the relationship between pretest and posttest scores. The steeper slope in 2023 indicates a stronger relationship between pretest and posttest scores when GenAI was not used, especially in the later stages of the competition. This graphical evidence supports our conclusion that GenAI use plays a positive role in maintaining students' performance across the CPS stages.

4.3. What individual and team-level factors mediate the effectiveness of human-AI collaboration?

The in-depth content analysis of the mentor participants' responses revealed several factors that mediate the effect of GenAI on students' CPS performance. The first is the student's general abilities ($N_{reference} = 2$), particularly thinking and expression skills. As Jun Zhang stated, "*For a student who has strong personal abilities, AI will become a stronger support. If their personal abilities are poor, for example, some art students may not have strong logical thinking skills, or their thinking may be too jumpy. These tools cannot help him generate better solutions."* This is because students with clear plans and steps for action can better use GenAI to collect and analyze data or background information, then screen, optimize, and present the content it suggests. For example, when integrating multiple creative ideas generated by GenAI, students with strong meta-cognitive abilities can quickly identify and combine relevant elements. In contrast, those with poor thinking abilities may struggle to make sense of the information.

Students' prior experience and AI literacy are other important factors ($N_{reference} = 7$). Those with prior experience in using GenAI ($N_{reference} = 2$), as Jun Zhang noted, "*with experience in GenAI usage, students have a more intuitive understanding of what GenAI can do and its limitation; they can use it better*." They can operate the tools more skillfully and select appropriate functions based on different needs. In contrast, students without GenAI exposure may struggle to use the tool effectively. Moreover, students' understanding of GenAI also affects their ability to use it ($N_{reference} = 4$). Xing Chen's observation that "*Many students are not clear about the uses and applications of GenAI, so they don't know how to ask questions*" indicates that a lack of knowledge about GenAI leads to ineffective use. Additionally, unrealistic expectations of GenAI ($N_{reference} = 1$), as noted by Ran Li, can lead students to abandon the tool when results do not meet expectations quickly.

Group dynamics is a significant factor ($N_{reference} = 2$). A positive team environment can promote communication and learning among students, accelerating the mastery and application of GenAI. Jun Zhang's comment, "*There should be a group of people who get ahead first. They were already using AI tools themselves, so they can drive others in the team to use GenAI, whether it is for research or quickly learning some knowledge they don't master,*" highlights the importance of a collaborative team atmosphere in facilitating the effective use of GenAI.

Student demographic background also influences the application of GenAI ($N_{reference} = 3$). Students from different majors use GenAI differently and achieve varying results ($N_{reference} = 2$). For example, students major in design may have an advantage in using GenAI for creative design, while students from other majors may need more time to master the relevant skills. Family background also plays a role ($N_{reference} = 1$). As Ran Li pointed out, students from diverse regions and family backgrounds have different access to new technologies and learning resources, which affects their use of GenAI.

5. Discussion

The findings of this study contribute context-specific evidence regarding the relationship between GenAI and college students' CPS performance and processes in an authentic setting. By comparing the CPS outcomes of students who used GenAI with those who did not, this research offers preliminary insights into how GenAI may be associated with differences in CPS performance, while also revealing several challenges and boundary conditions that warrant careful consideration.

5.1. The dual role of GenAI in the CPS process: catalyst and crutch

The results of the Linear Mixed Model indicate a significant associative relationship between GenAI use and the CPS process. Students who used GenAI demonstrated higher CPS performance, a pattern consistent with previous findings (Kim et al., 2024; Rafner et al., 2023; Urban et al., 2024). Rafner et al. (2023) found that AI-assisted visualization tools could enhance problem formuation and idea expansion during CPS. Extending this line of work, the present study suggests that the use of GenAI may be differentially associated with various stages of the CPS process in a real-world educational competition context. One possible interpretation is that GenAI's capacity to generate information, perspectives, and multimodal representations rapidly may have supported idea exploration and refinement, particularly in later stages of the CPS process. However, these interpretations remain speculative, as the underlying cognitive mechanisms were not directly examined. Accordingly, the observed associations should be interpreted with caution.

The interaction between GenAI and students' initial CPS levels is a particularly interesting finding. The results suggest that the use of GenAI was more strongly associated with improvements in CPS performance among students with lower pretest scores. Within the context of this study, GenAI appeared to function as a potential supportive resource for less-advanced students, narrowing but not eliminating performance differences relative to peers with higher baseline CPS abilities. Similar patterns have also been reported in prior work examining intelligent agent-based systems (Zha et al., 2024).

Rather than demonstrating a general equalizing effect, these findings provide preliminary, context-bound evidence that GenAI may help reduce performance disparities under specific instructional conditions. As such, claims regarding democratization or mitigation of educational inequities should be treated as tentative and contingent on task design, learner characteristics, and instructional framing.

5.2. Mediating factors: toward sustainable human-AI co-creativity

The findings further indicate that the effectiveness of GenAI in CPS is mediated by several learner- and context-related factors, including students' general cognitive abilities, prior experience, AI literacy, and team dynamics. Students with stronger logical reasoning and problem-solving skills appeared better able to critically evaluate and integrate AI-generated ideas, whereas students with weaker foundational skills were more likely to engage in superficial or inefficient use of the tool. This pattern aligns with existing CPS theories emphasizing the role of individual cognitive and metacognitive resources (Sternberg, 2006).

Consistent with Lim et al. (2023), the present study suggests that AI literacy may play an important role in shaping students' ability to use GenAI effectively. Interview data highlighted a gap between the growing demand for GenAI-related competencies and the limited availability of formal AI literacy instruction in current educational contexts. This structural limitation suggests that issues such as over-reliance or shallow use of GenAI should not be interpreted solely as individual shortcomings, but rather as reflections of broader systemic constraints.

Team dynamics further emerged as a relevant contextual factor. Our results align with Ouyang et al. (2023) by demonstrating that teams characterized by positive interdependence and role clarity tended to make more effective use of GenAI, whereas teams experiencing process conflict or low psychological safety struggled to integrate AI outputs productively. This finding resonates with prior work on human–AI collaboration and underscores the importance of social and organizational conditions in shaping AI-supported learning outcomes.

5.3. Challenges and limitations of GenAI in CPS

Despite the identified positive associations, the study also identifies several challenges associated with the use of GenAI in CPS. One recurring concern is students' tendency to overreliance on GenAI-generated content, which may erode independent critical thinking and problem-solving skills. As several mentors pointed out in the interviews, students who rely too heavily on GenAI without engaging in independent thinking often produce homogenized solutions that lack originality and depth, echoing Lim et al.'s (2023) warnings about AI-driven conformity and bias as reflected by Salvagno et al. (2023). This is also a concern raised by researchers such as Zhai et al. (2024) and Dwivedi et al. (2023), who argue that over-reliance on AI can stifle students' cognitive development and creativity. These findings suggest that instructional designs should emphasize critical evaluation, reflective use, and the integration of AI-generated ideas with students' own reasoning.

Additional challenges relate to the accuracy, relevance, and domain specificity of AI-generated content. Students often struggle to verify the accuracy of AI-generated data and references, increasing the risk of incorporating inaccurate or misleading information (Yan et al., 2024). Moreover, GenAI demonstrated limitations in addressing highly specialized or context-sensitive problems (Lanzagorta-Ortega et al., 2022), reinforcing the need for human expertise in domains requiring deep disciplinary knowledge (Wilson

& Daugherty, 2018).

6. Implications, limitations, and conclusion

The findings of this study have several tentative implications for educational practice. Firstly, the results suggest that incorporating GenAI into CPS-related activities may provide opportunities for students to engage with AI-supported idea generation and problem exploration. However, such integration should be accompanied by explicit instructional support that emphasizes critical evaluation and refinement of AI-generated content. Students may benefit from guidance on distinguishing potentially valuable ideas from those requiring further scrutiny, as well as on integrating AI-generated suggestions into their own thinking while maintaining independent judgment. In addition, the findings highlight the potential importance of balancing GenAI use with students' own problem-solving and reflective processes. Rather than functioning as a substitute for human cognition, GenAI may be more productively positioned as a supportive tool within carefully designed instructional contexts. Second, this study points to the relevance of AI literacy as a contextual factor shaping how students engage with GenAI in CPS tasks. Providing structured learning opportunities, such as workshops, training sessions, or courses, may help students better understand both the affordances and limitations of GenAI, thereby enabling more informed and reflective use. Finally, the results suggest that collaborative processes and team dynamics may play a role in shaping the effectiveness of GenAI-supported CPS activities. Instructional approaches that foster communication, coordination, and shared reflection within teams may facilitate more productive integration of AI tools. These implications should be interpreted cautiously, given the study's context-specific nature, and warrant further investigation across diverse educational settings.

This research acknowledges several key constraints. First, the interrater reliability of the CPS scores was moderate (ICC $= 0.59$), which may limit the precision of creativity measurement. Although such variability is common when evaluating open-ended creative products using expert judgments from diverse disciplinary backgrounds, it nonetheless introduces a degree of subjectivity into the assessment process. Future studies may consider combining expert ratings with complementary measures or increasing the number of raters to enhance measurement precision.

Second, the study adopted a natural experiment design by comparing two cohorts from consecutive years. While this design enhances ecological validity, the two cohorts may differ in ways beyond GenAI usage, such as prior experience, cohort composition, or contextual factors. As a result, the observed differences should be interpreted as associative rather than strictly causal. These limitations suggest that the findings should be understood as context-specific evidence rather than definitive causal claims. Subsequent controlled trials could be adopted to better isolate GenAI effects. In addition, the exclusive focus on short-term outcomes leaves unanswered questions about longitudinal skill development, warranting multi-wave studies to map AI's enduring cognitive impacts.

In conclusion, this study provides context-specific evidence that GenAI is positively associated with differences in college students' CPS performance in a realistic setting, particularly in relation to solution quality, elaboration, and originality. The findings further suggest that effective engagement with GenAI may depend on students' existing critical thinking and problem-solving skills, as well as their level of AI literacy. Accordingly, students may benefit from developing these competencies alongside technical familiarity with AI tools, thereby supporting collaborative skills and team processes. Within carefully designed instructional contexts, such an approach may help students engage more productively with GenAI while reducing risks of overreliance or superficial engagement with AI-generated content. Finally, future research is needed to address this study's limitations, including its short-term focus and potential cohort differences, and to examine further the longer-term cognitive and learning-related consequences of GenAI use while controlling for additional covariates. Such efforts would contribute to a more nuanced and cumulative understanding of how GenAI functions within educational contexts.

Declaration of generative AI in writing

During the preparation of this work the author(s) used ChatGPT in order to improve the language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Huanhuan Wang: Writing – original draft, Funding acquisition, Conceptualization. **Ting Da:** Writing – review & editing, Investigation. **Yuanrong Wang:** Formal analysis, Data curation. **Fenghao Luo:** Formal analysis, Data curation. **Yifan Zhang:** Writing – review & editing, Investigation. **Ahmed Tlili:** Resources, Methodology. **Dong Yang:** Resources, Methodology. **Yarong Wang:** Resources, Investigation. **Jiaxin Xu:** Resources, Investigation. **Xixian Zhu:** Resources, Methodology. **Man Wan:** Resources, Investigation. **Ronghuai Huang:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

We have nothing to declare.

Acknowledgement

This work was supported by the Fundamental Research Funds for the Central Universities in China [grant number 1233100004] and the Ministry of Education, China [grant numbers 24YJC880129 and 23JZDW12].

Appendix A. Project evaluation rubric

Dimensions	Score Range				
	Total	Excellent	Average	Pass	Fail
Problem-oriented	20	20–18	17–16	15–12	12–0
• Whether the problem identified is a real, clear, and specific problem in the field of education. • Whether the problem is solved based on users' real needs to implement people-oriented design.					
Novelty	30	30–27	26–24	23–18	17–0
• Whether the design can solve a new educational problem or provide a new idea. • Whether the design represents a novel and substantial breakthrough in ideation, imagination, or method. • Whether the design is original and not an existing commercialized product, or a prototype of a professional company or research unit.					
Tech and Sci Integration	15	15–14	13–12	11–9	8–0
• Whether the submitted design is sound, technically feasible, and complete. • Whether the new technologies used in the design are capable of delivering the intended functionalities.					
Application Prospect	15	15–14	13–12	11–9	8–0
• Whether the solutions are scientific and pedagogical. • Whether the solutions apply scientific research to real teaching scenarios. • Are the solutions interdisciplinary and integrated with industry and education? • Whether the solutions apply design thinking?					
Presentation and Expression	20	20–18	17–16	15–12	12–0
• Are the project materials complete and able to clearly express the problem, clarify the solution, and highlight the creativity? • Is the presentation of the project material harmonious and beautiful, with appropriate use of imagery and a high level of artistic expression? • Is the presentation of the project presented in an appropriate manner, with fluency and clarity of expression?					

Appendix B. Interview questions

1. Could you please introduce your professional background?
2. How much do you know about GenAI?
3. What is your view on using GenAI in creative problem-solving activities like this competition?
4. What is your overall impression of the works in this (2024) competition in terms of the quality of the solutions, innovativeness, and visual elegance?
5. What roles do you think GenAI played in this student activity?
6. How do you evaluate the students' application of GenAI in this competition? What aspects do you think were well done? What aspects do you think the application of GenAI was not good in?
7. What suggestions do you have for students using GenAI in creative problem-solving activities in the future?
8. Based on two years of competition observation, what do you think is the difference between using GenAI and not using GenAI?

Appendix C. Results of assumption checking

1. Checking normality of residuals

The normal Q–Q plot (Fig. 4) showed that most residuals aligned approximately with the diagonal line, although deviations were observed at the tails. The descriptive statistics (skewness = 0.795; kurtosis = 3.245) indicated a slightly right-skewed, leptokurtic distribution. However, given the sample size ($N = 82$) and the robustness of LMMs to minor violations of normality, this deviation is not considered problematic.

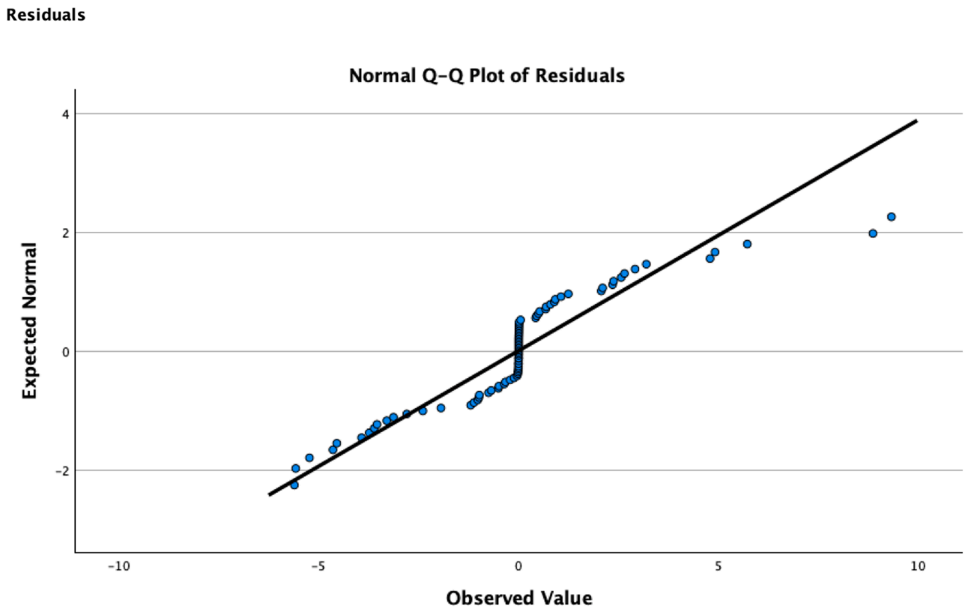

Fig. 4. Q-Q plot.

2. Linearity assumption checking

The scatterplot of residuals versus predicted values (Fig. 5) showed no apparent curvature or systematic pattern, suggesting that the relationship between predictors and the dependent variable was adequately linear.

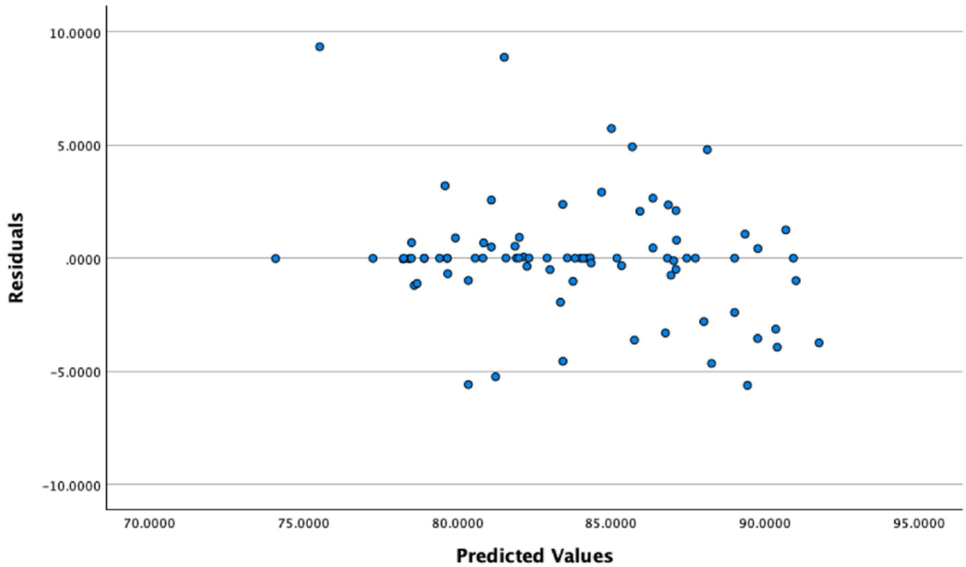

Fig. 5. Linearity assumption checking.

3. Independence of residuals

The residual plots did not indicate serial dependence or clustering. Given that the LMM accounts for repeated measures and random effects, the assumption of independence of residuals was reasonably satisfied.

4. Homoscedasticity (equal variances)

Inspection of the residuals versus predicted values plot (Figs. 6 and 7) indicated that the residuals were randomly and evenly dispersed around zero across the range of predicted scores. There was no visible funnel shape or trend, supporting the assumption of homoscedasticity.

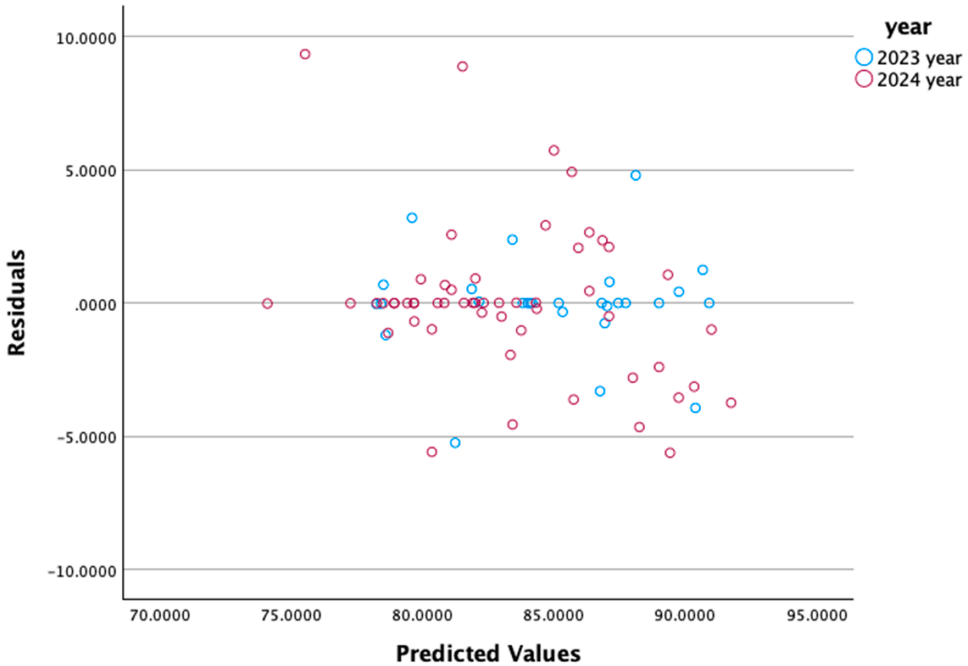

Fig. 6. Homoscedasticity assumption checking (Year).

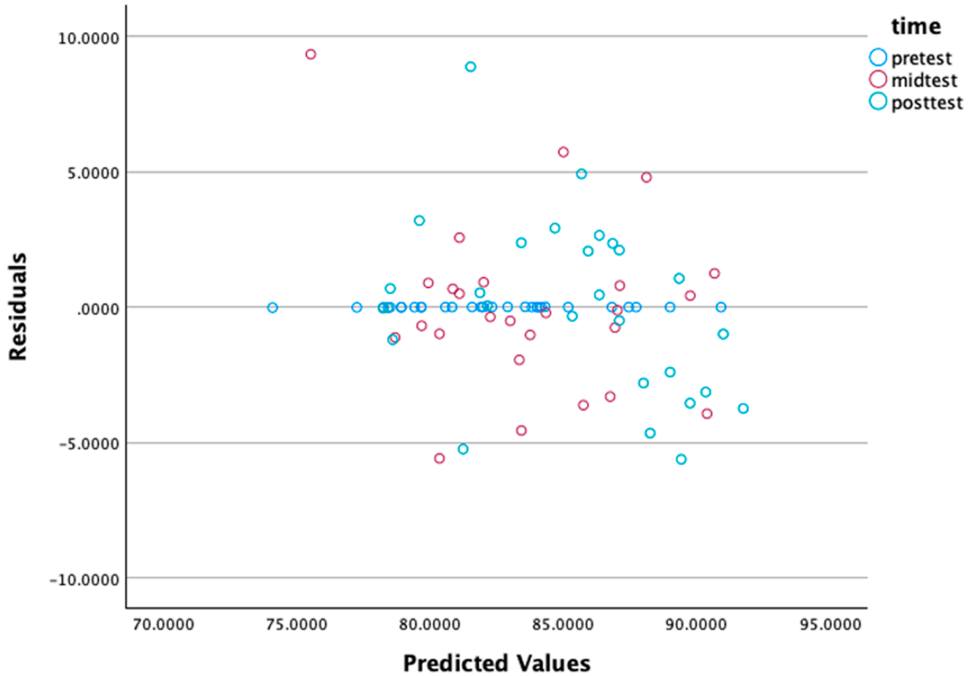

Fig. 7. Homoscedasticity assumption checking (Time).

Data availability

Due to the sensitive nature of the questions in this study, participants were assured that their raw data would remain confidential and would not be shared.

References

Al-Zahrani, A. M. (2024). Balancing act: Exploring the interplay between human judgment and artificial intelligence in problem-solving, creativity, and decision-making. *IgMin Research, 2*(3), 145–158.

Allwood, C. M., & Selart, M. (2010). *Decision making: Social and creative dimensions*. Dordrecht: Springer. https://doi.org/10.1007/978-94-015-9827-93

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43*(5), 997–1013.

Amabile, T. M. (1996). *Creativity in context*. Boulder, CO: Westview Press.

Amabile, T. M. (2018). *Creativity in context: Update to the social psychology of creativity*. Routledge.

Baer, J., & McKool, S. S. (2009). Research on and with the consensual assessment technique of creativity. *Creativity and reason in cognitive development* (pp. 211–228). Cambridge University Press.

Besemer, S. P., & O'quin, K. (1999). Confirming the three-factor creative product analysis matrix model in an American sample. *Creativity Research Journal, 12*(4), 287–296.

Couger, J. D. (1993). Essential element for viability of IS education: Improving our creativity. *Journal of Information Systems Education, 5*(4), 9.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage Publications.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion paper: "so what if ChatGPT wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Elik, E. (2023). Computational creativity: An investigation in stimulating human creativity with artificial intelligence in divergent thinking tasks.

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

Fiialka, S., Kornieva, Z., & Honcharuk, T. (2024). The use of ChatGPT in creative writing assistance. *XLinguae: European Scientific Language Journal, 17*(1).

Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does generative artificial intelligence impact student creativity? *Journal of Creativity, 34*(1), Article 100072. https://doi.org/10.1016/j.yjoc.2023.100072

Hwang, A. H. C., & Won, A. S. (2021). IdeaBot: Investigating social facilitation in human-machine team creativity. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–16).

Inclusive Innovation. (2017). Creative problem solving: A quick, down-and-dirty handbook. In *Proceedings of the inclusive innovation connector meeting*.

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. *Academy of Management Journal, 67*(1), 5–32.

Kaufman, J. C., Baer, J., & Cole, J. C. (2008). Expertise, domains, and the consensual assessment technique. *The Journal of Creative Behavior, 42*(4), 223–233.

Kharkhurin, A. V. (2024). A paradigmatic shift in the relationship between bilingualism and creativity: Plurilingual creativity approach. *Bilingualism: Language and Cognition*, 1–22.

Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies*, 1–36.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.

Kovač, D., Bakator, M., Palinkaš, I., Gaborov, M., & Ćoćkalo-Hronjec, M. (2023). Creative problem solving in modern organizations. In *Proceedings of the xiii international symposium engineering management and competitiveness* (pp. 118–123).

Lanzagorta-Ortega, D., Carrillo-Pérez, D. L., & Carrillo-Esper, R. (2022). *Inteligencia artificial en medicina: Presente y futuro, 158* pp. 17–21). Gaceta Médica de México.

Lim, J., Leinonen, T., Lipponen, L., Lee, H., DeVita, J., & Murray, D. (2023). Artificial intelligence as relational artifacts in creative learning. *Digital Creativity (Exeter), 34*(3), 192–210. https://doi.org/10.1080/14626268.2023.2236595

Liu, Z., Vobolevich, A., & Oparin, A. (2023). The influence of AI ChatGPT on improving teachers' creative thinking. *International Journal of Learning, Teaching and Educational Research, 22*(12), 124–139.

Marrone, R., Taddeo, V., & Hill, G. (2022). Creativity and artificial intelligence—A student perspective. *Journal of Intelligence, 10*(3), 65.

Mumford, M. D., Mobley, M. I., Reiter-Palmon, R., Uhlman, C. E., & Doares, L. M. (1991). Process analytic models of creative capacities. *Creativity Research Journal, 4*(2), 91–122.

Ouyang, F., Xu, W., & Cukurova, M. (2023). An artificial intelligence-driven learning analytics method to examine the collaborative problem-solving process from the complex adaptive systems perspective. *International Journal of Computer-Supported Collaborative Learning, 18*(1), 39–66.

Parnes, S. J. (1985). Creative problem solving. *Developing Minds: A Resource Book for Teaching Thinking*, 230–232.

Qawqzeh, Y. (2024). Exploring the influence of student interaction with ChatGPT on critical thinking, problem solving, and creativity. *International Journal of Information and Education Technology, 14*(4).

Rafner, J., Zana, B., Dalsgaard, P., Biskjaer, M. M., & Sherson, J. (2023). Picture this: AI-assisted image generation as a resource for problem construction in creative problem-solving. In *Proceedings of the 15th conference on creativity and cognition* (pp. 262–268).

Raz, T., Reiter-Palmon, R., & Kenett, Y. N. (2024). Open and closed-ended problem-solving in humans and AI: The influence of question asking complexity. *Thinking Skills and Creativity, 53*, Article 101598.

Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan, 42*(7), 305–310.

Runco, M. A. (2023). AI can only produce artificial creativity. *Journal of Creativity, 33*(3), Article 100063.

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing? *Critical Care, 27*(1), 1–5. https://doi.org/10.1186/s13054-023-04380-2

So, H. J., Lee, C., Kim, S., & Drajati, N. A. (2024). Enhancing teacher competency for creative problem solving with generative AI. *EdMedia+ innovate learning* (pp. 277–284). Association for the Advancement of Computing in Education (AACE).

Sternberg, R. J. (2006). The nature of creativity. *Creativity Research Journal*.

Tippett, C. D., & Yanez Gonzalez, R. (2022). Young children's creativity in the context of STEM learning experiences. edsK. J. Murcia, C. Campbell, M. M. Joubert, & S. Wilson (Eds.), Young children's creativity in the context of STEM learning experiences. *Children's creative inquiry in STEM. Sociocultural explorations of science education, 25*. https://doi.org/10.1007/978-3-030-94724-8_8.

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.

Weber, R. P. (1990). *Basic content analysis*. Newbury Park.

Wei, X., Wang, L., Lee, L. K., & Liu, R. (2025). The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation: An experimental study. *International Journal of Educational Technology in Higher Education, 22*(1), 23.

Wen, M. (2024). Interactive online classes in music education: The impact of online technologies on the level of creative thinking of students. *Current Psychology, 43*(15), 13619–13629.

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review, 96*(4), 114–123.

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour, 8*(10), 1839–1850.

Zha, S., Liu, Y., Zheng, C., XU, J., Yu, F., Gong, J., & XU, Y. (2024). *arXiv preprint arXiv: 2409.14228.*.

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*(1), 28–37. https://doi.org/10.1186/s40561-024-00316-7

Dr. **Huanhuan Wang** is a research assistant professor at Beijing Normal University. Her research interests focus on instructional design for personalized and online learning, achievement motivation, AI in education, etc.

Dr. **Ting Da** is a research associate at the Smart Learning Institute of Beijing Normal University. Her research focuses on the application of generative AI in education and the evaluation of LLMs in education.

Ms. **Yuanrong Wang** is a research assistant at the Smart Learning Institute of Beijing Normal University. Her research interests include early childhood education, cognitive development, and artificial intelligence in education.

Mr. **Fenghao Luo** is a research assistant at the Smart Learning Institute of Beijing Normal University. His research interests lie in future education and learning.

Dr. **Yifan Zhang** is a research associate at the Smart Learning Institute of Beijing Normal University. His research focuses on the application and evaluation of generative AI in education, learning sciences, and computational thinking.

Dr. **Ahmed Tlili** is an associate professor at Beijing Normal University. He serves as the Editor of Springer Series Future Education and Learning Spaces, and the Deputy-Editor-in-Chief of Smart Learning Environments.

Dr. **Dong Yang** is currently affiliated with the College of Education for the Future at Beijing Normal University. His research interests include student engagement in digital education, learning motivation, and STEM education, among others.

Ms. **Yarong Wang** is a research assistant at the Smart Learning Institute of Beijing Normal University. Her research interests include computer graphics and artificial intelligence in education.

Ms. **Jiaxin Xu** is a research assistant at the Smart Learning Institute of Beijing Normal University. Her research area is artificial intelligence in education.

Ms. **Xixian Zhu** is a research assistant at the Educational Research Centre, Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hung Hom, Hong Kong, China

Ms. **Man Wan** is a research assistant at the Smart Learning Institute of Beijing Normal University. Her research interests include educational technology, design and learning, and artificial intelligence in education.

Dr. **Ronghuai Huang** is a professor at the Faculty of Education of Beijing Normal University. He has mainly engaged in educational informatization, smart learning environments, technology-supported innovative teaching, and artificial intelligence in education.