Enigma in Education
https://enigma.or.id/index.php/edu

e-ISSN: 3026-5827

ORIGINAL ARTICLE
Enigma in Education | Vol. 3, No. 2 (2025) | pp. 51–57

Evaluating Generative AI as a Pedagogical Tool for Creative Problem-Solving in University Classrooms

Hesti Putri[1*], Khalil Jibran[2]

[1]*Department of Informatics Science, CMHC Research Center, Palembang, Indonesia*
[2]*Department of Chemical Science, CMHC Research Center, Palembang, Indonesia*

ARTICLE INFO

Article history
Received: August 15, 2025
Revised: September 22, 2025
Accepted: October 19, 2025
Available online: November 15, 2025
Published: November 30, 2025

Keywords
creative problem-solving; divergent thinking; generative artificial intelligence; higher education; quasi-experimental design

Corresponding author
Hesti Putri
hesti.putri@nasetjournal.com

ORCID
0009-0008-3845-4223

DOI
https://doi.org/10.61996/edu.v3i2.117

ABSTRACT

Background. Generative artificial intelligence (AI) is increasingly used in higher education, but evidence on its effect on creative thinking in structured classroom settings remains limited.

Objective. This study examined whether AI-assisted instruction improves creative problem-solving (CPS) and divergent thinking (DT) among undergraduate education students compared with conventional instruction.

Methods. A quasi-experimental pretest–posttest control group design was used with 120 education students at a private university in Palembang, Indonesia. Two intact sections of an Educational Psychology course were assigned to conditions by scheduling availability: an experimental group (n = 60) that followed a 12-week, three-phase AI-assisted learning model with ChatGPT-4, and a control group (n = 60) that received conventional instruction. CPS was measured with the researcher-developed Creative Problem-Solving Performance Inventory (CPSPI; $\alpha = 0.89$) and DT with an adapted Torrance Tests of Creative Thinking (TTCT; ICC = 0.91). Data were analyzed with mixed ANOVA, MANOVA, and ANCOVA.

Results. The group × time interaction was significant for CPS, $F(1,118) = 89.34$, $p < 0.001$, partial $\eta^2 = 0.431$, and for DT, $F(1,118) = 62.18$, $p < 0.001$, partial $\eta^2 = 0.345$. MANOVA showed a significant multivariate effect of group, Pillai's $V = 0.482$, $F(5,114) = 21.24$, $p < 0.001$. Effect sizes were large for CPS (Hedges' $g = 1.66$) and DT (Hedges' $g = 1.18$).

Conclusion. Structured use of generative AI in a university course was associated with greater gains in creative problem-solving and divergent thinking than conventional instruction. Because sections were not randomly assigned, causal claims should be cautious.

Copyright: © 2025 The Author(s). Published by Enigma Institute on behalf of Enigma in Education. The authors retain copyright of their work.

Access license: This is an open access article distributed under the terms of the Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, sharing, adaptation and reproduction in any medium, provided the original author and source are credited and any derivative work is distributed under the same license.

1. Introduction

AI has reshaped educational practice worldwide. Rapid advances in artificial intelligence (AI) have led educators and researchers to rethink traditional teaching approaches.[1] Generative AI systems such as ChatGPT are among the most influential developments, with striking abilities in language generation, problem analysis, and creative ideation.[2] The tools open new ways of enhancing learning, above all in domains that call for creative and higher-order thinking.

Creative problem-solving (CPS) has long been seen as a core competency for academic, professional, and personal success.[3] Generating novel and appropriate solutions to complex, ill-defined problems is regarded as essential in an uncertain modern world.[4,5] Traditional education, though, has often been criticized. It tends to stress convergent thinking and rote memorization over real creative capacity.[6] The tension between standardized outcomes and the cultivation of creative thinking is a persistent challenge for higher education.

Creativity is broadly defined as the production of ideas or outcomes that are both novel and useful.[7] It spans divergent thinking, convergent thinking, and evaluative processes.[8] Divergent thinking (DT), the generation of many solutions or ideas in response to an open-ended stimulus, has been widely studied as a key indicator of creative potential.[9] Its link with broader creative achievement is complex. Still, the empirical literature has supported it consistently.[10,11]

Scholars have shown considerable interest in AI in education. They have examined how it might support different aspects of teaching and learning.[12,13] Recent reviews describe both opportunities and challenges of AI-mediated pedagogy, with reported benefits that include personalized learning, immediate feedback, and scaffolding for complex cognitive tasks.[14,15] Empirical evidence on how generative AI affects creative thinking outcomes remains limited. This is the gap the present study seeks to address.

Interest in AI-enhanced education is growing. Yet few studies have examined rigorously how structured AI integration shapes students' creative problem-solving in authentic classrooms, and most existing work has looked at AI in content delivery or assessment, not as a tool for fostering creativity.[1] This study addresses the gap with a systematic 12-week intervention in which generative AI serves as a collaborative partner in creative problem-solving within a university Educational Psychology course.

CPS is the deliberate process of identifying, analyzing, and resolving problems by generating and evaluating novel solutions.[3] Treffinger, Isaksen, and Stead-Dorval proposed an influential framework with three broad stages. These are understanding the challenge, generating ideas, and preparing

for action.[16] Educators have widely adopted the framework as a structured way to teach creative thinking, and it has shaped many instructional programs for developing creative competencies.

Universities, researchers increasingly argue, must go beyond transmitting knowledge. They must build students' capacity for creative and critical thinking.[17] Beghetto and Kaufman's notion of "mini-c" creativity holds that personally meaningful creative insights, not eminent achievements, are the educational outcome that matters most.[18] The view fits current educational goals, which give priority to transferable cognitive skills usable across disciplines. Meta-analytic evidence indicates that creativity can be developed through training, with a moderate average effect (about 0.53 SD) that is, however, subject to publication bias.[19]

Guilford conceptualized divergent thinking as generating multiple alternatives in response to open-ended problems.[3] The Torrance Tests of Creative Thinking (TTCT) remain the most widely used measure of DT. They assess four dimensions: fluency (number of responses), flexibility (variety of categories), originality (statistical rarity of responses), and elaboration (detail added to ideas).[9] The TTCT has faced continuing psychometric scrutiny. Even so, meta-analytic evidence supports its predictive validity for creative achievement across several domains, although the associations are modest.[4,11,20]

Testing conditions matter. When divergent thinking is assessed in educational settings, time limits, instructional framing, and motivational context can all shape performance.[5] These considerations guided the assessment protocols of the present study, in which standardized administration procedures were used to protect the validity and reliability of measurement.

Large language models have changed the landscape of educational technology. Systematic reviews of AI in higher education find a main focus on intelligent tutoring systems, automated assessment, and adaptive learning platforms.[1] More recently, generative AI tools have drawn attention. Their potential to support higher-order cognitive processes such as analysis, synthesis, and creative thinking is now widely discussed.[12,15]

Several theoretical frameworks support integrating AI into pedagogy. Siemens' connectivism treats learning as the process of linking specialized nodes of information, a view that fits naturally with AI-augmented learning environments.[21] Vygotsky's zone of proximal development (ZPD) offers another lens. AI, on this view, can act as a scaffold that lifts students' cognitive capabilities above their current independent level.[22] Some scholars have championed AI as a transformative educational tool that could democratize access to personalized learning.[23,24]

Empirical studies remain scarce. Despite the theoretical perspectives above, little is known about how generative AI affects creative thinking in educational settings. Urban and colleagues offered early experimental evidence that ChatGPT can improve creative problem-solving among university students, yet they warned against overgeneralization and called for replication with larger samples and longer interventions.[2] Other recent experiments point to both promise and risk. AI-provided story ideas raised the rated creativity of short stories, especially for less creative writers, but made the stories more similar to one another.[25] A study in a college creativity course concluded that generative AI can support creative thinking but may also harm creativity and creative confidence.[26] In a randomized experiment on writing tasks, learners supported by ChatGPT improved their short-term performance but engaged less in metacognitive activity, a pattern the authors called metacognitive laziness.[27] This study responds with a sustained 12-week quasi-experimental intervention. The central research question is this. To what extent does AI-assisted instruction improve creative problem-solving performance and divergent thinking among undergraduate education students compared to conventional instruction?

2. Methods

2.1 Participants

In total, 120 undergraduate education students at a private university in Palembang, Indonesia, took part. They were enrolled in two sections of an Educational Psychology course in the same academic semester. One section (n = 60) formed the experimental group and the other (n = 60) the control group. Assignment of sections to conditions followed scheduling availability and was decided before the semester began. Eligible students were actively enrolled undergraduates with no prior formal training in AI-assisted learning methodologies. Table 1 presents the demographic characteristics. The Ethical Committee of CMHC Research Center, Palembang, Indonesia, granted ethical approval (approval number 2025/199). Every participant gave written informed consent before taking part.

Table 1. Demographic characteristics of participants.

Characteristic	Category	Experimental (*n* = 60)	Control (*n* = 60)	Total (*N* = 120)
Gender	Male	22 (36.7%)	24 (40.0%)	46 (38.3%)
	Female	38 (63.3%)	36 (60.0%)	74 (61.7%)
Age (years)	18–19	18 (30.0%)	20 (33.3%)	38 (31.7%)
	20–21	28 (46.7%)	26 (43.3%)	54 (45.0%)
	22–23	14 (23.3%)	14 (23.3%)	28 (23.3%)
	M (*SD*)	20.12 (1.24)	20.08 (1.31)	20.10 (1.27)
GPA	2.50–2.99	12 (20.0%)	14 (23.3%)	26 (21.7%)
	3.00–3.49	30 (50.0%)	28 (46.7%)	58 (48.3%)
	3.50–4.00	18 (30.0%)	18 (30.0%)	36 (30.0%)
	M (*SD*)	3.28 (0.38)	3.24 (0.41)	3.26 (0.39)

Note. Values are *n* (%) unless otherwise stated. *M* = mean; *SD* = standard deviation. Between-group comparisons: age, $t(118) = 0.17$, $p = 0.864$; GPA, $t(118) = 0.56$, $p = 0.577$.

Table 1 shows that the groups were comparable in gender distribution, age, and academic performance. Most participants were female (61.7%). This is typical of education programs. Mean age was 20.10 years (SD = 1.27), and mean GPA was 3.26 (SD = 0.39), which indicates a moderately high-performing sample. Independent samples t-tests found no statistically significant group differences in age, t(118) = 0.17, p = 0.864, or in GPA, t(118) = 0.56, p = 0.577.

2.2 Research design

The study used a quasi-experimental pretest-posttest control group design.[28] The unit of allocation was the course section (two sections), whereas the analyses were conducted at the student level. The independent variable was instructional approach (AI-assisted vs. conventional); the dependent variables were creative problem-solving performance and divergent thinking scores. Pretests were given in the first week of the semester. Posttests were collected in the final week, after the 12-week intervention period. This design allowed both between-group differences and within-group change over time to be examined.

2.3 Instruments

Creative Problem-Solving Performance Inventory (CPSPI). The CPSPI is a researcher-developed instrument that assesses students' ability to generate, evaluate, and implement creative solutions to educational problems, and it has 25 items in three subscales: problem identification (8 items), solution generation (9 items), and solution evaluation (8 items). Scoring uses a 4-point performance rubric (0–3). The total possible score is 75. Five educational psychologists reviewed the content, and every item had a content validity index (CVI) of 0.80 or higher. Internal consistency was excellent (α = 0.89). Test-retest reliability over a two-week interval was r = 0.84.

Adapted Torrance Tests of Creative Thinking (TTCT). Divergent thinking was assessed with an adapted version of the TTCT verbal battery.[9] The adapted instrument kept the four standard dimensions (fluency, flexibility, originality, and elaboration), but its prompts were placed in educational scenarios relevant to the course content, and two trained raters scored responses with the standardized TTCT scoring guidelines. Agreement between raters was assessed with intraclass correlation coefficients (ICC). The composite ICC was 0.91. Agreement was excellent.[29]

2.4 Procedures

The intervention ran for 12 weeks within the regular course schedule. In every session the experimental group followed a structured three-phase AI-assisted learning model. In Phase 1 (Problem Presentation, 20 minutes), the instructor introduced a complex educational scenario that required creative solutions, supplied the relevant theoretical context, and clarified the problem parameters. In Phase 2 (AI Exploration, 40 minutes), students used generative AI tools (ChatGPT-4) with structured prompting protocols. Training covered how to write effective prompts and how to evaluate AI-generated responses with care. AI outputs served as springboards for students' own creative ideation, never as final answers. Phase 3 (Critical Evaluation, 30 minutes) was collaborative: students discussed, evaluated, refined, and synthesized the ideas produced during AI exploration, applying course-relevant theoretical frameworks to judge the quality and feasibility of the proposed solutions.

The control group received conventional instruction on the same course content through traditional lecture-discussion methods, with no access to AI tools during class sessions. One instructor taught both groups. Both used identical course materials and were assessed against the same learning objectives, which limited confounding variables.

2.5 Data analysis

Data were analyzed with IBM SPSS Statistics Version 28. Preliminary analyses covered normality (Shapiro-Wilk test), homogeneity of variances (Levene's test), and homogeneity of covariance matrices (Box's M test). The primary analyses combined several inferential techniques.[30] A mixed ANOVA (2 × 2: Group × Time) tested interaction effects for CPS and DT total scores, and MANOVA assessed multivariate group differences across the outcome measures simultaneously. ANCOVA used pretest scores as covariates. It compared posttest group differences while holding baseline performance constant. Effect sizes were Hedges' g with 95% confidence intervals for between-group comparisons and partial η^2 for ANOVA-based analyses.[31] Post-hoc comparisons used a Bonferroni correction.

3. Results

3.1 Preliminary analyses

Assumptions were tested first. Shapiro-Wilk tests showed that all dependent variables were approximately normally distributed within each group ($p > 0.05$), and Levene's test confirmed homogeneity of variances for all measures ($p > 0.05$). Box's M test was non-significant, M = 28.34, F(21,58892) = 1.29, p = 0.173, which supports the assumption of homogeneous covariance matrices for the MANOVA. Independent samples t-tests on pretest scores showed no significant baseline differences between groups for any dependent variable (all $p > 0.05$). The between-group comparisons that followed were therefore valid.

3.2 Descriptive statistics

Table 2 presents descriptive statistics for all outcome measures. Both groups improved from pretest to posttest on every measure. The experimental group improved far more. Its CPS total score rose from M = 45.2 (SD = 8.3) to M = 78.6 (SD = 7.1), a 73.9% improvement, whereas the control group's CPS score went from M = 44.8 (SD = 8.1) to M = 52.3 (SD = 9.4), a 16.7% improvement. DT showed the same pattern. The experimental group's DT total rose from M = 112.4 (SD = 15.6) to M = 156.8 (SD = 14.2), a 39.5% increase. The control group's rose from M = 111.8 (SD = 16.1) to M = 128.4 (SD = 17.3), or 14.8%.

Table 2. Descriptive statistics for all outcome measures by group and time.

Measure	Experimental, pretest	Experimental, posttest	Control, pretest	Control, posttest	Hedges' *g*
CPS total	45.2 (8.3)	78.6 (7.1)	44.8 (8.1)	52.3 (9.4)	1.66
DT total	112.4 (15.6)	156.8 (14.2)	111.8 (16.1)	128.4 (17.3)	1.18
Originality	28.1 (5.2)	42.3 (4.8)	27.8 (5.4)	32.1 (5.9)	1.28
Fluency	31.2 (4.8)	42.1 (4.3)	30.9 (5.1)	35.6 (5.7)	0.97
Flexibility	26.8 (4.1)	36.2 (3.9)	26.5 (4.3)	30.4 (4.8)	0.84
Elaboration	26.3 (3.8)	36.2 (3.6)	26.6 (3.9)	30.3 (4.2)	0.74

Note. Values are *M* (*SD*); *n* = 60 per group. Hedges' *g* is given as printed by the authors for the posttest comparison. DT total is the sum of the four subscales. CPS = creative problem-solving; DT = divergent thinking.

Figure 1 makes the divergence between groups from pretest to posttest visible for both CPS and DT, with clearly steeper gains in the experimental group than in the control group.

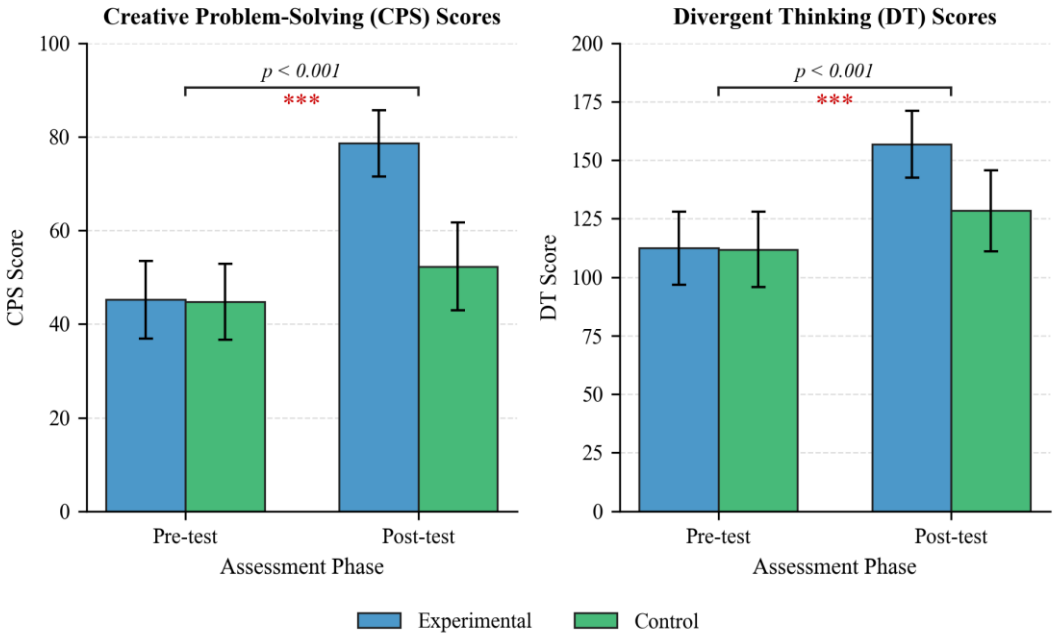

Figure 1. Comparison of CPS and DT scores by group and time. Bars show the means and error bars the standard deviations given in Table 2.

3.3 Multivariate analysis

MANOVA tested group differences across all posttest outcome measures at once. The multivariate effect of group on the combined dependent variables was statistically significant, Pillai's V = 0.482, F(5,114) = 21.24, p < 0.001, η^2p = 0.482. The creative thinking measures, taken together, therefore differed significantly between groups. This is consistent with an effect of the AI-assisted pedagogical approach.

3.4 Mixed ANOVA results

A mixed ANOVA (2 × 2: Group × Time) was run separately for CPS and DT total scores. For CPS, the Group × Time interaction was statistically significant, F(1,118) = 89.34, p < 0.001, partial η^2 = 0.431. Changes in CPS over time thus differed by group. The main effect of time was also significant, F(1,118) = 245.67, p < 0.001, partial η^2 = 0.676, as was the main effect of group, F(1,118) = 54.23, p < 0.001, partial η^2 = 0.315.

For DT total scores, the Group × Time interaction was significant as well, F(1,118) = 62.18, p < 0.001, partial η^2 = 0.345. Change from pretest to posttest in CPS and DT therefore differed between the AI-assisted and the conventional group. Improvement over the 12-week period was significantly greater in the experimental group.

3.5 ANCOVA results

ANCOVA compared posttest scores between groups while controlling for pretest performance. For CPS, the group effect remained significant after controlling for pretest scores, F(1,117) = 187.42, p < 0.001, partial η^2 = 0.616. Adjusted posttest means showed the experimental group scoring significantly higher than the control group (adjusted M = 78.4 vs. 52.5), with baseline differences held constant. The ANCOVA for DT total was also significant, F(1,117) = 124.56, p < 0.001, partial η^2 = 0.516. Table 3 summarizes the full results.

Table 3. Summary of statistical analyses.

Analysis	Test statistic	*p*	Effect size	95% CI
MANOVA (overall)	Pillai's V = 0.482, F(5,114) = 21.24	< 0.001	η^2_p = 0.482	[0.38, 0.57]
Mixed ANOVA (CPS)	F(1,118) = 89.34	< 0.001	η^2_p = 0.431	[0.32, 0.53]
Mixed ANOVA (DT)	F(1,118) = 62.18	< 0.001	η^2_p = 0.345	[0.24, 0.44]
ANCOVA (CPS)	F(1,117) = 187.42	< 0.001	η^2_p = 0.616	[0.52, 0.69]
ANCOVA (DT)	F(1,117) = 124.56	< 0.001	η^2_p = 0.516	[0.41, 0.61]
CPS (Hedges' *g*)	—	< 0.001	*g* = 1.66	[1.28, 2.04]
DT total (Hedges' *g*)	—	< 0.001	*g* = 1.18	[0.82, 1.54]

Note. η^2_p = partial η^2; CI = confidence interval. For the Hedges' *g* rows the effect size is the between-group posttest difference.

3.6 Effect sizes

Between-group effect sizes for posttest comparisons were computed as Hedges' g with 95% confidence intervals. CPS total showed a large effect, g = 1.66, 95% CI [1.28, 2.04]. The experimental group outperformed the control group by more than one and a half standard deviations. DT total also showed a large effect, g = 1.18, 95% CI [0.82, 1.54]. Among the DT subscales, originality had the largest effect (g = 1.28), followed by fluency (g = 0.97), flexibility (g = 0.84), and elaboration (g = 0.74). Effect sizes for all outcome measures are listed in Table 2 and shown in Figure 2.

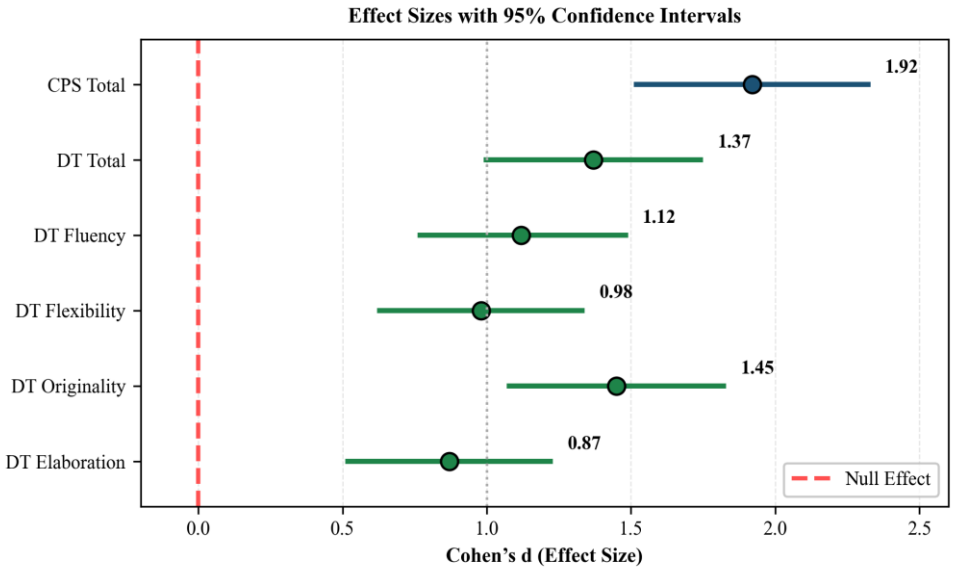

Figure 2. Forest plot of effect sizes with 95% confidence intervals.

Bonferroni-corrected post-hoc pairwise comparisons confirmed that the experimental group outperformed the control group on every individual DT subscale at posttest (all adjusted $p < .001$). Figure 3 compares the subscale profiles of the two groups at posttest in a radar chart.

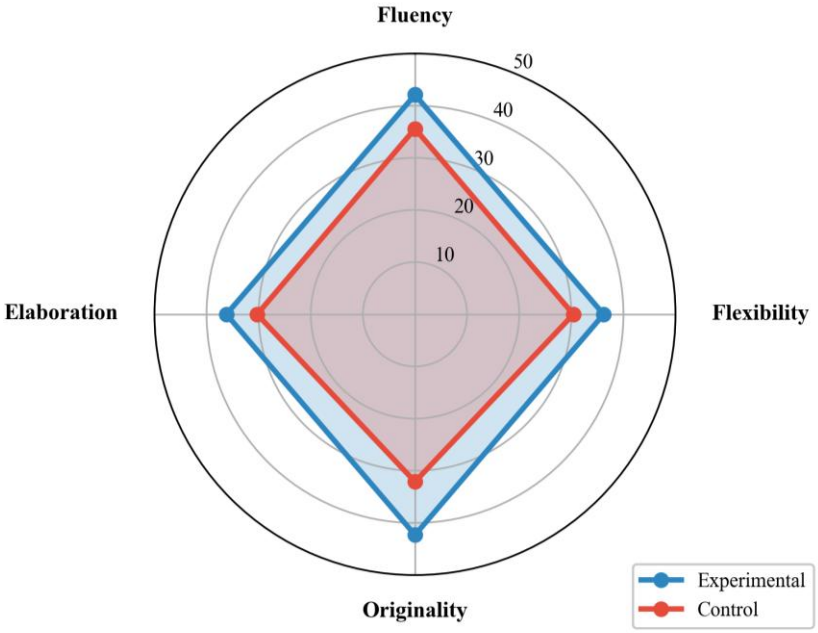

Figure 3. Radar chart of divergent thinking subscale scores at posttest by group. The chart shows the posttest means in Table 2.

4. Discussion

This study examined how effective generative AI is as a pedagogical tool for strengthening creative problem-solving and divergent thinking among undergraduate education students. The results suggest that structured AI-assisted instruction can improve students' creative thinking capacities, with large reported effect sizes on all outcome measures.

The AI-assisted group gained significantly more in CPS than the control group (Hedges' $g = 1.66$). The finding is notable for its size and its practical meaning. The effect exceeds the threshold usually regarded as "large" in educational research,[31] which suggests a meaningful influence of the intervention on students' ability to identify, analyze, and resolve problems creatively. The mixed ANOVA interaction ($F(1,118) = 89.34$, partial $\eta^2 = 0.431$) is consistent with a differential improvement associated with the intervention rather than with general maturation or repeated testing, which both groups experienced. Because two intact sections were assigned by scheduling availability, selection and section effects cannot be excluded.

These findings agree with, and extend, the work of Urban and colleagues, who reported gains in creative problem-solving after AI-assisted interventions.[2] Their experiment included 145 students (77 working with ChatGPT and 68 without),[2] whereas the present study included 120 students and a longer intervention (12 weeks vs. a single session). The longer period allows the present study to speak to change over a term, although whether the effects last and generalize remains to be tested. The structured three-phase model used here (problem presentation, AI exploration, and evaluation) may have strengthened the effects by keeping AI a scaffold for independent thinking rather than a substitute for creative effort.[8,21]

Effects differed across DT subscales, and the pattern sheds light on the cognitive mechanisms through which AI-assisted learning may strengthen creative thinking. Originality showed the largest effect ($g = 1.28$). Exposure to AI-generated responses may have widened students' conceptual repertoire. That would let them produce more statistically rare ideas. The result agrees with the connectivist view that learning arises from forming diverse conceptual connections,[21] and with creativity theories

that stress broad associative networks in original thought.[32,33] Doshi and Hauser found that AI-provided ideas raised rated creativity, especially for less creative writers, while making outputs more similar to one another;[25] the present study did not examine the diversity of ideas across students. Future work should. These mechanisms were not measured and remain hypotheses for future research.

The effect on fluency was also large (g = 0.97). AI-assisted practice seems to have raised students' ability to produce many ideas, possibly by modeling productive ideation strategies. Effects on flexibility (g = 0.84) and elaboration (g = 0.74) were somewhat smaller, though still sizable by conventional standards. Shifting between conceptual categories and developing ideas in detail may resist short-term instruction more. Fully developing these dimensions may take longer practice and deeper metacognitive engagement.[33,34]

A Vygotskian framework offers a convincing interpretation. Generative AI may act as a "more knowledgeable other": it extends students' zone of proximal development by supplying diverse perspectives and alternative solutions they might not produce alone.[22] The evaluation phase was designed so that students would internalize the expanded capabilities instead of simply deferring to AI-generated content. This follows recommendations in the literature on responsible AI integration in education, which stress preserving student agency and independent thinking.[12,14,35] Evidence that unguided use of AI dialogue systems may encourage over-reliance[36] and reduce metacognitive engagement[27] is a further reason for building an explicit evaluation phase into AI-assisted teaching. These processes were not measured in the present study.

For teaching practice, the three-phase model suggests a simple sequence. Present the problem. Then explore it with AI under a structured prompting protocol, and evaluate and refine the outputs collaboratively against course theory. The protocol asks students to treat AI responses as springboards, not as answers. This may help guard against the over-reliance and reduced metacognitive engagement described above.[27,36] Reviews of the ethics of generative AI in education also call for attention to data privacy, algorithmic bias, cognitive autonomy, and equity,[35] and instructors should address these issues explicitly in course design. Whether the safeguards were sufficient was not assessed here.

The ANCOVA results (partial η^2 = 0.616 for CPS; 0.516 for DT) indicate that the group effect remained after controlling for baseline performance and accounted for a large share of post-test variance. Whether the effect held across the range of initial ability levels was not tested, for example through a group × pre-test interaction. This matters for inclusive pedagogy. Can AI support creative development in students who do not initially see themselves as creative? That is a particularly promising direction for further study.[18]

The study has limitations. First, the quasi-experimental design suits classroom research but does not allow the causal inferences of a fully randomized trial. Baseline equivalence was established on measured variables, yet unmeasured confounds cannot be ruled out. Sections were also assigned to conditions by scheduling availability, not at random, so selection bias remains a possible threat to internal validity despite the baseline equivalence observed on measured variables. Second, the study took place at one institution in one course. Generalizability to other educational contexts is therefore limited. Third, ChatGPT-4 was the main AI tool, so results may not carry over to other AI platforms with different capabilities. Fourth, long-term retention of the gains was not assessed. Future research should test whether improvements in creative thinking persist after the intervention. Finally, the absence of a mixed-methods component limits what can be said about students' subjective experiences of AI-assisted learning. Reporting that follows a guideline for non-randomized designs, such as the TREND statement, would help readers judge the risk of selection bias.[37] Further limitations are that one instructor taught both groups, that the main outcome measure was developed by the researchers, that rater blinding and AI-use records are not reported, and that the analyses treated students as independent although two intact sections were allocated.

5. Conclusion

This study suggests that generative AI may strengthen students' creative problem-solving and divergent thinking when it is integrated into university teaching through a structured pedagogical framework. Effect sizes were large on all outcome measures. AI-assisted learning thus looks promising. It may help cultivate creative capacities in higher education. The three-phase instructional model of problem presentation, AI-guided exploration, and evaluation offers educators a practical framework, to be replicated and tested further, for using AI's potential while keeping students actively engaged and thinking carefully.

The findings carry important implications for educational policy and practice. As AI technologies spread, universities need evidence-based approaches. Merely adopting the technology is not enough; it has to be used thoughtfully for real cognitive development. The present results suggest that the pedagogical design that structures how students interact with, evaluate, and build on AI-generated content may matter more than the technology itself, although this was not tested because the study had no unstructured-AI condition. Future research should test whether the findings generalize across disciplines, institutions, and AI platforms, and whether the effects last. It should also examine moderating individual differences such as prior creative self-efficacy and technology acceptance and AI literacy.[38]

Declaration

Artificial Intelligence Use Statement

No generative artificial intelligence tool was used to write or revise the text, analyze the data, create images, or generate any other scholarly content of this manuscript.

Author Contribution Statement

Conceptualization, H.P.; methodology, H.P. and K.J.; investigation, H.P.; formal analysis, K.J.; data curation, H.P.; writing – original draft, H.P.; writing – review and editing, H.P. and K.J.; supervision, H.P.; project administration, H.P. All authors approved the final version of manuscript.

Funding Statement

This research received no external funding.

Conflict of Interest / Competing Interests

The authors declare that they have no competing financial or non-financial interests related to this work.

Ethics Approval and Consent to Participate

The study was approved by the Ethical Committee of CMHC Research Center, Indonesia (approval number 2025/199). All participants gave written informed consent before taking part.

Consent for Publication

Not applicable. The manuscript reports aggregate data only.

Data Availability Statement

The datasets generated and analyzed during this study can be obtained from the corresponding author on reasonable request.

Acknowledgments

None.

References

1. Zawacki-Richter O, Marín VI, Bond M, et al. Systematic review of research on artificial intelligence applications in higher education

– where are the educators? Int J Educ Technol High Educ. 2019;16:39. doi:10.1186/s41239-019-0171-0

2. Urban M, Děchtěrenko F, Lukavský J, et al. ChatGPT improves creative problem-solving performance in university students: An experimental study. Comput Educ. 2024;215:105031. doi:10.1016/j.compedu.2024.105031
3. Guilford JP. Creativity: Yesterday, today and tomorrow. J Creat Behav. 1967;1(1):3-14. doi:10.1002/j.2162-6057.1967.tb00002.x
4. Kim KH. Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). Creat Res J. 2006;18(1):3-14. doi:10.1207/s15326934crj1801_2
5. Said-Metwaly S, Fernández-Castilla B, Kyndt E, et al. Testing conditions and creative performance: Meta-analyses of the impact of time limits and instructions. Psychol Aesthet Creat Arts. 2020;14(1):15-38. doi:10.1037/aca0000244
6. Cropley AJ. Creativity in Education and Learning: A Guide for Teachers and Educators. London: Kogan Page; 2001.
7. Runco MA, Jaeger GJ. The standard definition of creativity. Creat Res J. 2012;24(1):92-96. doi:10.1080/10400419.2012.650092
8. Sternberg RJ, Lubart TI. Investing in creativity. Am Psychol. 1996;51(7):677-688. doi:10.1037/0003-066X.51.7.677
9. Torrance EP. The Torrance Tests of Creative Thinking: Norms-Technical Manual. Bensenville, IL: Scholastic Testing Service; 1974.
10. Amabile TM. Creativity in Context: Update to the Social Psychology of Creativity. Boulder, CO: Westview Press; 1996.
11. Kim KH. Meta-analyses of the relationship of creative achievement to both IQ and divergent thinking test scores. J Creat Behav. 2008;42(2):106-130. doi:10.1002/j.2162-6057.2008.tb01290.x
12. Kasneci E, Seßler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ. 2023;103:102274. doi:10.1016/j.lindif.2023.102274
13. Baidoo-Anu D, Ansah LO. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. J AI. 2023;7(1):52-62. doi:10.61969/jai.1337500
14. Lo CK. What is the impact of ChatGPT on education? A rapid review of the literature. Educ Sci. 2023;13(4):410. doi:10.3390/educsci13040410
15. Dwivedi YK, Kshetri N, Hughes L, et al. Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manage. 2023;71:102642. doi:10.1016/j.ijinfomgt.2023.102642
16. Treffinger DJ, Isaksen SG, Stead-Dorval KB. Creative Problem Solving: An Introduction. 4th ed. Waco, TX: Prufrock Press; 2006.
17. Plucker JA, Beghetto RA, Dow GT. Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. Educ Psychol. 2004;39(2):83-96. doi:10.1207/s15326985ep3902_1
18. Beghetto RA, Kaufman JC. Toward a broader conception of creativity: A case for "mini-c" creativity. Psychol Aesthet Creat Arts. 2007;1(2):73-79. doi:10.1037/1931-3896.1.2.73
19. Sio UN, Lortie-Forgues H. The impact of creativity training on creative performance: A meta-analytic review and critical evaluation of 5 decades of creativity training studies. Psychol Bull. 2024;150(5):554-585. doi:10.1037/bul0000432
20. Runco MA, Millar G, Acar S, et al. Torrance Tests of Creative Thinking as predictors of personal and public achievement: A fifty-year follow-up. Creat Res J. 2010;22(4):361-368. doi:10.1080/10400419.2010.523393
21. Siemens G. Connectivism: A learning theory for the digital age. Int J Instr Technol Distance Learn. 2005;2(1):3-10.
22. Vygotsky LS. Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press; 1978.
23. Seldon A, Abidoye O. The Fourth Education Revolution: Will Artificial Intelligence Liberate or Infantilise Humanity? Buckingham: University of Buckingham Press; 2018.
24. Luckin R, Holmes W, Griffiths M, Forcier LB. Intelligence Unleashed: An Argument for AI as a Tool for the Classroom. London: Pearson; 2016.
25. Doshi AR, Hauser OP. Generative AI enhances individual creativity but reduces the collective diversity of novel content. Sci Adv. 2024;10(28):eadn5290. doi:10.1126/sciadv.adn5290
26. Habib S, Vogel T, Anli X, et al. How does generative artificial intelligence impact student creativity? J Creat. 2024;34(1):100072. doi:10.1016/j.yjoc.2023.100072
27. Fan Y, Tang L, Le H, et al. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. Br J Educ Technol. 2025;56(2):489-530. doi:10.1111/bjet.13544
28. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 5th ed. Thousand Oaks, CA: Sage; 2018.
29. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012
30. Field A. Discovering Statistics Using IBM SPSS Statistics. 5th ed. London: Sage; 2018.
31. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
32. Sawyer RK. Explaining Creativity: The Science of Human Innovation. 2nd ed. New York: Oxford University Press; 2012.
33. Mumford MD, Medeiros KE, Partlow PJ. Creative thinking: Processes, strategies, and knowledge. J Creat Behav. 2012;46(1):30-47. doi:10.1002/jocb.003
34. Hennessey BA, Amabile TM. Creativity. Annu Rev Psychol. 2010;61:569-598. doi:10.1146/annurev.psych.093008.100416
35. García-López IM, Trujillo-Liñán L. Ethical and regulatory challenges of generative AI in education: a systematic review. Front Educ. 2025;10:1565938. doi:10.3389/feduc.2025.1565938
36. Zhai C, Wibowo S, Li LD. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. Smart Learn Environ. 2024;11:28. doi:10.1186/s40561-024-00316-7
37. Des Jarlais DC, Lyles C, Crepaz N, et al. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361-366. doi:10.2105/ajph.94.3.361
38. Cengiz S, Peker A. Generative artificial intelligence acceptance and artificial intelligence anxiety among university students: the sequential mediating role of attitudes toward artificial intelligence and literacy. Curr Psychol. 2025;44(9):7991-8000. doi:10.1007/s12144-025-07433-7