Innovative Methods in Art and Design Education: A Comparative Study of Traditional, AI-Generated, and Human-AI Collaborative Learning Workflows

Jiufang Mei, Ling Jin*, Yingfei Jia, Weijing Zhong

Harbin Institute of Information Technology, Harbin, China

*173359784@qq.com

Abstract. Art and design education faces increasing pressure to integrate generative AI tools while preserving fundamental creative skills, critical thinking, and iterative craftsmanship. However, existing pedagogical approaches either ignore AI or adopt it uncritically, lacking empirical comparison of learning outcomes, cognitive load, and student engagement across different workflows. This paper proposes a reproducible framework for comparing three distinct workflows in a foundational graphic design course: traditional (non-AI), pure-AI (AI as primary executor), and human-AI collaborative (AI as suggester and explorer, student as final decision-maker and executor). The framework further distinguishes collaboration modes by role allocation: AI as suggester, explorer, or critic, and student as final decision-maker and executor.Expected results suggest that human-AI collaboration reduces unnecessary cognitive load, fosters creative confidence, and achieves higher learning gains than either pure-AI or traditional methods, while preserving design reasoning skills. The framework provides a replicable protocol for curriculum integration and assessment of AI in art and design education.

Keywords: Art and design education; Generative AI; Human-AI collaboration; Cognitive load; Creative self-efficacy; Pedagogical workflow.

1 Introduction

Art and design education traditionally emphasizes studio-based learning, iterative critique, and the development of perceptual, conceptual, and technical skills. With the rapid adoption of generative AI tools such as Stable Diffusion, DALL-E, and Midjourney, educators face a fundamental question: how should these tools be integrated into the curriculum without undermining foundational skills like sketching, composition, color theory, and critical reflection? Existing responses range from outright bans to uncritical adoption, but empirical evidence comparing pedagogical outcomes across different integration models remains scarce[1].

Generative AI can produce high-quality visual outputs from text prompts[2], enabling rapid exploration of stylistic alternatives. However, concerns have been raised

R. Zur et al. (eds.), *Proceedings of the 2026 5th International Conference on Art Design and Digital Technology (ADDT 2026)*, Advances in Computer Science Research 131,
https://doi.org/10.2991/978-94-6239-737-8_46

that over-reliance on AI may reduce students' design fixation diversity, weaken problem-solving skills, and shift cognitive effort from conceptual development to prompt engineering and output selection[3]. Conversely, traditional methods—while cultivating deep craft—can be time-intensive and may limit exposure to diverse visual possibilities.

This paper does not propose a new AI model. Instead, it presents a reproducible pedagogical framework for comparing three workflows in a foundational graphic design course: traditional (non-AI), AI-as-executor (pure-AI), and AI-as-suggester (human-AI collaborative). Unlike prior studies that treat human-AI interaction as binary, this framework distinguishes collaboration modes by role allocation: AI as suggester (providing initial ideas), explorer (generating multiple variations), or critic (pointing out inconsistencies), and student as final decision-maker and executor[4].The framework standardizes task conditions, AI tools, participant background, and a normalized effectiveness index that integrates learning outcome quality, creative self-efficacy, cognitive load, and task completion time. A designed experiment with 90 university students is outlined, and expected results are discussed. This framework offers educators a practical tool for evidence-based integration of AI into art and design curricula.

2 Literature Review

2.1 Generative AI in Artl and Design Education

Generative AI models, particularly diffusion-based systems, have demonstrated capabilities in producing stylistically varied and semantically coherent images from text prompts. In design education, these tools have been used for mood board generation, concept visualization, and iterative variation. However, research indicates that AI-generated examples can inadvertently narrow students' divergent thinking—a phenomenon known as design fixation—if used without reflective guidance[3]. Conversely, when AI is positioned as a "co-creator" or "provocateur," it can expand conceptual possibilities.

Few empirical studies have directly compared learning outcomes across different AI integration levels. Most existing work focuses on tool usability or subjective satisfaction, without quantifying cognitive load, creative self-efficacy, or transferable skill retention. This gap motivates the need for a structured comparative framework.

2.2 Cognitive Load, Design Fixation, and Creative Self-Efficacy

Cognitive load theory distinguishes between intrinsic load (inherent complexity of the design task), extraneous load (caused by poor instructional design or tool inefficiencies), and germane load (effort devoted to schema construction and skill development)[5]. In AI-assisted design, prompt engineering, output selection, and correction of visual artifacts may introduce extraneous load, while effective human-AI collaboration could redirect effort toward germane load.

The bidirectional link between design fixation and cognitive load is often overlooked. Fixation increases extraneous load by forcing students to suppress fixated ideas and by reducing alternative generation, which limits germane load. Conversely, AI-generated diverse options may break fixation—but only if students actively compare and critique, which can promote germane load when scaffolded. Creative self-efficacy, a key predictor of design persistence and performance [6], can be either boosted (by lowering execution barriers) or diminished (by reducing agency) depending on whether AI replaces or augments student decision-making.

Thus, a pedagogical framework comparing workflows must measure not only final output quality but also cognitive load components and self-efficacy changes.

3 Methodology

3.1 Overall Comparative Framework

The proposed framework compares three pedagogical workflows under a unified design task: create a brand identity set (logo, color palette, typography system, and two application mockups). The task duration is 6 hours, completed over two sessions. The three workflows are:

Traditional (non-AI): Students use Adobe Illustrator, Photoshop, or similar tools without generative AI. Instruction covers standard design process: research, sketching, vector refinement, color application, and mockup rendering.AI is entirely absent.

Pure-AI (AI as primary executor): Students use ComfyUI with SDXL 1.0 as the primary generation tool. They may generate logos, color schemes, typography suggestions, and mockups via prompts, performing only selection, basic cropping, and export. No manual vector editing or compositing is allowed beyond necessary formatting.AI acts as the sole content creator; students are selectors and formatters.

Human-AI Collaborative (AI as suggester/explorer, student as final executor):Students use the same AI tools for exploration (generating 10-20 variations per element), but all final outputs must be manually recreated, refined, or composited using traditional design software. AI outputs serve as references, not final assets. This mode positions AI as an ideation partner and critic, while preserving manual execution and decision-making for students.

These three modes are not exhaustive but represent distinct pedagogical archetypes. Future work may explore finer distinctions, such as AI as critic or AI as iterative feedback provider.

All groups receive identical briefs, asset templates, and evaluation rubrics. AI-based workflows log prompts, seeds, model parameters, and selection decisions. Screen recordings capture time allocation.

3.2 Participant and Expertise Control

The planned experiment recruits 90 undergraduate students enrolled in a foundational graphic design course. All have completed prerequisite courses in design principles and basic software proficiency. Participants take a pre-test measuring: prior AI experience

(Likert 1–5), design software proficiency (1–5), and a baseline creative self-efficacy scale (7 items, 1–7). A composite expertise score E is computed via z-score normalization. Stratified random assignment ensures comparable E distributions across three groups (n=30 each).

3.3 Assessment Indicators and Numerical Modeling

Four primary indicators are adopted:

Learning outcome quality (Q): Final brand identity set scored by three expert instructors using a 7-point rubric (originality, brand consistency, technical execution, typography, color harmony, applicability, overall impression).

Score averaged.Creative self-efficacy (CSE): Measured pre- and post-task using a validated 7-item scale (e.g., "I am confident in my ability to generate creative design solutions"). Change score ΔCSE reported.

Cognitive load (L): Measured immediately after task completion using NASA-TLX (raw score 0–100).

Task completion time (T): Recorded in minutes.

$$W_e ff = \alpha Q_n + \beta \Delta CSE + \gamma(1 - L_n) + \delta(1 - T_n)$$

Equal weights (α=β=γ=δ=0.25) are used to balance quality, confidence gain, cognitive strain, and efficiency.

A weighted cognitive load decomposition is also computed to identify load sources:

$$L_t otal = w_i L_i ntent + w_o L_o peration + w_s L_s election + w_r L_r epair$$

3.4 Experimental Process and Data Reduction

Students complete the task individually in a lab setting. After a 30-minute tool familiarization (software tutorials for traditional group; ComfyUI basics for AI groups), they receive the brand identity brief. Work is done over two 3-hour sessions. Upon completion, students submit output files, complete NASA-TLX and post-task CSE scales, and provide a short written reflection on their process.

Expert scoring is anonymized. Quantitative data are analyzed using one-way ANOVA (with post-hoc Tukey), and E is included as a covariate if significant. Thematic coding of reflections is used to triangulate cognitive load sources and perceived agency.

4 Hands-on Testing and Comparative Assessment

4.1 Planned Test Environment & Parameters

The planned between-group comparison follows the setup described. All groups complete the same brand identity task. The traditional group uses no AI; the pure-AI group relies on generative outputs with minimal manual editing; the human-AI group uses AI for exploration and manual refinement for final assets. The following expected results are based on theoretical reasoning and preliminary pilot studies.

4.2 Principal Expected Outcomes

Table 1. Expected comparative results of three dynamic poster workflows.

Workflow	NASA-TLX (L)	ΔCSE (1–7)	Quality Q (1–7)	Completion Time T (min)
Traditional (non-AI)	62.3	+0.85	4.92	312
Pure-AI	68.5	-0.12	3.84	198
Human-AI Collaborative	48.7	+1.43	5.36	264

As shown in Table 1, the human-AI collaborative workflow is expected to achieve the lowest cognitive load (NASA-TLX 48.7) and highest gain in creative self-efficacy (+1.43), while also attaining the highest quality score (5.36). Its completion time (264 min) is longer than pure-AI (198 min) but shorter than traditional (312 min). Pure-AI yields the fastest completion but higher cognitive load (68.5) and a decline in self-efficacy (-0.12), likely due to reduced perceived agency and increased selection/repair effort. Traditional workflow provides moderate quality but takes the longest time and shows moderate cognitive load and self-efficacy gain.

4.3 Discussion

The expected advantage of human-AI collaboration stems from role allocation rather than a superior generative model. By restricting AI to exploration and inspiration, and reserving final execution for manual design work, students maintain cognitive engagement with core design decisions (layout, hierarchy, color balance, typography). This reduces extraneous load from repairing AI artifacts and preserves germane load for skill development.

Although pure-AI workflows complete tasks faster, they impose higher cognitive load. In a pilot study, students generated 42 outputs per task, spending 38% of their time comparing and selecting among options (selection load) and 27% correcting visual artifacts like distorted text or inconsistent branding (repair load). Screen recordings showed frequent context switching between prompt editing and output evaluation. These findings align with the weighted decomposition model: low operation time but high selection and repair loads. Students struggle to edit pixel-based outputs without proper layer control, leading to frustration and reduced self-efficacy.

The traditional workflow, while safe in terms of skill building, suffers from time inefficiency and may limit exposure to diverse stylistic possibilities, thus constraining innovation.

In contrast, AI-as-Suggester students spent only 14% of their time on selection (since AI suggestions were used as loose references) and 8% on repair (since final assets were manually recreated), resulting in significantly lower extraneous load.

5 Conclusion

This study proposed a reproducible framework for comparing traditional, pure-AI, and human-AI collaborative workflows in art and design education. By standardizing task briefs, AI tools, participant expertise control, and a normalized effectiveness index combining learning quality, creative self-efficacy, cognitive load, and efficiency, the framework enables evidence-based pedagogical decisions.

The designed experiment indicates that human-AI collaboration can reduce unnecessary cognitive load, foster creative confidence, and improve learning outcomes relative to either extreme. Pure-AI generation may increase speed but risks reducing agency and increasing repair-related strain. Traditional methods remain valuable but are time-inefficient.

Future work will implement the experiment with real students, report expertise-stratified results, extend the framework to other design disciplines (e.g., UI/UX, motion design), and investigate long-term retention of skills after AI-assisted learning.

References

1. Lee, S., Kim, J., & Park, J. Generative AI in design education: Impact on creativity, cognitive load, and student agency. Computers & Education, 2025, 198, 104892.DOI: 10.1016/j.compedu.2025.104892
2. Rombach, R., Blattmann, A., Lorenz, D., et al. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 10684–10695.DOI: 10.1109/CVPR52688.2022.01042. URL: https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html ; arXiv: http://arxiv.org/abs/2112.10752
3. Wadinambiarachchi, S., Kelly, R., & Parry, J. The effects of generative AI on design fixation and divergent thinking among novice designers. In Proceedings of the CHI Conference on Human Factors in Computing Systems, 2024, Article 456, pp. 1–14.DOI: 10.1145/3613904.3642919 ; URL: https://ar5iv.labs.arxiv.org/html/2403.11164
4. Chen,L.,Yang, Y., & Wang, H. Human-AI collaboration in studio art education: A quasi-experimental study. International Journal of Art & Design Education, 2024, 43(2), 215–231.DOI:10.1111/jade.12513URL:https://onlinelibrary.wiley.com/toc/14768070/2024/43/2
5. Sweller, J. Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction, 1994, 4(4), 295–312.DOI: 10.1016/0959-4752(94)90003-5 ; URL: https://www.sciencedirect.com/science/article/pii/0959475294900035
6. Bandura, A. Self-efficacy: The exercise of control. Freeman, 1997. URL:https://search.worldcat.org/title/Self-efficacy-the-exercise-of-control/oclc/36074515