Intelligent & Human Futures
WISDOM ACADEMIC https://doi.org/10.63808/ihf.v2i3.351 Volume 2 , Issue 3

Research Article

Artificial Intelligence and Future-Oriented Aesthetic Education: Design Thinking in Middle School Art Curriculum

Yijun Feng[1,*]
[1]Academy of Fine Arts, Capital Normal University, Beijing 100084, China
[*]**Correspondence:** Yijun Feng, 1091900274@qq.com

Abstract

Middle school art curricula in many systems still center on skill reproduction, which sits uneasily with the rapid diffusion of generative AI. This study develops a conceptual instructional model that embeds generative AI within the five-stage design thinking process (empathize, define, ideate, prototype, test) for an eighth-grade visual communication course. AI is assigned differentiated functions across the stages—analytical, retrieval, generative, assistive, and evaluative—while students retain responsibility for aesthetic judgment, justification of selections, and iterative revision. Three learning outcomes anchor the design and are operationalized in parallel measurement instruments: aesthetic judgment, creative problem-solving in visual communication, and stylistic self-awareness. Communicative intent is treated as the core sub-dimension of creative problem-solving rather than as a separate outcome, since intent is observable only through the design products and decisions that enact it. A quasi-experimental pretest-posttest pilot was conducted with 68 eighth-grade students (experimental group n = 34, control group n = 34) over a ten-week intervention at a single school, with both classes taught by the same art teacher. ANCOVA indicated that the experimental group outperformed the control group on aesthetic judgment, $F(1, 65) = 8.42$, $p = .005$, $\eta p^2 = .115$, and creative problem-solving, $F(1, 65) = 9.17$, $p = .003$, $\eta p^2 = .124$, with smaller yet significant gains on stylistic self-awareness. Thematic analysis of journals and interviews surfaced three patterns: shifted attention from execution to selection, prompt-based reasoning, and unease about authorship. Findings are presented as encouraging pilot evidence under specific local conditions, pending replication beyond the present site.

Keywords

Artificial intelligence in education, Design Thinking, Aesthetic education, Middle school art curriculum, Human–AI co-creation

1. Introduction

Middle school art curricula in many education systems continue to follow a transmissive model, with teachers demonstrating a technique and students replicating it. This method treats students as passive recipients of predetermined aesthetic standards (Tomljenović & Tatalović Vorkapić, 2020), and assessment that privileges technical correctness over conceptual originality limits the relevance of acquired skills to creative problem-solving outside the classroom.

Generative AI tools such as Midjourney and DALL-E have sharpened this tension. These systems produce sophisticated images from short prompts, unsettling the assumption that art competency is anchored in manual skill. Empirical work on Midjourney use in K-9 classrooms reports that the creative task shifts from hand-eye execution toward prompt crafting, critical evaluation of AI outputs, and aesthetic negotiation (Vartiainen et al., 2023). Reviews of AI in education observe that the educational value of such systems rests less on task completion than on their capacity to scaffold student reasoning and reflective engagement (Lampropoulos & Papadakis, 2025). Technical training retains value, yet technique alone no longer exhausts what an art curriculum should teach.

Middle school art instruction currently lacks a pedagogical framework that places creative judgment and problem-solving at its core while treating AI as a means rather than an end. Design thinking offers a candidate framework. A systematic review of 43 K-12 design thinking studies reported improvements in creative confidence, collaboration, and problem-solving, with the strongest gains at the secondary level (Li & Zhan, 2022). A subsequent meta-analysis of 25 design thinking interventions reported a weighted mean effect of $r = .436$ on student learning outcomes, with creative problem-solving consistently the most reliably affected outcome (Yu et al., 2024). Existing implementations in art classrooms have remained largely craft-based and have not examined how AI reshapes the operational logic of each design stage.

Building on these considerations, this study proposes an instructional model that combines generative AI with design thinking for middle school art classes, foregrounding aesthetic judgment and creative problem-solving rather than technical reproduction. A quasi-experimental pilot with 68 eighth-grade students examined the model's impact on three target outcomes: aesthetic judgment, creative problem-solving in visual communication, and stylistic self-awareness. Section 2 presents the theoretical base and instructional model. Section 3 describes the methodology. Section 4 reports results. Section 5 discusses implications, risks, and limitations.

2. Theoretical Framework and Instructional Model Development

2.1. Three theoretical sources anchor the model

The instructional model draws on three theoretical sources addressing how learning is organized, how AI tools function within that process, and what counts as a meaningful outcome.

Procedural support comes from design thinking theory as codified at the Stanford d.school, with its five-step process of empathize, define, ideate, prototype, and test. The key concept adopted here is that learning begins with a problem rather than a method, and that movement between problem and solution is non-linear and requires iteration. In an art curriculum, this means moving from "how to draw X" toward "how to express Y to audience Z." Within design thinking, the orchestration of problem framing, divergent ideation, and convergent evaluation has been treated as a structured form of creative problem-solving (Li & Zhan, 2022; Yu et al., 2024).

Constructivist learning theory grounds the role assigned to AI tools. Vygotsky's zone of proximal development holds that learners achieve more when supported by scaffolding bridging current and potential capability, and Papert's constructionism develops this idea by emphasizing knowledge built through tangible artifacts. AI scaffolding accordingly enhances students' capacity to generate visual outputs without replacing conceptualization, selection, and critical evaluation: a student who cannot yet produce a professional poster layout may use AI-generated drafts as material for aesthetic decisions while the intellectual work remains with the student.

The third foundation concerns target outcomes. Drawing on Parsons' (1987) four-stage model of aesthetic development, Eisner's (2002) arts-based conception of cognition, and the creative problem-solving framework of Treffinger et al. (2000), three distinguishable components are articulated. (i) Aesthetic judgment is the capacity to articulate why one design choice better serves a given communicative purpose than another, rooted in Parsons' (1987) account of mature judgment that moves beyond personal preference toward reasoned evaluation in light of intention and context. (ii) Creative problem-solving in visual communication is the capacity to move from a vague design problem to a defensible visual solution through iterative cycles of definition, ideation, prototyping, and refinement. From design thinking research, it is operationalized as the procedural integration of problem framing, generation of alternatives, comparative selection, and iterative refinement (Li & Zhan, 2022; Yu et al., 2024). From arts-based theory, it incorporates communicative intent—the deliberate arrangement of visual elements in service of a message directed at an

Intelligent & Human Futures
WISDOM ACADEMIC https://doi.org/10.63808/ihf.v2i3.351 Volume 2 , Issue 3

identified audience. In this study, communicative intent is not measured as a separate outcome but treated as the central sub-dimension of creative problem-solving and operationalized as the "communicative clarity" criterion within the rubric (Section 3.4), since intent is observable only through the design product and decisions taken to realize it. (iii) Stylistic self-awareness, drawing on Eisner's (2002) argument that artistic cognition involves the cultivation of a discriminating eye, captures the metacognitive dimension of aesthetic learning: recognizing one's own emerging visual preferences and reflecting on their communicative effects.

On this basis, the study addresses four research questions. RQ1 asks whether students taught with the AI-integrated design thinking model show greater gains in aesthetic judgment than students taught with the standard art curriculum. RQ2 asks whether the model produces greater gains in creative problem-solving in visual communication, with communicative intent operationalized through the communicative clarity rubric criterion. RQ3 asks whether the model produces greater gains in stylistic self-awareness. RQ4 asks what patterns of aesthetic decision-making, AI tool use, and self-perception as a visual communicator emerge under the model. RQ1–RQ3 are addressed quantitatively; RQ4 is addressed through thematic analysis of journals and interviews.

2.2. The model integrates AI across five design thinking phases

The model has five phases, each organized along three tracks: what the student does, what role the AI tool plays, and what role the teacher plays. The architecture appears in Figure 1.

Figure 1

AI-Integrated Design Thinking Instructional Model for Middle School Art Curriculum

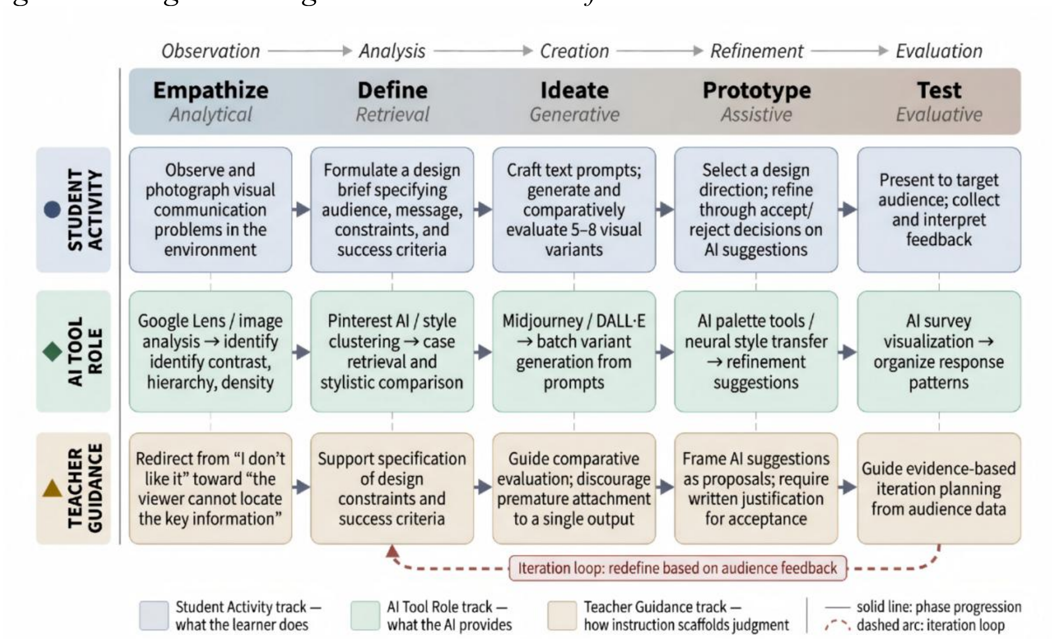

Figure 1 makes three analytical features of the model visible. First, the AI function type changes systematically across phases, moving from analytical and retrieval functions during problem framing, through generative and assistive functions during production, to an evaluative function during testing; this progression mirrors the cognitive arc from observation through creation to interpretation. Second, the teacher guidance track shows that scaffolding is differentiated rather than uniform: redirecting attention from preference to function in the empathize phase, requiring written justification for accepted AI suggestions in the prototype phase, and guiding evidence-based iteration in the test phase. Third, the dashed arc from test back to define encodes the iterative logic and identifies the point at which audience evidence may overturn the original design brief.

In the empathize phase, students identify real-life visual communication problems—ambiguous signs, disorganized notices, confusing infographics—and AI image recognition exposes compositional features such as contrast, hierarchy, and density. The teacher redirects subjective preference ("I don't like how it looks") toward functional analysis ("the viewer cannot locate the key information within three seconds"). The define phase consolidates observations into a design brief specifying audience, message, visual constraints, and success criteria, with AI case retrieval supporting comparison across stylistic directions. During ideation, generative tools produce five to eight visual variants per direction from student-authored prompts; quantity encourages comparative judgment. The prototype stage combines manual refinement with AI-assisted style transfer and palette generation; teachers frame AI suggestions as proposals requiring written justification before acceptance. In the test phase, completed works are presented to the target audience, AI visualization tools organize feedback patterns, and interpretation and revision planning remain with the student. Generative AI is deliberately withheld from the test phase, since its pedagogical value lies in student interpretation rather than automated synthesis.

The specific AI tools, function categories, student operations, and expected outputs for each phase are summarized in Table 1, which represents the intended design. The actual classroom enactment is reported in Table 2.

Table 1

AI Tool Function Mapping Across Five Instructional Phases (Intended Design)

Phase	AI Tool Example	Function Type	Student Operation	Expected Output
Empathize	Google Lens; AI image analysis (school iPad)	Analytical	Photograph visual problems; AI identifies	Problem identification report with annotated

			compositional elements (contrast, hierarchy, density)	images
Define	Pinterest (style clustering); curated case bank	Retrieval	Input keywords; compare and categorize retrieved cases	Design brief specifying audience, message, constraints, and success criteria
Ideate	Midjourney V6; DALL·E 3 (school-managed)	Generative	Craft prompts; generate 5–8 visual variants per direction	Annotated variant set with written rationale for top selections
Prototype	Color palette tools (Coolors, Khroma); style transfer	Assistive	Apply AI suggestions; accept or reject each modification	Refined design draft with decision log of accepted/rejected suggestions
Test	AI-assisted survey visualization (form summary)	Evaluative	Collect audience feedback; AI organizes response patterns	Iteration plan with specific revisions based on audience data

3. Methodology

3.1. Quasi-experimental pretest-posttest pilot design

The research design was a quasi-experimental pretest-posttest pilot with a non-equivalent control group. Two intact eighth-grade classes from the same school were assigned to experimental and control conditions. The experimental group followed the AI-integrated design thinking curriculum (Section 2.2); the control group followed the school's standard art curriculum on the same thematic unit (visual communication and poster design), so that between-group comparisons reflected instructional approach rather than content difference. The intervention spanned ten weeks. The study is framed throughout as a school-based pilot under specific local conditions rather than as evidence of general effectiveness; with one class per condition, treatment effects are confounded with class-level differences and cannot be cleanly separated from teacher-by-class interactions, school context, and cohort composition.

Two design choices reduced teacher effects. First, both classes were taught by the same art teacher—an in-service art specialist with eight years of experience who had completed in-service training in educational technology in the prior year—removing between-teacher differences as alternative explanations. Second, the two classes were timetabled on different days and in different rooms. To address teacher-side carryover, the teacher kept condition-specific lesson plans and slide decks separate and did not introduce AI tools or AI vocabulary in control sessions. Implementation fidelity was monitored via a weekly fidelity log and five non-consecutive structured classroom observations by the first author (one per fortnight, covering all five phases). Fidelity averaged 88% (range 82–94%), with the lowest rating during the first ideate-phase observation when prompt drafting overran. The protocol was approved by the Academic Ethics Committee of Capital Normal University (approval number: [CNU-AEC-2025-0186]); written informed consent was obtained from the legal guardians of all participants, and written assent was obtained from the students themselves, prior to data collection.

3.2. Participants

Participants were 68 eighth-grade students aged 13–14 years from a public middle school in an eastern Chinese city. The school was selected for two pragmatic criteria: a functional computer laboratory with stable internet access and administrative willingness to permit curricular adjustment. Each condition contained 34 students from intact classes, with balanced gender distribution (experimental: 18 female, 16 male; control: 17 female, 17 male). Independent-samples t-tests confirmed baseline equivalence: aesthetic judgment, $t(66) = -0.48$, $p = .631$; creative problem-solving, $t(66) = -0.46$, $p = .645$; stylistic self-awareness, $t(66) = -0.24$, $p = .813$. None of the participants reported prior hands-on experience with generative AI tools.

3.3. Curriculum implementation

The experimental curriculum comprised 20 instructional sessions of 45 minutes each (two sessions per week over ten weeks). The control class received an equivalent dose on the same thematic unit. The two conditions were matched on instructional time, content topic, final deliverable type (a campus-themed poster), and assessment task; they differed in the use of AI tools and design thinking process structure. As shown in Table 2, the five phases were distributed unevenly across the experimental curriculum, with the prototype phase receiving the longest allocation (six sessions) since refinement and decision-making on AI suggestions required sustained practice.

Table 2

Curriculum Implementation Schedule (Actual Classroom Enactment, Experimental Group)

Week	Phase	Key Activities	AI Tools Used	Sessions	Student Output
1–2	Empathize	Observe and photograph visual communication problems on campus and	Google Lens; AI image analysis	4	Problem identification report

		in the community; AI-assisted image analysis			
3	Define	Develop design briefs; retrieve and compare visual cases across styles	Pinterest; style clustering reference bank	2	Design brief
4–5	Ideate	Craft prompts; generate and evaluate 5–8 variants per direction	Midjourney V6; DALL·E 3	4	Annotated variant set with rationale
6–8	Prototype	Refine selected designs; apply and evaluate AI color and style suggestions	Coolors; Khroma; style transfer demos	6	Refined design with decision log
9–10	Test	Present to audience panel; collect feedback; draft iteration plan	Form survey with AI summary	4	Iteration plan and revised design

The control group followed the same thematic unit of campus poster design across 20 sessions, delivered through teacher demonstration, guided sketching, and manual execution without AI integration.

3.3.1. Implementation details, platform governance, and student safeguards

Because participants were minors and platform terms vary across deployments, operational details are reported in full. Tools and access: Google Lens via school iPads; Pinterest via a managed account with a teacher-curated feed; Midjourney V6 via a school-managed Discord workspace with a single licensed account under teacher supervision; DALL·E 3 via the school's ChatGPT account; Coolors and Khroma for palettes; a free style transfer demonstration site; and a standard form tool with AI summary feature for the test phase. All AI services were accessed exclusively from school-managed accounts on school devices, with no personal logins, and no personal data, school identifiers, or images of identifiable students were uploaded externally.

Each tool was introduced through a teacher-led demonstration with a short practice task at the start of the relevant phase; a printed prompt-writing guide was issued at the start of the ideate phase in Week 4. Three safeguards governed prompts and outputs: (i) all prompts were drafted in shared documents visible to the teacher, who reviewed them before submission; (ii) generated images were screened by the teacher before incorporation, with approximately six outputs across the ten weeks flagged and discarded for stereotyped depictions and used as discussion material rather than concealed (see Section 5.3); (iii) source photographs from the empathize phase were restricted to environmental and signage subjects, with any image containing identifiable faces of classmates excluded from upload. All five tool categories were available throughout the relevant phases without service interruption.

3.4. Data collection

Three quantitative instruments were administered at pretest and posttest, each operationalizing one target construct.

The Aesthetic Sensitivity Test was adapted from the Visual Aesthetic Sensitivity Test (VAST; Götz et al., 1979), which uses forced-choice comparisons between paired designs differing in compositional balance, hierarchy, and color harmony. Adaptation proceeded in three steps: 18 of the original 30 paired items were retained on content-fit grounds; 12 new items were constructed by the research team to cover figure-ground contrast, typographic hierarchy, and color-coded information grouping; and three doctoral-level art education specialists rated each item on a 1–4 content validity scale, with only items receiving a mean of 3 or above from at least two of three specialists retained. The final 30-item instrument was piloted with 24 non-participating eighth-graders, with Cronbach's $\alpha = .83$. A representative item presents two poster layouts differing in headline placement relative to the focal image, with the student selecting the layout that better directs viewer attention to the message.

Creative problem-solving was assessed through a parallel design task scored on a four-dimension rubric adapted from Treffinger et al.'s (2000) creative problem-solving framework, integrating procedural categories of design thinking (Li & Zhan, 2022) with product-centered dimensions used in K-12 design assessment (Yu et al., 2024). The four dimensions are problem definition, solution originality, communicative clarity, and iterative refinement, each rated on a five-point scale. The communicative clarity dimension carries the operational weight of the communicative intent construct introduced in Section 2.1 and assesses how well the visual product enacts a defensible relationship among audience, message, and visual form. Two raters, blinded to group assignment, independently scored all tasks. Cohen's κ reached .78 for problem definition, .72 for solution originality, .81 for communicative clarity, and .74 for iterative refinement, all exceeding the .70 threshold. Sample descriptors appear in Table 3.

Stylistic self-awareness was measured through a ten-item Likert questionnaire adapted from existing reflective awareness scales used in arts and design education research, capturing three facets: recognition of one's own visual preferences (e.g., "I can describe the visual style I tend to favor"), recognition of their communicative effects (e.g., "I can explain how my visual choices affect what the viewer notices first"), and reflective adjustment in light of audience response (e.g., "When audience feedback indicates a different reading than I

WISDOM ACADEMIC https://doi.org/10.63808/ihf.v2i3.351 Volume 2 , Issue 3

intended, I revise my design choices”). Pilot reliability was Cronbach’s α = .79. The construct lineage runs from Eisner’s (2002) account of artistic cognition through subsequent reflective awareness scales in design education, adapted here to the AI-supported design context.

Qualitative data comprised weekly design journals from all 34 experimental students and semi-structured interviews (25–40 minutes each) with eight students purposively selected to represent the range of pretest aesthetic judgment scores.

Table 3

Sample Rubric Descriptors for the Creative Problem-Solving Task

Dimension	Score 1 (Emerging)	Score 3 (Developing)	Score 5 (Proficient)
Problem definition	Problem stated only as a topic; no audience or success criteria.	Audience and message identified; success criteria implicit or partial.	Brief specifies audience, message, constraints, and explicit success criteria.
Solution originality	Solution closely mirrors a single retrieved or generated reference.	Solution combines elements from two or more references with limited recombination.	Solution shows defensible recombination or transformation in relation to the brief.
Communicative clarity (operationalizes communicative intent)	Intended message cannot be inferred without verbal explanation.	Message can be inferred but requires extended viewing or recovers only partially.	Message unambiguously legible to target audience within three seconds; hierarchy aligned to intent.
Iterative refinement	No documented revision; final product is the first variant.	One revision documented; rationale stated only in execution terms.	Two or more revisions documented; each justified in audience- or message-related terms.

3.5. Analyses

Between-group differences on posttest scores were examined using ANCOVA with pretest scores as covariate. Assumptions of normality, homogeneity of variance, and homogeneity of regression slopes were verified through Shapiro-Wilk tests, Levene’s tests, and group-by-covariate interaction terms; no violations were detected. Effect sizes are reported as partial eta-squared (small/medium/large thresholds: .01/.06/.14). Analyses were conducted in SPSS 27 with α = .05. Qualitative data were analyzed thematically following Braun and Clarke’s six-phase procedure. A second researcher independently coded a random 25% subset of transcripts; coding discrepancies were resolved through discussion until consensus.

4. Results

4.1. Quantitative outcomes

Descriptive statistics and ANCOVA outcomes appear in Table 4. ANCOVA controlling for pretest scores indicated a significant advantage for the experimental group on aesthetic judgment, $F(1, 65) = 8.42$, $p = .005$, $\eta p^2 = .115$ (medium effect); creative problem-solving, $F(1, 65) = 9.17$, $p = .003$, $\eta p^2 = .124$ (medium effect); and stylistic self-awareness, $F(1, 65) = 4.83$, $p = .032$, $\eta p^2 = .069$ (small to medium effect).

Table 4

Descriptive Statistics and ANCOVA Results for the Three Outcome Variables

Variable	EG Pretest M (SD)	EG Posttest M (SD)	CG Pretest M (SD)	CG Posttest M (SD)	F	p	ηp²
Aesthetic judgment	18.32 (3.41)	23.76 (3.08)	18.71 (3.25)	20.53 (3.17)	8.42	.005	.115
Creative problem-solving	12.44 (2.18)	16.91 (2.35)	12.68 (2.09)	14.12 (2.27)	9.17	.003	.124
Stylistic self-awareness	32.15 (5.62)	37.88 (5.31)	32.47 (5.48)	34.06 (5.39)	4.83	.032	.069

Note. EG = experimental group (n = 34); CG = control group (n = 34). Score ranges: 0–30 (aesthetic judgment), 4–20 (creative problem-solving), 10–50 (stylistic self-awareness).

Per-dimension statistics for the creative problem-solving rubric appear in Table 5. The experimental group showed the largest gains on problem definition (gain = 1.45) and iterative refinement (gain = 1.10), an intermediate gain on communicative clarity (gain = 1.18), and a more modest gain on solution originality (gain = 0.74). Control group gains did not exceed 0.42 on any dimension. The pattern indicates that the model trained the procedural and audience-oriented dimensions of creative work more strongly than novelty generation.

Table 5

Pretest and Posttest Means for the Four Creative Problem-Solving Rubric Dimensions

Rubric dimension	EG pretest M	EG posttest M	EG gain	CG pretest M	CG posttest M	CG gain
Problem definition	3.05	4.50	1.45	3.10	3.52	0.42

Solution originality	3.17	3.91	0.74	3.20	3.51	0.31
Communicative clarity	3.12	4.30	1.18	3.18	3.58	0.40
Iterative refinement	3.10	4.20	1.10	3.20	3.51	0.31
Total (sum of 4 dimensions)	12.44	16.91	4.47	12.68	14.12	1.44

Note. Each dimension rated on a 1–5 scale; total scores range 4–20. EG and CG totals match Table 4.

4.2. Three themes from journals and interviews

Three recurring themes emerged across the experimental group. First, students' attention shifted from execution to selection: early entries described frustration with drawing skill (clean lines, perspective), while later entries moved toward comparison-based reasoning, with students articulating why one AI variant fit the intended audience better than another. One student wrote in week six that the choice between two generated posters mattered more than the brushwork, since the wrong choice would fail the viewer regardless of execution. Second, prompt-based reasoning emerged as a form of design thinking: students moved from short descriptive prompts toward longer prompts specifying audience, mood, and compositional constraint, and six of the eight interviewed students described prompt revision as similar to redrafting writing—suggesting that prompt crafting functioned as an externalization of aesthetic intent—though two students reported feeling constrained by the prompt interface as a filter blocking ideas they could not verbalize. Third, unease about authorship surfaced: several students reported ambivalence about claiming AI-generated outputs as their own work even after substantial refinement, and this unease was more pronounced among students with stronger pretest drawing ability who saw the shift from manual production as a loss of identifiable craft. The tension did not resolve over the ten weeks, indicating that the model surfaces rather than settles the authorship question.

4.3. Quantitative and qualitative convergence

The two strands converged. Measurable gains in aesthetic judgment and creative problem-solving were accompanied by student accounts of attentional reorientation from execution to selection; intermediate gains on the communicative clarity dimension were accompanied by qualitative evidence of prompt-based externalization of audience-directed intent. The smaller effect on stylistic self-awareness is consistent with the qualitative finding that authorship unease remained unresolved, since stylistic self-awareness presupposes a settled sense of ownership the ten-week intervention did not fully establish. The divergence between strong gains on procedural dimensions and weaker gains on originality suggests that the model, in its current form, supports structured reasoning more effectively than novelty generation.

5. Discussion

5.1. Findings under specific local conditions

The results provide encouraging initial evidence consistent with the central design choice: positioning AI as a source of raw material subject to student judgment rather than a producer of finished work. Within the bounds of a pilot at one school with two intact eighth-grade classes over ten weeks in a single domain, the medium effects on aesthetic judgment and creative problem-solving indicate that an instructional sequence organized around selection, justification, and iterative revision can produce measurable gains on cognitive dimensions not directly addressed in the control condition. The weighted mean effect of r = .436 in a meta-analysis of 25 design thinking interventions (Yu et al., 2024) provides a reference point, and the effect sizes obtained here fall within a comparable range. Prior work on constructivist AI integration in art education has reported that students with limited technical ability engage in meaningful critical analysis when instruction positions AI as a starting point rather than an end point (Pavlik & Pavlik, 2024), consistent with the qualitative theme of attentional reorientation. Taken together, these findings are consistent with a plausible and testable instructional model and offer encouraging initial evidence; they do not yet establish a settled curricular solution.

5.2. Technical skill retains value alongside cognitive judgment

The findings invite a measured claim. Technical skill retains instructional value even when generative tools are available, since manual execution develops perceptual discrimination, sustained attention, and embodied understanding of material constraints that inform aesthetic judgment. The smaller gains on stylistic self-awareness and the persistent authorship unease among students with stronger drawing ability suggest that complete displacement of technical training risks undermining rather than advancing aesthetic development. This reading aligns with Park's (2023) call to reconceptualize human–AI relations in art education as collaborative and symbiotic rather than purely instrumental, recognizing AI not as a substitute for artistic skill but as a partner whose value depends on what students bring to the encounter. The model is therefore better understood as one that rebalances rather than replaces the place of technical work, with technique and judgment treated as complementary.

5.3. Ethical, bias-related, and access-related concerns

A complete account of AI integration in art education must attend to risks the instructional model alone cannot resolve. Generative models trained on image corpora that over-represent certain cultural, racial, and gender patterns reproduce these patterns in ways that can reinforce stereotypes when uncritically adopted. In this study, three of eight interviewed students commented that prompts involving occupational roles tended to return images skewed toward particular demographic profiles, and a consistent pattern of stereotyped depiction was visible in the approximately six AI outputs flagged and discarded by the teacher (Section 3.3.1). Teachers addressed these instances through explicit class discussion rather than concealment, but the pattern highlights the need for sustained critical framing. Over-reliance on AI outputs also presents a risk, since efficient variant generation can crowd out slower forms of visual exploration. Recent work cautions that sustainable classroom integration depends on data privacy, age-appropriate use, and teacher mediation rather than tool availability alone (Uğraş et al., 2024). Infrastructure inequality remains a structural concern, since reliable devices, network access, and platform permission differ sharply across school systems (Strycker, 2020). The authorship question also has pedagogical consequences, since assessment that credits AI-assisted work without explicit treatment of provenance can erode students' sense of creative ownership.

5.4. Implementation depends on teacher preparation and infrastructure

K-12 art teachers report limited technology experiences in their own initial preparation, and the instructional uses art teachers most frequently reported in Strycker's (2020) study were presentation and resource access rather than technologies that scaffold higher-order cognitive work. Reviews of educational AI emphasize that teacher capacity to orchestrate AI interactions, rather than tool availability, determines whether benefits are realized (Lampropoulos & Papadakis, 2025; Uğraş et al., 2024). The present model further requires familiarity with prompt engineering, output evaluation, and pedagogical rationale for each AI role assignment. Teacher preparation programs addressing these demands are a precondition for scaling beyond favorable study sites.

5.5. Limitations

Five limitations bound the contribution to that of an exploratory pilot. The study was conducted at a single school with two intact classes, constraining generalizability and limiting the extent to which classroom-level confounds can be ruled out, even with the same teacher delivering both conditions. The ten-week duration captured only immediate post-intervention outcomes. The task domain was limited to visual communication and poster design. The single eighth-grade cohort in one cultural and institutional setting is unlikely to be representative. Instrument adaptation, while supported by pilot reliability, was not accompanied by a full validation study, and the stylistic self-awareness scale would benefit from confirmatory factor analysis on a larger sample. These concerns align with broader limitations in design-based research (Tinoca et al., 2022). Multi-site replication, longer follow-up, and broader domain coverage are priorities for subsequent work.

6. Conclusion

This study has proposed and pilot-tested an instructional model that integrates generative AI within a design thinking framework for middle school art education. The model organizes learning into five iterative phases—empathize, define, ideate, prototype, test—with AI assigned differentiated analytical, retrieval, generative, assistive, and evaluative functions, and with student attention directed toward selection, justification, and revision. Quasi-experimental evidence from 68 eighth-grade students taught over ten weeks by the same teacher across both conditions indicated medium gains on aesthetic judgment and creative problem-solving and smaller gains on stylistic self-awareness in the experimental condition. Qualitative evidence further suggested a reorientation of attention from execution to comparative judgment, prompt crafting as a form of audience-directed intent, and unresolved unease about authorship that the intervention surfaced rather than settled.

The contribution is threefold. The empirical contribution is a pilot dataset showing that a ten-week AI-integrated design thinking sequence is feasible at school scale, produces measurable proximal effects on three theoretically distinct outcomes, and surfaces specific patterns of student response. The theoretical contribution is a tightened three-component account of the outcomes AI-supported aesthetic education ought to target, with communicative intent specified as the central sub-dimension of creative problem-solving rather than a parallel outcome. The practical contribution is a phase-by-phase model assigning AI a different functional role at each stage, withholding generative AI from the test phase to preserve student interpretation, and treating decision logs and written justifications as first-class assessment artifacts. Three implementation implications follow: the prototype phase warrants sustained time; teacher preparation must extend beyond tool familiarity to pedagogical rationale; assessment should capture process artifacts alongside final products.

The findings are bounded by a single school site, two intact classes, a ten-week window, the visual communication and poster domain, and a single eighth-grade cohort. Extension to other art domains, longer time horizons, multi-site replication, and successive AI tools awaits future research.

Acknowledgments

Not applicable.

Author contributions

Y.F. is the sole author of this manuscript and performed all roles: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, and project administration. The author has read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data availability

The data presented in this study are available on request from the corresponding author. The data are not publicly available because they were collected from minor participants and contain information that could compromise their privacy.

Declarations

Conflict of interest

The author declares that there is no conflict of interest.

Ethics approval

The study protocol was reviewed and approved by the Academic Ethics Committee of [Capital Normal University] (approval number: [CNU-AEC-2025-0186]; date of approval: 1 September 2025). The study was conducted in accordance with the Declaration of Helsinki and the relevant institutional guidelines and regulations.

Consent to participate

Written informed consent was obtained from the legal guardians of all participating students, and written assent was obtained from the students themselves, prior to data collection.

Consent for publication

Not applicable. The manuscript contains no identifiable personal data or images of individual participants.

Additional information

Received: 9 March 2026

Accepted: 11 May 2026

Published online: 3 September 2026

Publisher's Note

Wisdom Academic Press Ltd. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Eisner, E. W. (2002). *The arts and the creation of mind*. Yale University Press.

[2] Götz, K. O., Lynn, R., Borisy, A. R., & Eysenck, H. J. (1979). A new visual aesthetic sensitivity test: I. Construction and psychometric properties. *Perceptual and Motor Skills, 49*(3), 795–802. https://doi.org/10.2466/pms.1979.49.3.795

[3] Lampropoulos, G., & Papadakis, S. (2025). The educational value of artificial intelligence and social robots. In G. Lampropoulos & S. Papadakis (Eds.), *Social robots in education* (Studies in Computational Intelligence, Vol. 1194, pp. 3–15). Springer. https://doi.org/10.1007/978-3-031-82915-4_1

[4] Li, T., & Zhan, Z. (2022). A systematic review on design thinking integrated learning in K-12 education. *Applied Sciences, 12*(16), 8077. https://doi.org/10.3390/app12168077

[5] Park, Y. S. (2023). Creative and critical entanglements with AI in art education. *Studies in Art Education, 64*(4), 406–425. https://doi.org/10.1080/00393541.2023.2255084

[6] Parsons, M. J. (1987). *How we understand art: A cognitive developmental account of aesthetic experience*. Cambridge University Press.

[7] Pavlik, J. V., & Pavlik, O. M. (2024). Art education and generative AI: An exploratory study in constructivist learning and visualization automation for the classroom. *Creative Education, 15*(4), 601–616. https://doi.org/10.4236/ce.2024.154037

[8] Strycker, J. (2020). K-12 art teacher technology use and preparation. *Heliyon, 6*(7), e04358. https://doi.org/10.1016/j.heliyon.2020.e04358

[9] Tinoca, L., Piedade, J., Santos, S., Pedro, A., & Gomes, S. (2022). Design-based research in the educational field: A systematic literature review. *Education Sciences, 12*(6), 410. https://doi.org/10.3390/educsci12060410

[10] Tomljenović, Z., & Tatalović Vorkapić, S. (2020). Constructivism in visual arts classes. *Center for Educational Policy Studies Journal, 10*(4), 13–32. https://doi.org/10.26529/cepsj.913

[11] Treffinger, D. J., Isaksen, S. G., & Dorval, K. B. (2000). *Creative problem solving: An introduction* (3rd ed.). Prufrock Press.

[12] Uğraş, H., Uğraş, M., Papadakis, S., & Kalogiannakis, M. (2024). ChatGPT-supported education in primary schools: The potential of ChatGPT for sustainable practices. *Sustainability, 16*(22), 9855. https://doi.org/10.3390/su16229855

[13] Vartiainen, H., Tedre, M., & Jormanainen, I. (2023). Co-creating digital art with generative AI in K-9 education: Socio-material insights. *International Journal of Education through Art, 19*(3), 405–423. https://doi.org/10.1386/eta_00143_1

[14] Yu, Q., Yu, K., & Lin, R. (2024). A meta-analysis of the effects of design thinking on student learning. *Humanities and Social Sciences Communications, 11*, 742. https://doi.org/10.1057/s41599-024-03237-5