المجلة السعودية للفن والتصميم، 2026، المجلد 6، العدد 1، 309 – 348
https://pnu.edu.sa/ar/Departments/ScientificSocieties/SaudiArtDesignJournal/Pages/home.aspx

عدالة تقييم الإبداع في الإنتاج الفني المدعوم بالذكاء الاصطناعي مقارنة بالإنتاج الفني التقليدي

Assessment Equity of Creativity in AI-Assisted Art Production Compared with Traditional Art Production

Dol 10.57194/ 2351–006–001–009

نوال الغامدي
ORCID ID https://orcid.org/0009-0004-7582-4503
Email: naghmade@uqu.edu.sa
قسم الفنون البصرية، كلية التصميم والفنون، جامعة أم القرى، المملكة العربية السعودية.

Nawal Alghamdi
ORCID ID https://orcid.org/0009-0004-7582-4503
Email: naghmade@uqu.edu.sa
(PhD) Department of Visual Arts, College of Designs and Arts, Umm Al-Qura University, Saudi Arabia

رقم البحث Article No	الاستقبال Received	القبول Accepted	النشر Published
2026 – 19	6 March 2026	20 May 2026	June 2026

الملخص

هدفت هذه الدراسة إلى مقارنة نتائج الأداء الفني وتصورات الطلاب نحو عدالة التقييم والرضا التعليمي والإبداع والدافعية بين مجموعتين: إحداهما تعتمد أساليب الرسم والتلوين اليدوي، والأخرى تستخدم أدوات التصميم المدعومة بالذكاء الاصطناعي، في سياق مقرر (التعبير اللوني) بجامعة أم القرى بالمملكة العربية السعودية. اعتمدت الدراسة تصميمًا شبه تجريبي تكوَّن من مجموعتين غير متكافئتين مع قياس بعدي، وشملت العينة 40 طالبة جامعية قُسِّمت بالتساوي إلى مجموعة الرسم اليدوي (حجم العينة=20) ومجموعة التصميم بمساعدة الذكاء الاصطناعي (حجم العينة=20)، اشتغلت كلتاهما ضمن إطار موضوعي موحد على مدار 13 أسبوعًا. وقُيِّم الأداء الفني باستخدام بطاقة تقييم مُرجَّحة المعايير تغطي أربعة أبعاد: الفكرة المبتكرة، والخطوط والتكوين الفني، والتعبير اللوني، والمعالجة البصرية، وذلك وفق بروتوكول تقييم مُحايد. وقِيْسَتْ تصورات الطالبات عبر استبيان مُعتمد يتناول ثلاثة محاور: الإنصاف، والرضا عن التجربة التعليمية، والإبداع والتحفيز. وتشير النتائج إلى أن دعم الذكاء الاصطناعي أسهم في تعزيز توليد الأفكار المبتكرة وتوسيع الاستكشاف المفاهيمي في المراحل الأولى من العملية الإبداعية، دون أن يمتد هذا الأثر إلى المعايير الفنية التنفيذية أو تصورات الطلاب المتعلقة بالعدالة والرضا والدافعية. وبناء على ذلك، يُوصى بتوظيف أدوات الذكاء الاصطناعي في التعليم الفني داعمًا للمرحلة التأسيسية الإبداعية، مع صياغة معايير تقييم واضحة وبنية تقييمية منظمة تحفظ للطالب إحساسه بالتقييم المنصف.

الكلمات المفتاحية:
التفكير الإبداعي، أساليب تدريس الفنون في التعليم العالي، الذكاء الاصطناعي في دورات الاستوديو.

Abstract

This study aimed to compare artistic performance outcomes and student perceptions of assessment equity, educational satisfaction, creativity, and motivation between two groups: one using manual drawing and colouring methods, and the other using AI-assisted design tools, within the context of a Colour Expression course at Umm Al-Qura University, Saudi Arabia.

A quasi-experimental non-equivalent groups design with post-test measures was adopted. The study sample consisted of 40 female undergraduate students, equally divided into a manual drawing group (n = 20) and an AI-assisted design group (n = 20), both working within a unified thematic framework over 13 weeks.

Artistic performance was evaluated using a criterion-weighted rubric covering four dimensions: innovative idea, lines and artistic composition, colour expression, and visual processing, applied under a blind assessment protocol. Student perceptions were measured using a validated post-experiment questionnaire addressing three dimensions: fairness, satisfaction with the educational experience, and creativity and motivation.

The findings suggest that AI support contributed to enhancing the generation of innovative ideas and broadening conceptual exploration in the early stages of the creative process, without this effect extending to technical execution criteria or students' perceptions of fairness, satisfaction, and motivation. Accordingly, it is recommended that AI tools be integrated into art education as a scaffold for the foundational creative phase, alongside clearly defined assessment criteria and a well-structured evaluation framework that preserves students' sense of equitable assessment.

Keywords
Place, Perception, Space, Urban, sculpture Architecture

Introduction

In the past few years, the integration of generative artificial intelligence (GenAI) into visual arts education has redefined the relationship between creativity, technology, and pedagogy. As AI systems evolve from mere computational tools into co-creative partners, they have transformed how artistic processes are conceptualized, executed, and evaluated (Ansone et al., 2025; Miranda, 2025; Al-Kawmani, 2025). Within art and design education, GenAI now mediates the creative encounter between the learner and the digital medium, offering new affordances for experimentation, visualization, and ideation (Yang & Shin, 2025; Su & Mokmin, 2024).

Yet, while these technologies expand the creative horizon, they simultaneously challenge traditional pedagogical models that emphasize human authorship, material practice, and interpretive depth (Huh et al., 2025; Sáez-Velasco et al., 2024). Recent literature highlights an emerging paradigm shift in art education, where educators and students must negotiate between human-centered expression and algorithmically generated aesthetics (Liu & Zhu, 2025; Liu & Wang, 2025). On one hand, AI-driven tools enhance accessibility, feedback, and sustainability in artistic production (Su & Mokmin, 2024); on the other, they introduce ethical and evaluative complexities concerning originality, transparency, and the boundaries of creative agency (Zhang et al., 2025; Miranda, 2025).

Empirical evidence also points to the cognitive and affective dimensions of AI use in visual arts classrooms, revealing both the potential and the challenges of integrating AI imagery into creative pedagogy (Lin & Chen, 2024; Effects of AI-generated Images in Visual Art Education by Bian et al., 2025). These concerns have become particularly salient in higher-education art programs, where GenAI increasingly informs visual research, assessment, and cross-disciplinary collaboration (Fang, 2026). Consequently, there is a growing need for pedagogical frameworks and assessment models that reconcile technological innovation with artistic integrity, ensuring that AI integration supports rather than supplants the human creative process (Ilieva et al., 2023; Adamakis & Rachiotis, 2025; Lye & Lim, 2024).

The rapid integration of generative artificial intelligence (GenAI) into visual arts education has fundamentally transformed artistic production processes. Assessment frameworks, however, have not evolved at a comparable pace. Most evaluative systems in art education remain grounded in manual, material-based paradigms that prioritize individual authorship, technical execution, and tactile skill. As a result, conventional rubrics struggle to account for the hybrid, collaborative, and process-driven nature of AI-assisted artmaking, leading to persistent ambiguity in judgments of originality, authorship, and educational value (Messer, 2024; Al-Kawmani, 2025; Loivaranta et al., 2025). Recent studies acknowledge the affordances of GenAI for ideation, visual exploration, and expanded creative agency, while simultaneously documenting sustained skepticism among students, artists, and educators regarding fairness, transparency, and authenticity in the evaluation of AI-assisted artworks (Lovato et al., 2025). These concerns suggest that perceived assessment fairness may be influenced not only by creative outcomes, but also by the visibility, clarity, and structure of evaluative criteria. Nevertheless, existing research has largely approached these issues through survey-based or conceptual lenses, offering limited insight into how assessment design operates within formal educational settings.

Parallel efforts to automate artistic evaluation using artificial intelligence, such as the ArtMentor framework, have explored the use of multimodal large language models to assess visual artworks (Albarakat et al., 2024; Liu & Wang, 2025). While such systems demonstrate technical potential, they remain algorithmically oriented and detached from pedagogical contexts, overlooking affective, ethical, and aesthetic dimensions that are central to visual arts education. Consequently, these approaches fall short of addressing how assessment practices shape learners' perceptions of fairness, trust, and creative legitimacy.

Accordingly, the central research problem is the lack of pedagogically grounded, AI-responsive assessment models that can explain and address discrepancies in perceived fairness between manual and AI-assisted art practices. To address this gap, the present study proposes and empirically tests a blind, criterion-weighted rubric within a quasi-experimental design to evaluate both manual and AI-assisted artworks.

By recalibrating how creative learning outcomes are defined and assessed, this research seeks to contribute a more equitable, transparent, and pedagogically coherent evaluation framework for contemporary visual arts education. Therefore, this study aims to compare manual drawing and AI-assisted design approaches in a higher education colour design course by evaluating student artistic performance across four criteria, namely innovative idea, lines and artistic composition, colour expression, and visual processing, using a criterion-weighted rubric, and by examining student perceptions of fairness, satisfaction, and creative motivation through a validated questionnaire. Specifically, the study seeks to determine whether the method of production, manual or AI-assisted, influences assessed artistic outcomes and subjective learning experience, in order to inform more equitable and effective integration of AI tools within art and design education.

Research Questions

This study addressed the following research questions:

RQ1: Are there statistically significant differences in artistic performance outcomes (innovative idea, lines and artistic composition, colour expression, and visual processing) between students using manual drawing methods and those using AI-assisted design tools in a higher education colour design course?

RQ2: Are there statistically significant differences in students' perceptions of assessment equity between the manual drawing group and the AI-assisted design group?

RQ3: Are there statistically significant differences in students' satisfaction with the educational experience and creative motivation between the two production groups?

Research Significance

This study holds theoretical and practical significance across several dimensions relevant to art education, assessment design, and the integration of emerging technologies in higher education.

Theoretically, the study contributes to the growing body of literature examining how AI-assisted design tools intersect with creative learning outcomes, particularly in

studio-based disciplines where artistic performance has traditionally been evaluated through manual production methods. By empirically comparing the two production modalities within a controlled quasi-experimental framework, the study provides evidence-based insights into whether and how AI tools affect the dimensions of creative performance, including innovative ideation, compositional decision-making, and colour expression. This addresses a notable gap in the literature, where most existing scholarship either advocates for or critiques AI in education without rigorously measuring its impact on measurable artistic outcomes.

Practically, the findings carry direct implications for curriculum design and assessment policy in higher education art and design programmes. As AI-assisted design tools become increasingly accessible to students in Saudi universities and beyond, educators face pressing questions about how to evaluate student work equitably when production methods differ fundamentally. This study responds to that challenge by testing a criterion-weighted rubric under a blind assessment protocol and examining whether students themselves perceive the assessment process as fair. The results therefore offer actionable guidance to faculty, programme coordinators, and institutional policymakers seeking to update assessment frameworks to reflect the realities of digital and hybrid art production.

Contextually, the study is significant for its setting within a Saudi higher education institution, contributing to the localised evidence base needed to support Vision 2030's emphasis on creativity, innovation, and technology integration in education. Research situated within the Saudi art and design context remains relatively limited, and this study adds empirical depth to a field where culturally contextualised findings are essential for meaningful policy and pedagogical decisions.

Research Hypothesis

There are no statistically significant differences in artistic performance outcomes between students using manual drawing methods and those using AI-assisted design tools in a higher education colour design course.

There are no statistically significant differences in students› perceptions of assessment equity between the manual drawing group and the AI-assisted design

group.

There are no statistically significant differences in students› satisfaction with the educational experience and creative motivation between the manual drawing group and the AI-assisted design group.

Operational Definitions

Assessment Equity: Operationally defined in this study as students' self-reported perceptions that the evaluation criteria, scoring process, and outcomes were fair, transparent, and free from bias based on the method of artistic production (manual or AI-assisted), as measured by the Fairness subscale of the post-experiment questionnaire.

Creativity: Operationally defined as the student's capacity to generate original and innovative visual ideas within the given thematic framework, assessed through the Innovative Idea criterion of the weighted rubric (scored 0–6 points) and the Creativity and Motivation subscale of the post-experiment questionnaire.

Criterion-Based Assessment: Operationally defined as a structured evaluation approach using a weighted rubric with four pre-defined artistic criteria (Innovative Idea, Lines and Artistic Composition, Colour Expression, and Visual Processing) yielding a total maximum score of 30 points, applied uniformly to all artworks under a blind assessment protocol.

AI-Assisted Art Production: Operationally defined as a production process in which students used generative AI tools (via text prompts) to produce initial visual compositions, which were then printed in black-line format and completed manually with acrylic paint on canvas, as implemented in Group B of this study.

Research Limitations

Geographical Location:

The study was conducted exclusively at Umm Al-Qura University, located in Makkah, Saudi Arabia. As a single institutional context within one region of the Kingdom, the findings may not fully generalise to art and design programmes at other Saudi universities, nor to higher education institutions in different cultural, linguistic,

or socioeconomic contexts. Variations in infrastructure, access to AI tools, faculty readiness, and institutional culture across different universities mean that caution should be exercised when extrapolating the results beyond this specific setting.

Case Study Limitations

The study was confined to one course, the Colour Expression course, within a specific academic programme, and involved a sample of 40 female undergraduate students, equally divided between two groups. This relatively modest sample size, while sufficient for the statistical analyses employed, limits the statistical power available to detect small effect sizes across all outcome dimensions and reduces the robustness of subgroup comparisons. Furthermore, the all-female composition of the sample, which reflects the gender structure of the host institution, means that findings cannot be assumed to extend to mixed-gender or male-only educational environments. The 13-week timeframe of the intervention, while structured and consistent across both groups, represents a single semester and may not capture longer-term effects of sustained AI tool use on artistic development, creative identity, or assessment perception.

Literature Review

The Evolving Landscape of Creativity Assessment in Art Education

The question of how to assess creativity fairly has long occupied art educators, but the rapid integration of generative artificial intelligence (GenAI) into studio-based practice has rendered this challenge considerably more complex. Traditionally, creativity assessment in visual arts education has been grounded in rubric-based frameworks that reward demonstrable technical skill, compositional decision-making, and the originality of human expression (Onsman, 2016). These rubrics reflect an implicit assumption that artistic value is rooted in the labour, intention, and material engagement of an individual maker. However, as scholars have noted, creativity itself remains “an ill-defined and debated concept in the arts, design, architecture, and related engineering” disciplines, making any standardised evaluation inherently interpretive (Abdellatif & El-Wakeel, 2025). Rubrics provide a structured framework

for assessing dimensions such as ideation, originality, and visual communication, yet even these instruments remain vulnerable to subjective interpretation, especially when evaluators are confronted with work whose production process is unclear or algorithmically mediated (Khan et al., 2025).

The emergence of AI-assisted art production has intensified this ambiguity. Tools such as DALL-E, Midjourney, and Stable Diffusion enable students to generate sophisticated visual outputs with minimal manual dexterity, thereby decoupling the technical execution of an artwork from the conceptual and aesthetic intent underlying it. Studies examining how observers evaluate human-made versus AI-generated art consistently reveal a "source attribution bias": audiences and assessors tend to rate artworks lower on perceived skill, creativity, and monetary value when they believe the work was created by AI, even when the visual quality is objectively comparable (Horton et al., 2023). This cognitive bias presents a direct threat to assessment equity, as it suggests that grading outcomes in art education may be shaped less by the intrinsic quality of the work than by knowledge of its mode of production. Cunningham et al. (2025) further confirmed that explicit biases against AI-generated art are linked to implicit perceptual mechanisms, raising the possibility that even well-intentioned assessors cannot fully neutralise their evaluative prejudice toward algorithmically produced imagery.

Tensions Between Authenticity, Authorship, and Evaluative Equity

At the heart of the assessment equity debate lies a deeper philosophical tension regarding authorship and originality. In traditional art pedagogy, the value of a creative work is inseparable from the intentionality and effort of the individual artist; the creative process is considered as educationally significant as the product itself. The introduction of GenAI complicates this calculus by distributing creative agency between the human prompter and the generative model, producing what Zhang and Xu (2025) describe as a renegotiation of the creator's role, where AI functions as "a partner that stimulates ideation and expands creative boundaries" rather than a passive tool. While participants in their survey broadly acknowledged AI's capacity to enrich the creative process, they simultaneously expressed concerns about stylistic

homogenisation and the erosion of traditional authorship norms — concerns that directly impinge on how originality can be defined and assessed within an educational context.

These tensions are not merely academic. A large-scale survey of 516 professional artists conducted by Lovato et al. (2024) found that a majority believed AI outputs should not be attributed solely to model creators and that transparency in the use of AI during artistic production is an ethical imperative. This community-level demand for disclosure and transparency has direct implications for art education: if assessors are unaware of whether a submitted work was created manually or with AI assistance, the resulting grade may neither reflect the student's actual learning nor uphold principles of evaluative fairness. Peters and Angelov (2025) have similarly argued that the arrival of GenAI compels educators to redesign assessment tasks to ensure that student work reflects genuine creative engagement, free from what they describe as a new form of academic dishonesty that conventional plagiarism-detection systems are insufficient to address.

Adding further complexity, research suggests that GenAI tools, despite their impressive fluency in visual and textual production, exhibit a "fixation bias" comparable to human creators — tending to cluster ideas within conventional categories rather than generating truly original outputs (Desdevises, 2025). This finding has important pedagogical implications: if AI-assisted students are assessed on criteria such as innovative ideation or compositional novelty, they may be advantaged by the sheer volume of options the technology generates, even if the underlying originality of the prompting process is limited. Assessors who cannot distinguish the student's conceptual contribution from the system's generative output may therefore be making judgments of creative merit that are systematically misleading.

Assessment Design Responses: Rubrics, Blind Evaluation, and Equity Frameworks

Given these challenges, researchers and practitioners have increasingly called for new assessment paradigms that can accommodate the hybrid and collaborative nature of AI-assisted artmaking without privileging either mode of production unfairly. One prominent approach involves redesigning rubrics to shift evaluative

weight from technical execution toward process-oriented criteria such as conceptual development, critical reflection, and transparency of method (Fan, 2026). In this framework, students may be asked to submit process journals documenting their prompts, iterations, and reflective responses to AI-generated content, ensuring that the assessed artefact captures the student's intellectual engagement rather than merely their prompt-engineering skills.

Empirical work in this area suggests that rubric design significantly mediates students' perceptions of assessment fairness. A quasi-experimental study employing blind evaluation of AI-generated critiques in art history education found that rubric-based assessments in which artworks were anonymised and raters were blinded to group assignments produced more consistent and equitable outcomes, with AI-assisted learners achieving higher rubric scores on technical dimensions such as colour, composition, and brushstroke analysis (Abdulmajid et al., 2025). These findings echo those of Bian et al. (2025), whose controlled study in visual arts education used a blinded, criterion-weighted rubric to mitigate evaluator bias, finding that raters' assessments of technical and expressive dimensions of student paintings were more reliable when group identity was concealed during scoring.

A further strand of research highlights student perceptions of fairness as an independent variable worthy of empirical attention. When learners perceive an assessment system as transparent, equitable, and aligned with the actual demands of the task, their satisfaction and motivation tend to increase regardless of production modality (Abdellatif & El-Wakeel, 2025). It may therefore be helpful to introduce clearly weighted criteria that incorporate both process and product into the assessment process to serve a dual pedagogical function, reducing evaluator bias and enhancing students' sense that they are being assessed on what they have actually learned rather than on which tools they used. Taken together, the literature points toward a consensus that assessment equity in AI-assisted art production requires not simply updating existing rubrics but fundamentally reconceptualising what creative learning looks like in a technological environment where the boundaries of authorship, effort, and originality are constantly being renegotiated.

Methodology

Research Design

This study employed a quasi-experimental non-equivalent groups design with post-test measures to examine differences in artistic performance and student perceptions between manual and AI-assisted production methods. The independent variable was the method of artistic production (manual versus AI-assisted), while the dependent variables included criterion-based artistic evaluation scores and self-reported perception measures (fairness, satisfaction, and creative motivation). A quasi-experimental rather than true experimental design was adopted due to institutional constraints that precluded random assignment; group allocation was based on logistical considerations (laptop availability and voluntary consent) rather than performance-based criteria. It is acknowledged that the absence of a formal pre-test to verify baseline equivalence in artistic ability represents a methodological limitation; however, several compensatory controls were applied to strengthen internal validity, including shared academic cohort membership, identical syllabi, unified instructional conditions, and a blind assessment protocol.

Given institutional and logistical constraints, random assignment was not feasible. However, several methodological controls were implemented to strengthen internal validity. Both groups were drawn from the same academic cohort, followed an identical syllabus, operated under a unified thematic framework, and were supervised by the same instructor over the same 13-week instructional period. All participants completed equivalent submission requirements under comparable temporal and material conditions.

A post-test-only structure was adopted, and a blind assessment protocol was implemented to minimize evaluator bias. Expert evaluators independently assessed anonymized artworks without knowledge of the production method. Although baseline artistic ability was not formally measured through a pre-test, academic cohort equivalence and controlled instructional conditions were used to reduce potential selection bias. The study integrated rubric-based quantitative evaluation data with

post-experiment questionnaire responses to enable structured comparison of artistic outcomes and perceived assessment fairness. Figure 1 shows the quasi-experimental research design comparing manual and AI-assisted art production in the Colour Expression course (VA2504).

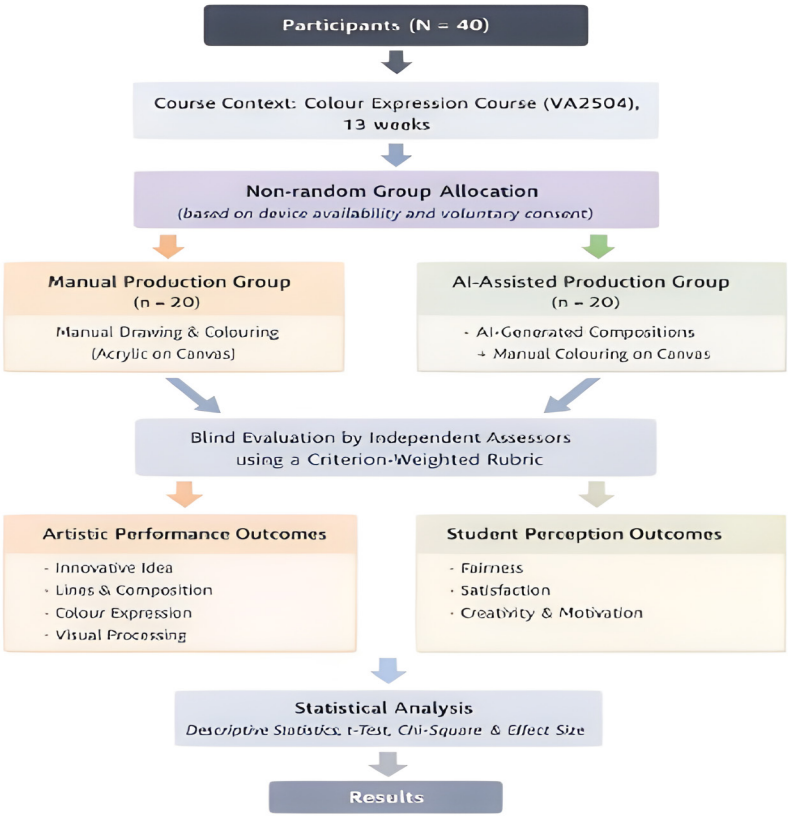

Figure 1. Quasi-experimental research design comparing manual and AI-assisted art production in the Colour Expression course (VA2504).

The study population consisted of female undergraduate students enrolled in the fourth-year Visual Arts/Art Education program at Umm Al-Qura University. Participants were primarily enrolled at Level 5, with a smaller proportion from Level 6 within the same academic program. A total of 40 students from a single existing course section (Section 1) participated in the study.

Group allocation followed the natural instructional structure of the course and was

not based on random assignment. Participants were divided into two equal groups (n= 20 per group) based on logistical considerations, specifically the availability of personal laptop devices and their voluntary consent to participate in the AI-assisted component of the study. Not all students agreed to participate in the AI-assisted condition due to limited or unstable internet access. Consequently, allocation criteria were technical rather than performance-based and were unrelated to prior artistic competence.

Both groups were enrolled in the same course (VA2504), followed an identical syllabus, received the same instructional content, and were supervised by the same instructor throughout the 13-week period. All participants belonged to the same academic program and had completed comparable prerequisite coursework in drawing and colour design. Examination of academic level distribution revealed no statistically significant differences between the two groups ($p > 0.05$), suggesting comparable academic standing.

To further reduce potential bias inherent in quasi-experimental non-randomized designs, all artworks were evaluated using a blind assessment protocol in which evaluators were unaware of the production method. Despite these controls, baseline artistic skill and prior academic performance were not formally measured, which represents a methodological limitation of the study.

Study Procedures

The study was conducted within the Colour Expression course (VA2504) over a 13-week instructional period. All participants engaged in a unified thematic framework titled "Saudi Life Glimpses – Saudi Art and Traditional Crafts." This standardised theme ensured conceptual comparability across both production conditions while allowing for individual creative interpretation.

Students in both groups operated under identical temporal constraints, submission requirements, and evaluation criteria to maintain procedural equity. The Manual Production Group (Group A) developed and executed their artworks entirely by hand, from conceptualization and compositional planning to final execution. Students were required to produce original designs and were restricted to the exclusive use of canvas and acrylic paints. Figure 2 shows an example of students' studio work, demonstrating their engagement with manual drawing and colouring.

Figure 2. Example of students' work in the studio: Manual Drawing and Colouring.

The AI-Assisted Production Group (Group B) generated initial visual compositions using generative AI tools through student-created text prompts. All prompts were documented to ensure transparency of the creative process. To preserve comparability of material execution, AI-generated outputs were printed in pencil-outline or black-line format only, without coloured printing. Students were then required to manually complete all colouring and refinement stages using acrylic paints on canvas. The use of mobile phones for AI generation was prohibited to maintain consistency in technological access conditions. Figure 3 shows an example of students' studio work, demonstrating their engagement with AI-assisted design and colour selection.

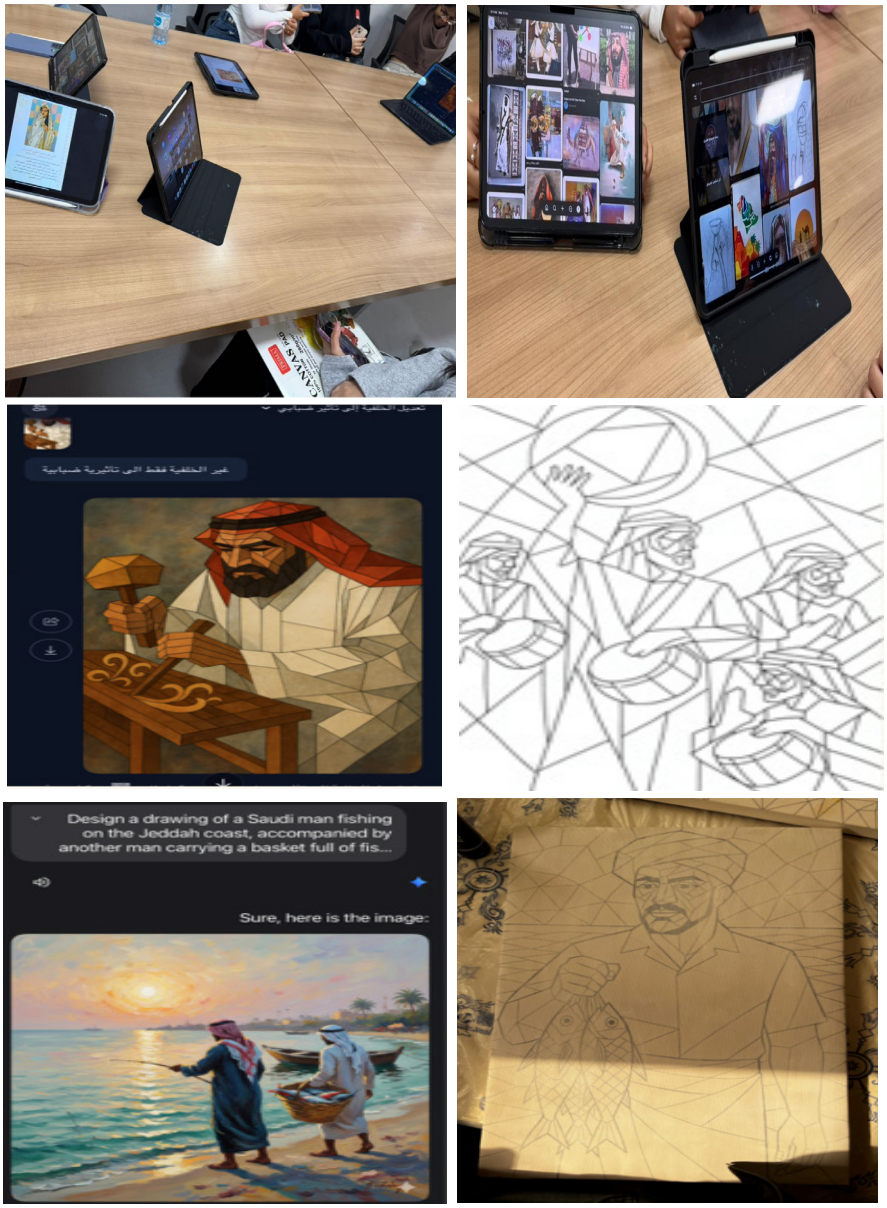

Figure 3. Example of students' work in the studio: AI-Assisted Design and Colour Selection.

Upon completion, all artworks were anonymized through numerical coding to prevent identification of the production method. Each student provided a brief conceptual explanation of her work to the evaluation committee without disclosing whether the piece was manually produced or AI-assisted.

An expert committee of art education specialists conducted independent evaluations using a standardised criterion-weighted rubric aligned with official assessment protocols. All artworks were assessed individually under blind conditions, with evaluators unaware of the production method employed. Scores were recorded directly on structured evaluation forms based on predetermined artistic criteria.

Assessment Instrument

Evaluation Rubric

A criterion-weighted assessment rubric was developed to evaluate student artworks across four core artistic domains: Concept and Innovation (maximum 6 points), Linework and Composition (maximum 6 points), Colour Expression (maximum 8 points), and Visual Treatment, Technical Execution, and Final Presentation (maximum 10 points), yielding a total maximum score of 30 points.

The weighting structure was designed to reflect the pedagogical priorities of the Colour Expression course, with greater emphasis placed on visual treatment and technical execution while preserving substantial evaluative weight for conceptual originality and compositional structure. Each criterion included clearly defined performance descriptors to guide consistent scoring.

Evaluators independently assigned scores within the predefined ranges for each domain. Individual criterion scores were subsequently aggregated to produce a total artwork score for comparative statistical analysis. The rubric was applied uniformly across both production conditions to ensure equitable assessment standards.

Validity of the Assessment Instrument

The content validity of the artwork evaluation rubric was established through expert review by six specialists in art education and visual arts. The experts evaluated the rubric criteria—Concept and Innovation, Linework and Composition, Colour Expression, and Visual Treatment and Technical Execution—across four dimensions:

clarity of wording, relevance to the study objectives, comprehensiveness of artistic domain coverage, and linguistic appropriateness for the target population. Feedback from the expert panel was systematically incorporated into the final version of the instrument.

Content validity was quantified using the Content Validity Index (CVI) based on a four-point relevance scale (1 = not relevant to 4 = highly relevant). All criteria achieved Item-Level CVI (I-CVI) values of 0.833 or higher, exceeding the recommended threshold of 0.78 for panels of six experts. The Scale-Level CVI calculated using the Average method (S-CVI/Ave) was 0.984, while the Universal Agreement index (S-CVI/UA) reached 0.905, indicating strong inter-expert agreement. These findings provide robust evidence of content validity for evaluating both AI-assisted and manually produced visual artworks.

Reliability Considerations

Formal inter-rater reliability coefficients were not calculated in the present study. However, multiple procedural safeguards were implemented to enhance scoring consistency. Evaluators applied a standardised criterion-weighted rubric containing clearly defined performance descriptors. All assessments were conducted independently under blind conditions, and evaluators adhered to established judging protocols to ensure consistent interpretation of criteria. While the absence of statistical reliability coefficients represents a methodological limitation, these structured evaluation procedures were designed to reduce subjectivity and minimize evaluator bias, thereby supporting a reasonable degree of scoring consistency across assessments.

Data Collection

Data collection took place immediately following the 13-week production period during the final evaluation session. All 40 artworks were displayed under standardised exhibition conditions and assessed using the previously described criterion-weighted rubric. Evaluation scores were recorded independently by committee members on structured assessment forms under blind conditions.

Subsequently, participants completed a post-experiment questionnaire designed

to measure their perceptions of assessment fairness, creative engagement, challenges encountered during the production process, and overall satisfaction with their respective instructional condition. Questionnaire responses were collected anonymously to ensure candid reporting.

Data Analysis

All data were coded and entered into IBM SPSS Statistics (Version 30). Descriptive statistics, including means, standard deviations, frequencies and percentages, were calculated to summarize the distribution of scores across both production groups.

Inferential analyses were conducted to examine differences between the manual and AI-assisted groups. The normality of score distributions was assessed using the Shapiro–Wilk test. For variables meeting normality assumptions, Independent Samples t-Tests were performed to compare mean scores across artistic criteria and total rubric scores. For variables that violated normality assumptions, the non-parametric Mann–Whitney U test was applied. Pearson's Chi-square test was used to examine associations between production method and categorical performance ratings. Effect sizes (Cohen's d) were calculated to estimate the magnitude of group differences. Statistical significance was established at $p < 0.05$.

Results

Socio-demographic Characteristics of the Participants

Table 1 presents the socio-demographic characteristics of the 40 student participants. The sample was equally divided by project type, with 20 students completing manual drawing and colouring and 20 completing AI-assisted design and colour selection. Regarding academic level, the majority were enrolled at Level 5 (57.5%), followed by Level 4 (25.0%), Level 7 (15.0%), and Level 6 (2.5%).

Table 1: Socio-demographic Characteristics of Participants (N = 40)

Variables	Frequency	Percentage (%)
Academic Level		
Level 4	10	25.0
Level 5	23	57.5
Level 6	1	2.5

Variables	Frequency	Percentage (%)
Level 7	6	15.0
Project Type		
Manual Drawing and Colouring	20	50.0
AI-Assisted Design and Colour Selection	20	50.0

N = total sample size.

Student Perceptions about the Educational Experience

Table 2 presents the distribution of Student Perceptions about the Educational Experience. Regarding fairness, the clarity of evaluation criteria was the highest-rated item (mean= 3.72; SD= 1.26), followed by fair recognition of personal effort in colour selection and execution (mean= 3.55; SD= 1.40), availability of tools and materials (mean= 3.43; SD= 1.28), overall fairness of the evaluation regardless of execution method (mean= 3.18; SD= 1.30), and perceived bias between manual and AI-assisted designs, which recorded the lowest mean in this axis (mean= 3.00; SD= 1.24), with the axis total mean reaching 3.38 (SD= 0.89). In terms of satisfaction with the educational experience, developing colour expression skills ranked highest (mean= 3.77; SD= 1.23), followed by adequate time allocation (mean= 3.65; SD= 1.19), receiving sufficient support during the experience (mean= 3.63; SD= 1.17), overall satisfaction with the experience (mean= 3.62; SD= 1.39), and the desire to repeat a similar experience (mean= 3.28; SD= 1.41). Satisfaction with the evaluation method applied to students' artistic work recorded the lowest score in this axis (mean= 3.10; SD= 1.58), with the axis total mean of 3.51 (SD= 1.17). Among the creativity and motivation, the discovery of new artistic expression methods was the top-rated item (mean= 3.75; SD= 1.19), followed by understanding the relationship between manual work and technology in the artistic field (mean= 3.68; SD= 1.29), motivation for creative thinking in design and colours (mean= 3.65; SD= 1.17), perceiving the relationship between technology and artistic imagination in a new way (mean= 3.65; SD= 1.29), and increased confidence in artistic abilities (mean= 3.55; SD= 1.30). The perception of sufficient freedom to express personal ideas recorded the lowest score in this axis (mean= 3.23; SD= 1.53), with the axis total mean of 3.58 (SD= 1.21). Across all three axes, every total mean exceeded the scale midpoint of 3.00, reflecting

generally positive student perceptions of the experience.

Table 2: Distribution of Student Responses Across Three Axes – Fairness, Satisfaction, and Creativity & Motivation (N= 40)

Items	Scale					Mean	SD
	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree		
	n (%)	n (%)	n (%)	n (%)	n (%)		
Fairness						3.38	0.89
1. Were the evaluation criteria clear and understandable for your design work?	3 (7.5)	4 (10.0)	8 (20.0)	11 (27.5)	14 (35.0)	3.72	1.26
2. To what extent did you feel that your personal effort in colour selection and execution was fairly valued in the evaluation?	5 (12.5)	4 (10.0)	9 (22.5)	8 (20.0)	14 (35.0)	3.55	1.40
3. Did you notice any preference or bias between manual designs and AI-assisted designs?	5 (12.5)	10 (25.0)	10 (25.0)	10 (25.0)	5 (12.5)	3.00	1.24
4. Did you encounter any shortage of tools, materials, or technical/ material difficulties that affected the fairness of the evaluation or your participation?	3 (7.5)	8 (20.0)	8 (20.0)	11 (27.5)	10 (25.0)	3.43	1.28
5. Did you feel that the overall evaluation of the experience was fair to all students, regardless of the method of execution?	6 (15.0)	5 (12.5)	12 (30.0)	10 (25.0)	7 (17.5)	3.18	1.30
Satisfaction with the Educational Experience						3.51	1.17

Items	Scale					Mean	SD
	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree		
	n (%)	n (%)	n (%)	n (%)	n (%)		
1. I am satisfied with the educational experience I participated in.	5 (12.5)	4 (10.0)	6 (15.0)	11 (27.5)	14 (35.0)	3.62	1.39
2. The experience helped me develop my colour expression skills.	3 (7.5)	4 (10.0)	5 (12.5)	15 (37.5)	13 (32.5)	3.77	1.23
3. The available time was sufficient to complete the work.	3 (7.5)	4 (10.0)	7 (17.5)	16 (40.0)	10 (25.0)	3.65	1.19
4. I received adequate support during the experience.	3 (7.5)	2 (5.0)	13 (32.5)	11 (27.5)	11 (27.5)	3.63	1.17
5. I would like to repeat a similar experience in other courses.	8 (20.0)	2 (5.0)	10 (25.0)	11 (27.5)	9 (22.5)	3.28	1.41
6. I feel satisfied with the way my artistic work was evaluated.	11 (27.5)	3 (7.5)	8 (20.0)	7 (17.5)	11 (27.5)	3.10	1.58
Creativity and Motivation						3.58	1.21
1. The experience motivated me to think creatively in design and colours.	2 (5.0)	6 (15.0)	6 (15.0)	16 (40.0)	10 (25.0)	3.65	1.17
2. The experience helped me discover new methods of artistic expression.	2 (5.0)	6 (15.0)	4 (10.0)	16 (40.0)	12 (30.0)	3.75	1.19
3. The experience increased my confidence in my artistic abilities in design and colouring.	3 (7.5)	7 (17.5)	7 (17.5)	11 (27.5)	12 (30.0)	3.55	1.30

Items	Scale					Mean	SD
	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree		
	n (%)	n (%)	n (%)	n (%)	n (%)		
4. The experience helped me understand the relationship between manual work and technology in the artistic field.	3 (7.5)	6 (15.0)	5 (12.5)	13 (32.5)	13 (32.5)	3.68	1.29
5. The experience made me see the relationship between technology and artistic imagination in a new way.	3 (7.5)	6 (15.0)	6 (15.0)	12 (30.0)	13 (32.5)	3.65	1.29
6. The experience gave me sufficient freedom to express my own ideas.	8 (20.0)	6 (15.0)	7 (17.5)	7 (17.5)	12 (30.0)	3.23	1.53

SD= Standard Deviation.

Comparison of Domain Scores about the Educational Experience Between Manual and AI-Assisted Groups

Table 3 presents the results of Independent Samples t-Tests comparing the mean scores of the manual drawing group and the AI-assisted design group across the three study axes. No statistically significant differences were found between the two groups on any of the three axes ($p > 0.05$). The AI-assisted group scored marginally higher on all three axes; however, the effect sizes (Cohen's d) were small, ranging from -0.178 to -0.303, indicating minimal practical differences between the groups.

Table 3: Comparison of Domain Scores Between Manual and AI-Assisted Project Groups (N= 40)

Domain	Manual Drawing Mean (SD)	AI-Assisted Design Mean (SD)	t	df	P-value	Cohen's d
Fairness	3.24 (0.996)	3.51 (0.775)	-0.957	38	0.172	-0.303
Satisfaction	3.33 (1.192)	3.68 (1.157)	-0.942	38	0.176	-0.298
Creativity & Motivation	3.48 (1.208)	3.69 (1.226)	-0.563	38	0.288	-0.178

SD= Standard Deviation. p-values based on equal variances assumed (Levene's test non-significant for all domains).

Effect size interpretation: small < 0.5, medium 0.5–0.8, large > 0.8.

Association between Project Type and Artistic Evaluation Criteria

Table 4 presents the Chi-square test results examining the association between project type (manual vs. AI-assisted) and student performance ratings across four artistic evaluation criteria by the evaluators. Figure 4 presents examples of student artwork from the manual drawing group, while Figure 5 presents examples from the AI-assisted design group. Regarding the innovative idea criterion, the results showed a statistically significant association with project type ($\chi^2 = 8.908$, $p = 0.012$). Among students rated as Good, 3 (42.9%) were from the manual drawing group and 4 (57.1%) were from the AI-assisted group. Among those rated Very Good, the majority were from the manual drawing group, with 13 students (76.5%) compared to only 4 (23.5%) from the AI-assisted group. In contrast, among students rated Excellent, 12 (75.0%) were from the AI-assisted group and only 4 (25.0%) were from the manual drawing group, indicating that AI-assisted students were notably more likely to achieve the highest performance rating on this criterion.

Concerning the Lines and Artistic Composition criterion, no statistically significant association was found ($\chi^2 = 5.016$, $p = 0.081$). Among students rated Good, 2 (66.7%) were from the manual drawing group and 1 (33.3%) from the AI-assisted group.

Among those rated Very Good, 13 (65.0%) were from the manual drawing group and 7 (35.0%) from the AI-assisted group. Among students rated Excellent, however, the distribution reversed, with 12 (70.6%) from the AI-assisted group compared to 5 (29.4%) from the manual drawing group, suggesting a directional trend that did not reach statistical significance.

With respect to the Colour Expression criterion, the association with project type was also non-significant ($\chi^2 = 5.712$, $p = 0.057$), though it approached the conventional threshold. Among students rated Good, 3 (42.9%) were from the manual group and 4 (57.1%) from the AI-assisted group. Among those rated Very Good, the manual drawing group accounted for 10 students (76.9%) compared to 3 (23.1%) from the AI-assisted group. Among students rated Excellent, 13 (65.0%) were from the AI-assisted group and 7 (35.0%) from the manual drawing group, mirroring the pattern observed in the Innovative Idea criterion.

Regarding the Visual Processing criterion, no significant association was identified ($\chi^2 = 2.565$, $p = 0.464$). One student (100.0%) in the Average category was from the manual drawing group. Among students rated Good, 5 (45.5%) were from the manual group and 6 (54.5%) from the AI-assisted group. Among those rated Very Good, 6 (66.7%) were from the manual drawing group and 3 (33.3%) from the AI-assisted group. Among students rated Excellent, 11 (57.9%) were from the AI-assisted group and 8 (42.1%) from the manual drawing group, reflecting a relatively balanced distribution across both groups with no meaningful pattern.

Table 4: Association between Project Type and Artistic Evaluation Criteria (N = 40)

Criterion	Rating	Manual Drawing	AI-Assisted Design	Total	χ^2	p-value
		n (%)	n (%)	n (%)		
Innovative Idea Criterion	Good	3 (42.9)	4 (57.1)	7 (100.0)	8.908	0.012*
	Very Good	13 (76.5)	4 (23.5)	17 (100.0)		
	Excellent	4 (25.0)	12 (75.0)	16 (100.0)		

Criterion	Rating	Manual Drawing	AI-Assisted Design	Total	χ^2	p-value
		n (%)	n (%)	n (%)		
Lines and Artistic Composition	Good	2 (66.7)	1 (33.3)	3 (100.0)	5.016	0.081
	Very Good	13 (65.0)	7 (35.0)	20 (100.0)		
	Excellent	5 (29.4)	12 (70.6)	17 (100.0)		
Colour Expression	Good	3 (42.9)	4 (57.1)	7 (100.0)	5.712	0.057
	Very Good	10 (76.9)	3 (23.1)	13 (100.0)		
	Excellent	7 (35.0)	13 (65.0)	20 (100.0)		
Visual Processing Criterion	Average	1 (100.0)	0 (0.0)	1 (100.0)	2.565	0.464
	Good	5 (45.5)	6 (54.5)	11 (100.0)		
	Very Good	6 (66.7)	3 (33.3)	9 (100.0)		
	Excellent	8 (42.1)	11 (57.9)	19 (100.0)		

* Significant at $p < 0.05$. χ^2 = Chi-square statistic.




Figure 4. Examples of student artwork produced using Manual Drawing and Colouring. The coloring was applied manually.

Figure 5. Examples of student artwork produced using AI-Assisted Design and Colour Selection. The coloring was applied manually.

Comparison of Artistic Criterion Scores Between Groups

Table 5 presents the Independent Samples t-Test results comparing the two groups on four graded artistic criteria. The AI-assisted group consistently achieved higher mean scores across all criteria. Regarding the Innovative Idea criterion, the AI-assisted group recorded a higher mean (mean= 5.25; SD= 0.85) compared to the manual drawing group (mean= 4.80; SD= 0.83), and this difference reached statistical significance (t = -1.690, p = 0.050), with a medium effect size (Cohen's d= -0.534), representing the largest effect size observed in this analysis. Concerning the Lines and Artistic Composition criterion, the AI-assisted group also scored higher (mean= 5.15; SD= 0.88) than the manual drawing group (mean= 4.90; SD= 0.79), though the difference was not statistically significant (t = -0.949, p = 0.174), with a small effect size (Cohen's d= -0.300). With respect to the Colour Expression criterion, the AI-assisted group recorded a mean of 6.60 (SD= 1.70) compared to 5.95 (SD= 1.40) for the manual drawing group, and the difference also did not reach statistical significance (t = -1.323, p = 0.097), with a small to moderate effect size (Cohen's d= -0.418). Regarding the Visual Processing criterion, the AI-assisted group scored a mean of 7.80 (SD= 1.79) compared to 7.50 (SD= 1.85) for the manual drawing group, with no statistically significant difference (t = -0.521, p = 0.303) and the smallest effect size among all criteria (Cohen's d= -0.165). Overall, these findings suggest that AI-assisted tools offer a statistically significant advantage in ideation and creative originality, as evidenced by the Innovative Idea criterion, though this advantage did not extend to the remaining technical execution criteria.

Table 5: Comparison of Artistic Evaluation Criterion Scores Between Manual and AI-Assisted Groups (N= 40)

Artistic Criterion	Manual Drawing	AI-Assisted Design	t	df	p-value	Cohen's d
	Mean (SD)	Mean (SD)				
Innovative Idea	4.80 (0.834)	5.25 (0.851)	-1.690	38	0.050	-0.534

Artistic Criterion	Manual Drawing	AI-Assisted Design	t	df	p-value	Cohen's d
	Mean (SD)	Mean (SD)				
Lines & Artistic Composition	4.90 (0.788)	5.15 (0.875)	-0.949	38	0.174	-0.300
Colour Expression	5.95 (1.395)	6.60 (1.698)	-1.323	38	0.097	-0.418
Visual Processing	7.50 (1.850)	7.80 (1.795)	-0.521	38	0.303	-0.165

SD= Standard Deviation. p-values are two-sided.

Effect size interpretation: small < 0.5, medium 0.5–0.8, large > 0.8.

Challenges Faced by Students

Several practical challenges were encountered by students across both groups during the experience. Students in the manual drawing group reported the absence of a dedicated studio or workshop space as a primary difficulty, as the lack of appropriate facilities such as easels, water access for brush cleaning, and suitable working surfaces affected the quality and authenticity of their work. Consequently, many students were unable to complete their projects within the classroom setting and were required to finish their work at home, which reduced the integrity and consistency of the production environment. Students in the AI-assisted design group faced a different set of challenges, most notably the unavailability of reliable internet access within the classroom, which forced them to rely on personal mobile data. Additionally, many students lacked professional familiarity with AI design tools, and the financial cost of premium subscriptions meant that students were largely restricted to free versions of available tools, which offered limited functionality and were unable to execute all required design commands, thereby reducing the overall quality of some outputs.

Discussion

The findings of this study indicate that students generally perceived the assessment process as moderately fair, with the overall mean for Axis One reaching 3.38 (SD= 0.89). The highest-rated item concerned the clarity of evaluation criteria (M= 3.72,

SD= 1.26), suggesting that most students found the rubric relatively transparent. However, the item addressing perceived bias between manual and AI-assisted work yielded a notably neutral mean (M= 3.00, SD= 1.24), with responses distributed evenly across all five response options. This pattern of equivocal responses closely mirrors findings reported by Sarkar and Chakrabarti (2011), who identified subjectivity and inconsistency as persistent challenges in the assessment of student design work, particularly when evaluators must compare outputs produced through fundamentally different processes. The tension observed in the present study, between students who viewed the dual-method assessment as equitable and those who did not, is consistent with broader concerns raised in the literature about the opacity of AI-related grading decisions. As documented by Adadi and Berrada (2018), algorithmic opacity can undermine student trust even in settings where the actual evaluation is conducted by human assessors. The relatively low mean for overall fairness (M= 3.18, SD= 1.30) further suggests that while the design of the experience was perceived as generally balanced, a non-trivial proportion of students questioned whether their execution method placed them at a disadvantage.

Student satisfaction with the educational experience was moderate to positive, with a total axis mean of 3.51 (SD= 1.17). The highest satisfaction was reported for skill development in colour expression (M= 3.77, SD= 1.23), indicating that regardless of project type, students perceived genuine learning value in the activity. This result aligns with the findings of Lu (2025), who, in a comparative study of 77 undergraduate product design students in China, found that integrating AI tools into design courses significantly enhanced students' learning outcomes and creativity ratings compared to a control group working without AI assistance. Similarly, the willingness to repeat the experience in other courses (M= 3.28, SD= 1.41) reflects moderate enthusiasm, which is consistent with patterns reported by Balabdaoui et al. (2024) in a large-scale survey of over 4,798 university students, where the majority expressed a desire for AI to be more systematically integrated into their coursework, while a notable minority maintained reservations. The most polarised item in this axis concerned satisfaction with the evaluation method (M= 3.10, SD= 1.58), suggesting

that the assessment approach, despite its structured rubric, generated divided views. This finding resonates with observations by Bui and Barrot (2024), who found that students frequently question the appropriateness of assessment tools when subjective artistic qualities are involved, particularly when human and AI-produced work must be judged against a shared standard.

The creativity and motivation axis recorded the highest overall mean among the three domains (M= 3.58, SD= 1.21), suggesting that the dual-method design experience was perceived as broadly stimulating. The item concerning the discovery of new artistic methods received the highest mean (M= 3.75, SD= 1.19), while the item addressing freedom of expression scored the lowest (M= 3.23, SD= 1.53). The stronger scores for creative discovery are consistent with findings from Abdellatif and El-Wakeel (2024), who implemented an AI-assisted design system at an architecture school and reported that generative AI tools were most effective in the concept creation stage of the design process, precisely where exposure to unfamiliar visual outputs can stimulate students' ideational breadth. The relatively constrained scores for freedom of expression, however, point to a limitation that has been widely noted in the literature. Redefining Creative Education, a case-study analysis published by the Journal of Research in Innovative Teaching and Learning, reported that students working with AI-assisted design tools sometimes experienced an overabundance of generated options that complicated rather than facilitated their creative choices, and raised concerns that AI outputs could homogenise artistic style and erode the distinctiveness of individual student work (Alam, 2021). The present study's data suggest a similar dynamic: while AI tools expanded students' technical and conceptual repertoire, some students felt that the constraints of the AI-assisted workflow limited their personal artistic voice.

The Independent Samples t-Tests revealed a statistically significant difference between the manual drawing group and the AI-assisted design group on the Innovative Idea criterion ($t = -1.690$, $p = 0.050$, Cohen's $d = -0.534$), with the AI-assisted group achieving a notably higher mean score (M= 5.25, SD= 0.85) compared to the manual drawing group (M= 4.80, SD= 0.83). The effect size for this criterion was medium,

indicating that the advantage held by the AI-assisted group was not only statistically detectable but also of practical educational significance. This finding is consistent with what Lu (2025) reported in a comparable between-group design study, where AI tools provided the most pronounced gains in creativity-related dimensions rather than in technical execution. The Chi-square analysis reinforced this conclusion: the only significant association between project type and evaluation rating was observed for the Innovative Idea criterion ($\chi^2 = 8.908$, $p = 0.012$), where 75.0% of students rated as Excellent came from the AI-assisted group, while 76.5% of those rated Very Good came from the manual group. This divergence suggests that AI tools elevate the ceiling of creative output for some students while the majority of competent but non-exceptional performers remain in the manual group, a dynamic noted by Urban et al. (2024, who observed that students using generative AI tools produced more creative solutions to ill-defined problems, though not uniformly across ability levels. The remaining three criteria, namely Lines and Artistic Composition ($t = -0.949$, $p = 0.174$), Colour Expression ($t = -1.323$, $p = 0.097$), and Visual Processing ($t = -0.521$, $p = 0.303$), showed no statistically significant differences between the two groups, reinforcing the view that AI assistance primarily influences ideation and creative originality rather than technical craftsmanship.

Taken together, the findings of this study support the view that integrating AI-assisted design tools alongside traditional manual methods in art and design education produces broadly comparable educational outcomes in terms of student fairness perceptions, satisfaction, and creative motivation, while offering a statistically significant and practically meaningful advantage specifically in idea generation and innovative thinking. These conclusions are partially consistent with the position advanced by Sáez-Velasco et al. (2024), who, reviewing AI integration across multiple creative disciplines, argued that AI tools are most effectively positioned as complementary instruments that extend rather than replace traditional artistic skills. The significant finding on the Innovative Idea criterion, however, suggests that AI tools may do more than merely complement manual skills in some contexts, and that their integration into assessment frameworks warrants careful consideration to ensure equitable evaluation

standards. The absence of significant between-group differences on the remaining criteria also speaks to an important equity concern: students who lacked access to AI tools or were less familiar with them were not demonstrably disadvantaged on technical execution outcomes, which echoes findings from Kortemeyer et al. (2023) regarding equitable AI access in online educational environments. Nevertheless, several limitations temper these conclusions. The sample size of 40 students limits statistical power and generalisability, and the study was conducted within a single course context. Furthermore, the high standard deviation observed for the satisfaction with evaluation item (SD= 1.58) signals that the heterogeneity of student responses on assessment fairness warrants further qualitative investigation. Future research should examine larger and more diverse samples, incorporate longitudinal designs to track creative development, and explore whether specific AI tools differ in their pedagogical effectiveness across different artistic disciplines. It is also worth noting that the blind assessment protocol, while designed to ensure evaluator neutrality, may have inadvertently favoured AI-assisted works in some criteria, as evaluators assessed outputs solely as final visual products without knowledge of the production method. The marginally higher scores awarded to AI-assisted works on certain criteria may therefore reflect the visual polish and aesthetic consistency that AI tools tend to produce, rather than a genuine difference in student creative ability, which represents an inherent limitation of product-focused assessment in contexts where the learning process is equally valued. These findings suggest that AI tools can be strategically integrated into art education at the ideation stage, where they can effectively enhance creative thinking. Importantly, this integration does not appear to compromise the development of students' technical artistic skills. Instead, it enables educators to utilize AI as a supportive tool that expands conceptual exploration while maintaining the importance of manual practice and technical proficiency.

Conclusion

This study examined student perceptions of fairness, satisfaction, and creative motivation in a colour design experience integrating both manual drawing and AI-assisted design approaches. Overall, students held positive perceptions across all

three axes, with means exceeding the scale midpoint of 3.00, and the experience was most positively received in relation to creativity and motivation. The comparative analysis revealed no statistically significant differences between the two groups in terms of fairness, satisfaction, or motivation perceptions, indicating that neither approach produced a substantially superior educational experience. However, a statistically significant advantage in favour of the AI-assisted group was identified on the Innovative Idea criterion ($t = -1.690$, $p = 0.050$, Cohen›s $d = -0.534$), supported by the Chi-square analysis ($\chi^2 = 8.908$, $p = 0.012$), suggesting that AI tools hold particular promise in supporting creative ideation, even when they do not confer measurable advantages in technical execution criteria. These findings support a complementary approach in which AI-assisted methods are introduced alongside traditional manual skills, while highlighting the need for clearly communicated evaluation criteria that account for the different creative processes involved in each approach. Future studies with larger and more diverse samples are needed to confirm these findings and to better understand the long-term impact of AI integration on student creativity and perceptions of educational equity.

Recommendations for Future Practice

Based on student responses, several recommendations emerge for improving colour design and AI-assisted learning experiences in future courses. The most frequently cited suggestions centred on granting students greater freedom in choosing their project topics, artistic styles, and methods of expression, rather than prescribing a fixed artistic school or approach. Students also emphasised the importance of communicating evaluation criteria clearly before the project begins, so that expectations are understood from the outset. Additional recommendations included providing dedicated laboratory or studio facilities for the course to avoid reliance on home-based working, extending the time allocated to projects to allow for adequate revision and refinement, incorporating the teaching of artistic techniques and AI tools during contact hours rather than expecting students to acquire these independently, and ensuring that assessment is conducted by the course instructor rather than external evaluators, in order to maintain consistency and contextual understanding of each

student›s work.

Acknowledgements

The authors would like to thank the participants and evaluators for making this research possible.

Ethical Approval

Ethical approval was obtained from Umm Al-Qura University, Saudi Arabia.

References

- Abdellatif, R., & El-Wakeel, H. (2025). Assessing creative outcomes in studio-based learning: A comparative assessment of analytical rubrics. International Journal of Design Creativity and Innovation, 13(1), 4166-. https://doi.org/10.1080/21650349.2024.2426645/
- Abdulmajid, M., Alrashaidi, E., Alayar, M., & Alrashidi, N. (2025). AI as a pedagogical partner: enhancing art history education and quality assurance through AI-generated critiques. Quality Assurance in Education, 118-. https://doi.org/10.1108/QAE-030074-2025-
- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access, 6, 5213852160-. https://doi.org/10.1109/ACCESS.2018.2870052
- Adamakis, M., & Rachiotis, T. (2025). Artificial intelligence in higher education: A state-of-the-art overview of pedagogical integrity, artificial intelligence literacy, and policy integration. Encyclopedia, 5(4), 180. https://doi.org/10.3390/encyclopedia5040180
- Alam, A. (2021, November). Possibilities and apprehensions in the landscape of artificial intelligence in education. In 2021 International conference on computational intelligence and computing applications (ICCICA) (pp. 18-). Ieee. https://doi.org/10.1109/ICCICA52458.2021.9697272
- Albarakat, M. S., Almuzaini, S. S., & Alshebeili, M. S. (2024). The role of artificial intelligence in designing calligraphy artistic works inspired by Abdulhalim Radwi to promote Arab-Islamic identity. Saudi Art and Design Journal, 4(2). https://doi.org/10.57194009-002-004-2351/

- Al-Kawmani, W. A. (2025). Inspiration of Van Gogh's style to create artworks using artificial intelligence. Saudi Art and Design Journal, 5(1), 1–20. https://doi.org/10.57194001-001-005-2351/
- Ansone, A., Zālīte-Supe, Z., & Daniela, L. (2025). Generative artificial intelligence as a catalyst for change in higher education art study programs. Computers, 14(4), 154. https://doi.org/10.3390/computers14040154
- Balabdaoui, F., Dittmann-Domenichini, N., Grosse, H., Schlienger, C., & Kortemeyer, G. (2024). A survey on students' use of AI at a technical university. Discover Education, 3(1), 51. https://doi.org/10.1007/s44217-024-4-00136
- Bian, C., Wang, X., Huang, Y., Zhou, S., & Lu, W. (2025). Effects of AI-generated images in visual art education on students' classroom engagement, self-efficacy and cognitive load. Humanities and Social Sciences Communications, 12(1), 114-. https://doi.org/10.1038/s415992-05860-025-
- Bui, N. M., & Barrot, J. S. (2025). ChatGPT as an automated essay scoring tool in the writing classrooms: how it compares with human scoring. Education and Information Technologies, 30(2), 20412058-. https://doi.org/10.1007/s10639-12891--024w
- Cunningham, C. V., Radvansky, G. A., & Brockmole, J. R. (2025). Human creativity versus artificial intelligence: Source attribution, observer attitudes, and eye movements while viewing visual art. Frontiers in Psychology, 16. https://doi.org/10.3389/fpsyg.2025.1509974
- Desdevises, J. (2025). The paradox of creativity in generative AI: High performance, human-like bias, and limited differential evaluation. Frontiers in Psychology, 16. https://doi.org/10.3389/fpsyg.2025.1628486
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). Opinion Paper:"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International journal of information management, 71, 102642. https://doi.org/10.1016/j.

ijinfomgt.2023.102642

- Fang, Z. (2026). Integrating generative AI in higher art education: A systematic review of tools, pedagogies, and practices. SN Computer Science, 7(2), 215. https://doi.org/10.1007/s4297904788--026-x
- Hartnett, M., Brown, C., Forbes, D., Gedera, D., & Datt, A. (2023). Enhanced or diminished attitudes: University students' agency. Computers & Education, 198, 104773. https://doi.org/10.1016/j.compedu.2023.104773
- Horton, C. B., White, M. W., & Iyengar, S. S. (2023). Bias against AI art can enhance perceptions of human creativity. Scientific Reports, 13, Article 18906. https://doi.org/10.1038/s415983-45202-023-
- Huh, M. B., Miri, M., & Tracy, T. (2025). Students' perceptions of generative AI image tools in design education: insights from architectural education. Education Sciences, 15(9), 1160. https://doi.org/10.3390/educsci15091160
- Ilieva, G., Yankova, T., Klisarova-Belcheva, S., Dimitrov, A., Bratkov, M., & Angelov, D. (2023). Effects of generative chatbots in higher education. Information, 14(9), 492. https://doi.org/10.3390/info14090492
- Jiang, R., & Chen, C. W. (2025, October). Multimodal llms can reason about aesthetics in zero-shot. In Proceedings of the 33rd ACM International Conference on Multimedia (pp. 66346643-). https://doi.org/10.11453746027.3754961/
- Jiang, Y., Fan, Y., & Liu, Z. (2025). Generative AI in Art Education: A Systematic Review of Research Trends, Tool Applications, and Outcomes (2019–2025). Education Sciences, 16(1), 47. https://doi.org/10.3390/educsci16010047
- Khan, S., Liow, W. T., & Ang, L. K. (2025, December). Human or AI? Comparing Design Thinking Assessments by Teaching Assistants and Bots. In 2025 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE) (pp. 18-). IEEE. https://arxiv.org/abs/2510.16069
- Lin, H., & Chen, Q. (2024). Artificial intelligence (AI)-integrated educational applications and college students' creativity and academic emotions: students and teachers' perceptions and attitudes. BMC psychology, 12(1), 487. https://doi.org/10.1186/s403590-01979-024-

- Liu, W., & Wang, Z. (2025). Evolution and Value of Arts Education Technology: A Comparative Analysis of Information Technology, Digital Technology, and Artificial Intelligence. Sage Open, 15(4), 21582440251395477. https://doi.org/10.1177215824402513954 77/
- Liu, Y., & Zhu, C. (2025). The use of deep learning and artificial intelligence-based digital technologies in art education. Scientific Reports, 15, 15859. https://doi.org/10.1038/s415989-00892-025-
- Loivaranta, T., Hautala, J., & Lundman, R. (2025). Spectrum of creative agencies in AI-based art: analysis of art reviews. Digital Creativity, 36(3), 223237-. https://doi.org/10.108014626268.2025.2491471/
- Lovato, J., Zimmerman, J. W., Smith, I., Dodds, P. S., & Karson, J. L. (2024). Foregrounding artist opinions: A survey study on transparency, ownership, and fairness in AI generative art. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 7, 905–916. https://doi.org/10.1609/aies.v7i1.31691
- Lovato, P., Ciston, S., & colleagues. (2024). Foregrounding artist opinions: A survey study on transparency, ownership, and fairness in AI generative art. ACM CHI Conference on Human Factors in Computing Systems. https://arxiv.org/abs/2401.15497
- Lu, W., Lalli, G. S., & Jiang, G. (2025). A comparative study on the influence of artificial intelligence (AI) on student creativity, learning outcomes and attitudes in product design education (PDE). European Journal of Education, 60(4), e70233. https://doi.org/10.1111/ejed.70233
- Lye, C. Y., & Lim, L. (2024). Generative artificial intelligence in tertiary education: Assessment redesign principles and considerations. Education Sciences, 14(6), 569. https://doi.org/10.3390/educsci14060569
- Messer, U. (2024). Co-creating art with generative artificial intelligence: Implications for artworks and artists. Computers in Human Behavior: Artificial Humans, 2(1), 100056. https://doi.org/10.1016/j.chbah.2024.100056
- Miranda, A. G. (2025). Educación, creating artística and artificial intelligence: una aproximación a los retos y oportunidades.. Tercio Creciente, (28), 715-. https://

orcid.org/00091413-9992-0000-

- Onsman, A. (2016). Assessing creativity in a 'New Generation'architecture degree. Thinking Skills and Creativity, 19, 210218-. https://doi.org/10.1016/j.tsc.2015.07.001
- Peters, M., & Angelov, D. (2025). Redefining assessment tasks to promote students' creativity and integrity in the age of generative artificial intelligence. International Journal for Educational Integrity. https://doi.org/10.1007/s4097900201--025-x
- Sáez-Velasco, S., Alaguero-Rodríguez, M., Delgado-Benito, V., & Rodríguez-Cano, S. (2024). Analysing the impact of generative AI in arts education: A cross-disciplinary perspective of educators and students in higher education. Informatics, 11(2), 37. https://doi.org/10.3390/informatics11020037
- Sarkar, P., & Chakrabarti, A. (2011). Assessing design creativity. Design studies, 32(4), 348383-. https://doi.org/10.1016/j.destud.2011.01.002
- Su, H., & Mokmin, N. A. M. (2024). Unveiling the canvas: Sustainable integration of AI in visual art education. Sustainability, 16(17), 7849. https://doi.org/10.3390/su16177849
- Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. Computers & Education, 215, 105031. https://doi.org/10.1016/j.compedu.2024.105031
- Yang, Z., & Shin, J. (2025). The impact of Gen AI on art and design program education. The Design Journal, 28(2), 310326-. https://doi.org/10.10801460692/5.2024.2425084
- Zhang, C., & Xu, S. (2025). Aesthetic experience and educational value in co-creating art with generative AI: Evidence from a survey of young learners. arXiv preprint. https://arxiv.org/abs/2509.10576
- Zhang, X., Yue, W. S., & Lee, K. C. S. (2026). Innovations and Challenges of AI Technology in Visual Art Education: A Systematic Review Based on General Systems Theory (2019–2024). Foro de Educacion, 24(1). https://doi.org/10.14201/fde.24110