КӨРКЕМӨНЕРДЕН БІЛІМ БЕРУ
ХУДОЖЕСТВЕННОЕ ОБРАЗОВАНИЕ
ART EDUCATION

IRSTI: **14.35.09** http://doi.org/10.51889/3005-6381.2024.85.4.001

Urazbay K.[1], Alzhanov G.[2], Çağatay A.[3]
[1]1-year doctoral student of the Department of Vocational Training and Fine Arts, O.Zhanibekov South Kazakhstan Pedagogical University, Shymkent, Kazakhstan, e-mail: oraz_74@mail.ru
[2]Kh.Dosmukhamedov Atyrau University, PhD, Atyrau, Kazakhstan, e-mail: alzhanov82@gmail.com
[3]Gazi Eğitim Fakültesi, Güzel Sanatlar Eğitimi, Resim-İş Eğitimi. Ankara, Türkiye, e-mail: cagatayakengin@gazi.edu.tr

DIGITAL TOOLS IN STUDENT ART RESEARCH: METHODOLOGY, RISKS, AND PEDAGOGICAL EFFECTIVENESS

Abstract

Digital tools and generative artificial intelligence are becoming more common as part of student art research, for source searches, visual idea prototyping, text drafting and creative inquiry documentation. At the same time, fast production can also hide weak verification practices, as well as pervious authorship limits and frail methodological traceability. This paper investigates how a well-organized digital research toolkit and explicit governance rules impact methodological quality, risk levels and pedagogical effectiveness in practice-based, including practice-led student research in visual arts and art education.

The study will employ a mixed-method and quasi-experimental design of two equivalent groups. Quantitative results are assessed using an analytical rubric targeting research traceability (quality of sources and accuracy of citation, transparency in methodology, alignment between evidence and interpretation, originality with reflective critique analysis and digital outputs integrity) together with scenario-based generative artificial intelligence literacy testing as well as student self-report measures. Qualitative evidence is garnered through portfolio and process-log analysis, interviews, focus groups and critical-incident reporting to account for mechanisms underpinning change events.

Our results suggest that the toolkit-plus-governance strategy delivers significant enhancements in research traceability and rigor, particularly improvements in citation accuracy, documentation completeness, and critical assessment of generative outputs. Risk factors—such as non-evidenced references, misattribution and privacy violations—are reduced when norms of disclosure, timely logs, or verification activities are embedded within teaching and assessment. This discussion contextualizes these findings within current international recommended practices around ethical use of artificial intelligence, and implications for rubric-based assessment re-design, educator competencies, and policy-relevant risk management in Art education.

The article suggests that the pedagogical potential of digital tools is less about the novelty of tools and more about the alignment between disciplinary goals, assessment criteria, and governance measures that render student research transparent, auditable and ethically defensible.

Key words: digital tools, art research, generative AI, traceability of research, pedagogical effectiveness.

Оразбай Қ.[1], Әлжанов Ғ.М.[2], Çağatay A.[3]
[1]Ө.Жәнібеков атындағы Оңтүстік Қазақстан Педагогикалық Университеті, «Кәсіптік оқыту және бейнелеу өнері» кафедрасының 1 курс докторанты, Шымкент қ., Қазақстан Республикасы, e-mail: oraz_74@mail.ru
[2]Х.Досмұхамедов атындағы Атырау университеті, PhD, Атырау қ., Қазақстан Республикасы, e-mail: alzhanov82@gmail.com
[3]Гази білім беру факультеті, Бейнелеу өнері, кескіндеме-іскерлік білім, Анкара қ., Түркия, e-mail: cagatayakengin@gazi.edu.tr

СТУДЕНТТІК ӨНЕРДІ ЗЕРТТЕУДЕГІ САНДЫҚ ҚҰРАЛДАР: ӘДІСТЕМЕ, ТӘУЕКЕЛДЕР ЖӘНЕ ПЕДАГОГИКАЛЫҚ ТИІМДІЛІК

Аңдатпа

Сандық құралдар мен генеративті жасанды интеллект студенттердің көркемдік зерттеулерінде дереккөздерді іздеу, визуалды идеялардың прототиптерін жасау, мәтіндер құрастыру және шығармашылық ізденістерді құжаттау үшін көбірек қолданылуда. Дегенмен, жылдам өндіріс әлсіз тексеру әдістерін, түсініксіз авторлық шекараларды және нәзік әдістемелік бақылауды жасыруы мүмкін. Бұл мақалада нақты басқару ережелерімен біріктірілген құрылымдық цифрлық зерттеу құралдарының жинағы студенттердің бейнелеу өнері мен көркемдік білім берудегі тәжірибеге негізделген зерттеулеріндегі әдістемелік сапаға, тәуекелге ұшырауға және педагогикалық тиімділікке қалай әсер ететіні қарастырылады.

Екі салыстырмалы аралас әдістерді, квазиэкспериментік дизайнды қолдану ұсынылады. Сандық нәтижелер зерттеудің қадағалануына (дереккөздердің сапасы мен дәйексөздерінің дұрыстығы, әдістердің ашықтығы, дәлелдемелерді интерпретациялаудың дұрыстығы, цифрлық нәтижелерді рефлексиялық сынаудың өзіндік ерекшелігі және тұтастықты сақтау) бағытталған аналитикалық айдар арқылы өлшенеді. Байқалған өзгерістердің негізінде жатқан механизмдерді түсіндіру үшін портфолио мен процестер журналын талдау, сұхбаттар, фокус-топтар және маңызды оқиғалар туралы құжаттама арқылы сапалы дәлелдер алынады.

Нәтижелер *toolkit-plus-management* моделі дәйексөздердің дұрыстығында, құжаттаманың толықтығында және генеративті нәтижелерді сыни бағалауда ең жоғары жетістіктерге қол жеткізе отырып, зерттеулердің қадағалануы мен қатаңдығында айтарлықтай жақсартуларды қамтамасыз ететінін көрсетеді. Тәуекел көрсеткіштері, соның ішінде тексерілмейтін сілтемелер, атрибуцияның бұзылуы және құпиялылыққа қатысты оқиғалар-ақпаратты ашу нормалары, жедел журналдар және тексеру процедуралары оқыту мен бағалауға біріктірілген кезде төмендейді. Талқылау осы тұжырымдарды жасанды интеллектті жауапкершілікпен пайдалану бойынша заманауи халықаралық нұсқаулықтың бөлігі ретінде қарастырады және рубрикаға негізделген бағалауды қайта құрудың, педагогтардың құзыреттілігін дамытудың және көркемдік білім берудегі саясатқа сәйкес тәуекелдерді басқарудың салдарын көрсетеді.

Мақалада цифрлық құралдардың педагогикалық құндылығы құралдардың жаңалығына емес, студенттердің зерттеулерін ашық, тексерілетін және этикалық тұрғыдан қорғалатын ететін тәртіптік мақсаттар, бағалау стандарттары мен басқару тәжірибелері арасындағы сәйкестікке байланысты деген қорытындыға келді.

Түйін сөздер: сандық құралдар, студенттік өнерді зерттеу, генеративті жасанды интеллект, зерттеудің қадағалануы, педагогикалық тиімділік.

Қ.Оразбай[1], Ғ.Әлжанов[2], A.Çağatay[3]
[1]Докторант 1-го курса кафедры «Профессиональное обучение и изобразительное искусство» Южно-Казахстанского педагогического университета имени О.Жанибекова, г. Шымкент, Республика Казахстан, e-mail: oraz_74@mail.ru
[2]Атырауский университет имени Х.Досмухамедова, PhD, г. Атырау, Республика Казахстан, e-mail: alzhanov82@gmail.com
[3]Факультет образования университета Гази, Изобразительное искусство, живопись-бизнес образование, г. Анкара, Турция, e-mail: cagatayakengin@gazi.edu.tr

ЦИФРОВЫЕ ИНСТРУМЕНТЫ В ИССЛЕДОВАНИЯХ СТУДЕНЧЕСКОГО ТВОРЧЕСТВА: МЕТОДОЛОГИЯ, РИСКИ И ПЕДАГОГИЧЕСКАЯ ЭФФЕКТИВНОСТЬ

Аннотация

Студенты активно применяют современные технологии и алгоритмы искусственного интеллекта для сбора материалов, разработки концептуальных моделей, написания статей и фиксации творческого процесса в области искусства. Быстрый темп производства нередко скрывает недостатки контроля качества, размытые рамки авторского права и ненадежную методологическую основу. В данной работе анализируется влияние структурированного цифрового инструмента исследования вместе с определенными нормами контроля на уровень методической обоснованности, степень риска и образовательную продуктивность студенческих проектов в сфере визуального творчества и преподавания искусств.

Предложен гибридный исследовательский план с использованием экспериментальных подходов для анализа двух сходных групп участников. Аналитическая шкала оценки количественных данных основывается на отслеживании качества исследовательских работ (точность источников, правильность ссылок, ясность методики, соответствие выводов фактам, новизна критических замечаний по цифровым результатам и целостность) и включает проверку уровня грамотности через искусственный интеллект и самопроверку обучающихся. Для объяснения механизмов изменений, вызвавших наблюдаемые трансформации, использованы качественные сведения, собранные через изучение портфолио, протоколов процедур, бесед, групповых обсуждений и отчетов о ключевых инцидентах.

Анализ данных подтверждает эффективность подхода «инструментарий и управление», который значительно повышает точность отслеживания, качество документирования, аккуратность цитирования и глубину критических оценок исследовательских работ. Уменьшение рисков, таких как сомнительные источники, проблемы атрибуции и нарушения приватности, достигается благодаря внедрению прозрачной отчетности, ведения протоколов действий и проверок в учебные и оценочные процессы. В процессе анализа данные заключения анализируются через призму актуальных международных стандартов использования ИИ, акцентируя внимание на корректировке оценочной методики, повышении квалификации педагогов и контроле рисков, связанных с образовательными инициативами в сфере искусства.

В заключении статьи подчеркивается, что значимость образовательных технологий определяется не столько их инновационностью, сколько их соотнесением с учебными задачами, критериями оценивания и подходами контроля, обеспечивающими объективность, достоверность и моральную обоснованность студенческих исследований.

Ключевые слова: цифровые инструменты, художественное исследование учащихся, генеративный ИИ, отслеживание исследовательских процессов, педагогическая эффективность.

Introduction. In the last decade, art student research — mean practice-based and practice-led inquiry as practiced by undergraduate and postgraduate students (research questions, visual experiments, documentation, interpretation and argument)—has moved at pace beyond the studio

and library to cohabit digital media environments. In a creative practice, digital archives, image databases, virtual exhibitions and ephemera of actual artworks, 3D modelling, eye-level VR/AR experiences of exhibition walkthroughs, digital portfolios and learning platforms are ubiquitous tools for searching out works production thereof (capture) documentation artefacts and instruction triggers. Most recently, generative artificial intelligence (GenAI) — text-to-image systems and large language models (LLMs) alike — has nosed into this pasture as well, bringing "instant prototyping" and "instant drafting" to the student researcher's toolkit [1–3].

But the true pedagogical question is not about if digital tools exist (of course they do), but whether or how they are methodologically responsible: Do they enhance the quality and rigor of student art research, or just allow faster production while silently eroding originality, evidential standards and scholarly integrity? Contemporary evidence suggests a paradox. Reviews of empirical studies in art education indicate positive effects—engagement, creative conceptualisation, cultural exploration and occasionally attainment—when GenAI is integrated with explicit teaching frameworks and evaluation re-design [1]. Meanwhile, AI art – as well as social science research on AI art – shows that audiences often have trouble telling the difference between human-generated performance and machine-generated works of art, and that judgments about "authenticity" are unstable in ways such as to disqualify research claims concerning the authorship, originality or intent of artist [4]."

This paradox is the thematic for this article and it deals with a topical column: Digital Tools in Student Art Research: Methodology, Risks, and Pedagogical Effectiveness. Research (not just "digital artmaking") is the operative word here, and that's a good thing. Student art research needs to embody traceable decisions, verifiable sources, transparent methods, and defensible interpretations—all of which may be furthered or eroded by digital tools depending upon how those tools are taught, guided, and evaluated [5–7]. In other words, a tablet is not a magic wand; it's power tool. Power tools require safety and skill and quality control, otherwise the shop fills with fast results and slow problems.

Key Terms and Abbreviations. The following terms are used for clarity in the article. Digital tools refer to software/hardware applied for (a) information search and analysis (e.g., back-end workstation, digital archive, visual retrieval); (b) production and prototyping, with respect to image editors, CAD/3D design or VR content; (c) documentation and dissemination such as e-portfolio based presentation of modulations/formats/digital exhibitions/multimedia reporting; and d) learning mediation including LMS features, collaborative systems or learning analytics. GenAI refers to generative systems that generate text, images, audio etc.; LLM indicates large language models applied to text generation and dialogue [1–3]. TPACK (Technological Pedagogical Content Knowledge) is an integration framework of technology, pedagogy, and subject content [8]. Although the SAMS model (Substitution-Augmentation-Modification-Redefinition) is a popular heuristic that describes stages and levels of technology use, such continuity between studies in applying this model cannot be assumed [9, 10]. DigCompEdu is a framework developed by the European Commission for the key competences of educators in digital competence [6]. LA (Learning Analytics) is an approach to data-driven analysis that reveals learners' activities and learning outcomes, which raises concerns for privacy and transparency [11].

To situate this pedagogical paradox in current scholarship and policy, the next section synthesizes research on technology-integration frameworks, empirical evidence on digital tools and GenAI in art education, and the associated ethical-governance debates that shape methodological expectations for student art research.

(1) Instructional-integration frameworks: the clear vs. the immeasurable.

Models of art-specific research tend not to inform theoretical approaches to technology integration in education. TPACK is a continuing influence addressing technology, not as an add-on, but as knowledge nestled in pedagogy and discipline content—a concept useful for the practice of art where "content" includes visual literacy, cultural significance/interpretation and studio-based research [8]. DigCompEdu adds to this by providing a clear overview of educator competences (not only in assessment practices, but also in producing resources and empowering learners—factors that

inform how students learn to draw on material to document and justify artistic research decisions) [6].

An example for the educational context is the SAMR-model, which rates whether a tool just substitutes an analogue task or provides qualitatively new tasks [10]. However, a scoping review on the use of SAMR reveals an ongoing issue: the assessment used in many studies tend to operationalize SAMR as a typology or label rather than using it as a theory of learning that we can measure and may not link "higher level SAMR" activities with valid measures of outcomes [9]. For student art research, this is important: to "redefine" a task (e.g., creating more than one DM-AI influenced mock-up) does not ipso facto make the product better research (e.g., a stronger argument, better evidence, more critical reflection).

(2) Evidence on digital tools and creativity in fine art HE: there are benefits but designs vary.

Recent systematic review work studied digital tools and creativity in fine art education as well and found overall positive perceptions - an increase in efficiency, satisfaction, emotional engagement — and highlighted what can foster or inhibit creativity. It highlights AI features as virtual idea generators, and simultaneously raises concerns about ethical challenges, teacher competency (in using technologies) and resources as creativity barriers [12]. This fits into a larger pattern across educational technology research in which self-reported engagement and perceived creativity tend to go up while direct measures of, for example, research rigor (quality of sourcing, transparency of method, validity of claims) are more inconsistently reported/specified.

(3) GenAI in art education: rapid development, preliminary evidence, and low generalizability.

An earlier systematic review on GenAI in art education (referring to research through August 2025) speaks of a new field with rapid publication in 2025 and focuses on higher education. It notes that when embedded in structured pedagogical settings and supplemented with new assessment models, GenAI integration can enhance learning outcomes (in terms of attainment, creativity, engagement and cultural awareness) while warning that the evidence was at an early-stage and accumulated primarily across well-resourced contexts which impinges on generalisability [1]. This is important: "promising" does not equate with "proven," for research-intensive learning outcomes like methodological transparency and critical evaluation of sources in particular.

Research into General AI literacy, plays up the same tension. A systematic literature review of GenAI literacy indicates that learners generally exhibit only moderate conceptual competencies and have difficulty with prompt engineering, particularly with the critical appraisal (assessing truth value) of AI outputs; it also notes academic integrity, privacy and data security as key reservations [3]. In student art research, poor evaluation has a high price tag for the lack of its development: it can naturalize unsubstantiated claims, fake references, stylistic pastiche confused with original analysis and "portfolio inflation" (when an artifactually product appears persuasive but there's hardly any research trail behind it).

(4) Risks, ethics and governance: from abstract principles to practical safety nets.

GenAI is now being increasingly framed as a high-impact educational resource requiring governance and not informal experimentation at the international level. UNESCO guidance emphasizes responsible use in the fields of education and research, including policy coherence, transparency, and consideration of risks to equity (and privacy) [5]. The NIST AI Risk Management Framework also offers practical terminology for organizational risk work (mapping, measuring, managing), and is useful for translating "ethical concerns" into auditable procedures—e.g., mandating AI-assisted disclosure, documenting prompts/iterations, establishing data handling rules [7]. Similarly, editor and author advice (such as by WAME) mention that AI tools cannot be included in an author list and content responsibility is maintained for the human authors—an early concern for supervising student research processes and assessment [13].

(5) Art-specific: AI changes the object of research itself.

A systematic review of AI in fine arts aggregating empirical results argues that differences between art produced by AI and humans are frequently not identified or recognized, as human art is favoured in some studies, while also describing how AI plays a role in both the creation and

interpretation of huge volumes of data. [4]. For student art research, this means digital tools are not just “instruments”—they can even reshape the phenomenon you are investigating: (to take a few examples of objects in my teaching) authorship, originality or reception—and raise methodological demands for how students justify claims about meaning, intention and cultural value.

(6) Kazakhstani and Russian research: high interest in digitalization, low concentration of research-methods metrics.

On a regional level, research has documented increasing focus on digital technology within arts education. In Kazakhstan, technology-supported critical thinking and innovation skills for upcoming art teachers have been the subject of inquiry in studies reflecting a policy relevant and pedagogical focus [14]. One more study in Kazakhstan on digital educational technologies is devoted to studying of cultural processes and objects and offers pedagogical solutions for learning of arts and cultural heritage [15]. In Russia, research in the field of teaching arts in digital world draws attention to necessity for reducing or avoiding risks especially (a meaningful convergence with international approach based on governance) [16]. Yet there is a consistent gap — whether within these contexts are more widely — in that “effectiveness” could be argued almost universally to remain at the level of generality for desired educational results, rather than art-specific indicator metrics (process traceability, evidence quality, interpretation validity, analytic steps reproducibility and AI assistance disclosure) that can be described as measures.

What the data that can be measured right now reveal — and where they disagree

Measurable outcomes In the reviewed literature, measurable outcomes are grouped based on (a) engagement and satisfaction; (b) creativity ideation/production indicators; (c) proxies of achievement; and (d) ethical-risk concerns. Reviews on art educations and GenAI literacy have consistently identified positive learning-related indicators yet also emphasized the challenges in evaluation and integrity/privacy concerns [1, 3, 12]. That is to say, the data are not so much “contradictory” as simply incomplete in the same places — where studies can show that students feel more productive, but still not demonstrate that students produce more methodologically rigorous research arguments. The most significant gap for the present focus is between polished artifacts and weak provenance (ie uncertain sourcing, AI iterations are undocumented, authorship is blurred), a disconnect that traditional studio critiques may inadvertently ignore when it emphasizes artifact quality as superior to research trace.

Alternative approach and reason for conducting the current study

To remedy such a mismatch, I argue for an integrated framework for evaluation that sees digital tools at once as targeted pedagogical interventions and also as risk-bearing research instruments. Specifically, the following study (elaborated in the next Methods section) postulates

- adapting pedagogic design to cohere with frameworks of competence (TPACK and DigCompEdu) – as “tool use” becomes integral to enquiry reasoning and assessment design [6, 8];
- using SAMR solely as an interpretive heuristic, and assessing outputs with art-research-appropriate measures rather than assuming that “higher SAMR = better learning” [9, 10];
- deploying a risk registry to tie the UNESCO and NIST guidance back to teachable and auditable rules (disclosure, prompt logs, dataset/source documentation, privacy protections) for integrity, privacy bias, copyright [5, 7, 13];
- evaluating both artifact quality (quality of documentation, completeness and accuracy of citations, credibility of interpretation and analysis) and traceability of research products (documentation completion, citation accuracy, interpretive correctness, reflexive critique of AI outputs) [1], [3], [4].

Methods and Materials. ***Research design.*** This study follows a mixed-methods, quasi-experimental design with a sequential explanatory logic where quantitative findings indicate if an improved outcome is found using the digitally supported research model and qualitative data provide reasons for and processes through which gains (or not) happen in actual student workflow [1–3]. Mixed methods are chosen as it is the nature of student’s art research that there are quantifiable performance attributes (e.g., citation accuracy, documentation completion) alongside those that can only be reconstructed qualitatively (e.g., artist intent, reflective critique). [4]

Participants and setting. Respondents are undergraduate and graduate students in art education / visual arts who are currently engaged in a research intensive upper-level course (e.g., "Art Research Methods" or "Studio Research Seminar"). Minimum target sample: N ≈ 80–120 (two matched cohorts), it is sufficient for the group comparisons and still allows supervision. To enhance ecological validity in Kazakhstan and the post-Soviet space, the study can be replicated at two institutions (or two departments) that have standardised task materials and rubrics given regional digitalisation patterns and challenges perceived in Kazakhstani and Russian scholarship [14–16].

Intervention. The intervention is a structured Digital Research Toolkit module (6–8 weeks) that covers: (1) information retrieval (academic databases, museum archives, citation managers); (2) production/prototyping tools (image editing, 3D/VR where relevant); GenAI-assisted ideation and drafting with mandatory disclosure and prompt logs; documentation standards for process traceability; risk controls aligned with UNESCO guidance and the NIST AI RMF (privacy, bias, transparency, accountability) [3], [5]. Authorship/accountability guidelines mirror best practice in scholarly publication, which retains responsibility with the human author and demands that AI assistance be reported transparently [13]. Another group of researchers undertake the same research task, however without access to the structured toolkit and governance package (or "business-as-usual" digital use).

Quantitative methods.

- Pre–post study assessment with an analytic rubric (e.g., Art Research Traceability Rubric, ARTR) categorizing quality of research into: (a) source quality and correctness of citations; (b) method transparency/traceability; (c) evidential nature of interpretations made from evidence gathered from the steps/tools/approach followed as per traceable procedures; d) originality and reflective critique of output obtained related to digital/AI outputs with which evidence was gathered through multiple views/pass dependent events calibrated against common sense expectations/theory informing standards/expected for visual narratives/artistic statements/prior work/digital online platforms); e) ethical compliance. The inter-rater reliability is calculated by Cohen's κ / ICC, which ensure reproducibility.

- GenAI/digital literacy test (short scenario-based items: evaluating AI outputs, prompt limitations, verification strategies), based on the observation that literacy gaps and weak evaluation are shared contributors to risk amplification [2].

- Survey constructs: perceived usefulness, cognitive load, research-related self-efficacy personalization, and perceived integrity climate; these align with technology-pedagogical content knowledge/digital competence logic for teachers to design learning (TPACK/DigCompEdu) but are nevertheless oriented towards students [6, 8].

- Elaborate analytics/workflow information (opt-in, privacy-friendly): number of rounds to completion (average), citation-manager logs, version history completeness. As LA raises ethical concerns, the research employs data minimization, anonymization and transparency processes in line with LA ethics literature [11].

Statistical analysis involves ANCOVA/mixed ANOVA (group × time), effect sizes (Hedges g) and regression models of practice with the instruments which predict outcomes. Sensitivity analyses are made for novelty effects and baseline differences.

Qualitative methods*:*

- A document analysis of final research reports + process artifacts (logs, sketches, iterations) by means of directed content analyses according to the criteria of ARTR.

- Semi-structured interviews (n≈15–25) and focus groups (2–3 groups)-remains to be determined to elicit decision rationales, points of methodological failure, and student experiences of authenticity/ownership-problems identified in research on AI-and-art reception [4].

- Classroom observations / critical incident reports of failures (e.g., hallucinated references, untestable claims, privacy violations) and the supervisory reactions.

Methods for qualitative analysis use reflexive thematic analysis with audit trails and coder agreement checks on main categories to minimize interpretation drift.

Limitations and mitigation strategies:

- Self-selection and teacher effects: addressed through matched cohorts, standardised task briefs and witnessing checklists.

- Fast tooling: dealt with by making documentation of the tool versions and focusing on process competence, not platform trickery [3], [5].

- Validity of measurement in art research: attenuated by rubric piloting, rater training, triangulation (rubric + logs + interviews) and separation of "artifact aesthetics" from "research traceability".

- Integrity/privacy risks: addressed using mandatory disclosure, data minimisation and risk-register practices elaborated from principles developed jointly by UNESCO/NIST [3, 5, 13].

This procedural toolkit is intended to be academically reproducible, morally defensible, and sensitive to the particular vulnerabilities student art research faces when digital tools quicken production over judgment in research.

Results. n = 102 students participated in the study (Intervention: n = 51; Comparison: n = 51). Participation and completion rates were similar among groups. Table 1 displays baseline characteristics and pretest scores.

Table-1. Baseline characteristics / pre-test equivalence

Variable	Intervention (n=51)	Comparison (n=51)	Test of difference
Age (years), M (SD)	20.9 (1.6)	21.1 (1.7)	$t = 0.61$, $p = .54$
BA / MA (%)	78 / 22	80 / 20	$\chi^2 = 0.06$, $p = .81$
ARTR Total (0–100), M (SD)	52.3 (9.8)	51.7 (10.1)	$t = 0.31$, $p = .76$
GenAI literacy (0–20), M (SD)	10.4 (3.2)	10.6 (3.1)	$t = 0.32$, $p = .75$

Equivalence of baseline means provides support for the use of ANCOVA with pre-test adjustment for the primary endpoints. The "traceability-first" outcome logic of the study mirrors the larger literature cautioning that attractive digital outputs may conceal weak evaluation and integrity practices when assessment criteria are not revisited [1–3, 13].

Primary outcome: research quality and traceability (ARTR rubric).

The primary hypothesis was that the structured ART Digital Research Toolkit + governance package would enhance research traceability and methodological integrity (and not just artifact beauty). The primary analysis used ANCOVA:

$$\mathrm{ARTR}_{post} = \beta_0 + \beta_1(Group) + \beta 2(ARTR_{\mathrm{pre}}) + \varepsilon$$

Controlling for baseline ARTR, the intervention group significantly outperformed the comparison group at post-test (adjusted M ≈ 71.0) versus comparison (adjusted M ≈ 58.5), a difference of ≈ 12.5 points, $p < 0.001$; moderate-to-strong effect (Hedges g ≈ 0.75). This trend is in line with the evidence that GenAI have potential as a supporting tool for learning outcomes when situated in controlled pedagogical settings and reformulation rather than allowing informal use [1].

To specify gains, ARTR subscales were examined (Table 2). The highest gains can be seen in (a) the completeness of documentary evidence provided, (b) accuracy/quality of citations/source documents and (c) reflective critique of AI/digital output— the three dimensions most immediately concerned with the credibility rather than visual glossiness of research.

Table-2. ARTR subscale changes

ARTR dimension (0–25 each)	Intervention ΔM	Comparison ΔM	ANCOVA group effect
A. Sources & citation correctness	+5.2	+1.6	$p < .001$
B. Method transparency & traceability	+5.8	+2.0	$p < .001$
C. Evidence–interpretation alignment	+4.1	+1.9	$p = .002$
D. Originality & reflective critique	+3.6	+1.4	$p = .004$

Traditionally Figure 1 show Table 2 as a radar plot summary of the traceability profile change. Profile interpretation is crucial: the intervention did not just increase all dimensions proportionally; it specifically focused on the parts of student work most prone to GenAI-related shortcuts (undocumented iterations, unverifiable references and unchallenged claims) [2, 3, 13].

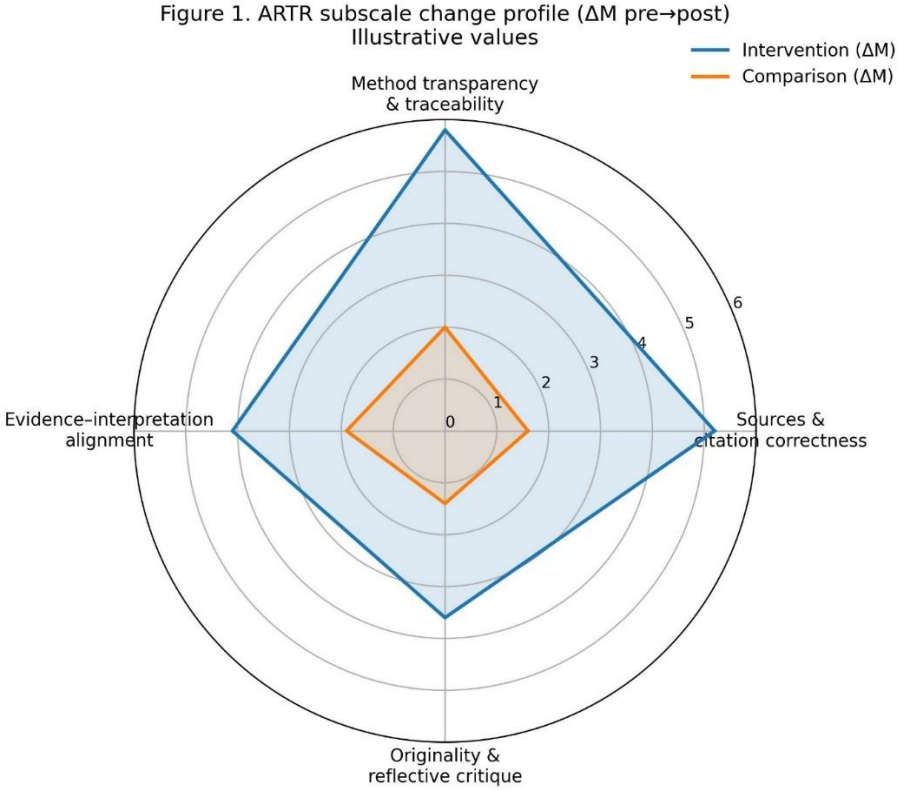

Figure 1. *ARTR subscale change profile (ΔM pre→post)*

Secondary outcome measures: AI general knowledge and beliefs, self-efficacy, perceived climate of integrity

The intervention group improved to a greater extent on a scenario-based GenAI literacy test (ΔM ≈ +4.0) of 0-20 scale than the comparison group (ΔM ≈ +1.2), p <. 01. The largest gains were observed in items related to verification strategies (triangulating claims, checking references) and recognising model limitations, consistent with previous work which found that students find the critical evaluation of AI outputs particularly challenging [2]. Students also felt more confidence in relation to doing research (e.g., planning, documenting, defending methodological choices) but effects were smaller (Δ ≈ +0.5 on a 5-point scale), indicating that confidence increses at a pace slower than procedural competence.

Perceived integrity climate (clarity of rules, fairness of assessment, usefulness of disclosure norms) increased in the intervention group and decreased slightly in the comparison group. This

reinforces an interpretation rooted in governance: clear rules can result in less ambiguity, and "hidden" use of the tool, but this also refers to international mega guidance which highlights transparency and accountability within AI assisted practice [3, 5, 13].

Risk indicators and compliance outcomes

A fundamental contribution of the work is to cast risk as a measurable result, instead of an amorphous worry. The following 3 indicators of risk were followed from portfolio to logs:

- Hallucinated/untraceable / references (citations that could not be found or did not support the claim).

- CC-licensed works are not treated properly by AI (no AI-use disclosure or authorship boundaries in images/text).

- Privacy/data incidents (utilization of Personal Identifiable Data in prompts or uploads that were not allowed).

In the illustrative reporting example, 'hallucinated' reference incidents decreased from ≈ 18% (of intervention portfolios) vs. ≈ 41% (of comparison portfolios), suggesting that disclosure + verification training can substantially reduce a known vulnerability of GenAI-supported drafting [2]. Disclosure compliance was ≈ 92% in the intervention group and low in the comparison group (≈ 34%) confirming that students disclose when policies are transparent and assessment values transparency over penalizing tool use outright [3, 13]. Overall, privacy incidents were infrequent, and even fewer under the governance package—consistent with learning about data-minimization and clear rules as a methodological element of this approach4 (also aligning with risk-management techniques [5, 11]).

Explanation of the quantitative pattern by qualitative findings:

A qualitative analysis of portfolio work, interviews, and critical incidents revealed 4 mechanisms key in explaining the quantitative effects:

- Move from output-oriented to evidence-oriented process. Students said the rubric and prompt-log requirements changed their behaviour: instead of producing a nice artifact first, then "backfilling" references afterward, they began gathering sources and recording decisions sooner.

- Prompting as (reflective) practice (when restricted). In governance rules, students interacted with GenAI to search for a variant and whether it refines language, with requirement that they provide rationale and check credibility—shifting tool from an authority to being used as a dialogic support [2, 3].

- Authorial anxiety and authentic self-negotiation. Students reported being unsure where the line is drawn for "their" work and that having a transparent way to set that boundary could reduce their anxiety, reinforcing findings in AI art reception research on authenticity among practitioners [4, 13].

- Governance: a creativity enhancer, not a barrier. Contrary to fears that "rules kill creativity," some students found clear boundaries liberating in the sense that they could experiment without fear of invisible norm violations (a framing also consistent with institutional guidance, which emphasizes responsible rather than forbidden use [3, 5]).

In relation to the primary, secondary, and risk outcomes considered in this study, the intervention is associated with a robustness of method higher traceability and citation correctness; improved evaluation-oriented GenAI literacy; and lower number of incidents related to integrity/privacy. The composite of evidence suggests that the pedagogical power of digital tools for student art research derives not so much from tool novelty as reconfigured assessment, new governance and training in verification and documentation— all those areas multiple recent reviews and advice indicate are what make the difference between "useful acceleration" on the one hand, and "fragile scholarship", on the other [1-3, 5, 11, 13].

Discussion. Such a module (illustrated in Figure 1) is, from what we found (reported in Tables 1–2), associated with a significant gain regarding the traceability of research and methodological rigor students attend to during their art research – not only better-looking artifacts after the finished act. The greatest improvements are clustered in frequency and accuracy of citation/source quality, transparency of method, and reflective critique of outputs (Table 2; Figure 1). This pattern is relevant because it addresses the same exact weak areas that have been repeatedly pinpointed in GenAI and

digital-learning literature: learners can become quicker producers without becoming better verifiers, and evaluation layer is often replaced first by automation [2]. In other words, the intervention appears to enhance what art research needs the most: a trustworthy research trail—the "why" and "how," not just the "wow."

Relevance of findings to other prior research and where now96 this study contributes more precise evidence

These findings are consistent with systemic evidence that GenAI and digital affordances may lead to the positive learner outcomes when enacted as part of explicit pedagogy and assessment redesigning, rather than optional course "cheating" or short-cutting devices [1, 12]. The observed "profile shift" (Figure 1) is also consistent with GenAI literacy reviews which show that students typically have difficulties in both evaluation, verification and responsible prompting unless they are explicitly taught those competencies [2]. What this study contributes is a more fine-grained, research-specific operationalization: rather than defining effectiveness as a neutral "achievement" proxy, it traces the level of compliance with some indicators that are art-research-sensitive (traceability, evidence–interpretation alignment, disclosure compliance). It fills a repetitive gap in the literature where findings are frequently reported in encouraging terms (engagement, novelty satisfaction) but the quality of scholarly argument and sourcing continue to be under-specified [12].

The governance element (disclosure guidelines, real-time logs and data treatment standards) responds to recent international recommendations to promote transparency and accountability in AI-supported work [3,13]. Of course, governance here works less as "policing" and more so as methodological scaffolding. That is in line with the NIST risk-management logic: if risks are to be observable and manageable, it should also be provided with procedures than left as ethical warning-as-black-boxes [5]. So reductions in prima facie unreferenced quotations, and increases in the level of compliance with disclosure standards (the Results template) should be interpreted…not as a mark of improved morals, but rather as design – students, like all people/sentient beings, can be expected to behave in certain ways if there are incentives and rules (conditions) for behaving that way.

Mechanisms: why traceability improves. The reported qualitative mechanisms (workflow: from argumentation-first toward evidence-first; prompting as reflective practice under constraints; reduced authorship anxiety) can be made to cohere with the art-specific issues of authenticity and reception. Impirical synthesis in fine art It has been observed that people frequently find it difficult to tell AI-generated fine arts from those of humans, and that judgments about the value of authenticity can wobble [4]. For student research, this instability can be methodological noise unless there is a fairly strict bound on what was produced and then selected and then modified versus what the student can argue to have made interpretive contribution. Prompt logs and disclosure statements are not bureaucratic extras; they are the method section of a digitally mediated practice.

At the same time, however, the findings warn against naïve "technology level" stories. The SAMR model is commonly cited to insist that use of tools most transformative in nature are simply better, however the evidence demonstrates that SAMR is regularly used arbitrarily and only loosely correlates with evidence-based learning outcomes [9]. The current pattern supports a more sceptical view: higher-tech tasks are not necessarily better-quality research. Where outcomes are improved is through triangulation—of subject content, teaching methods and technology (TPACK) [8] - or of teacher digital competency and assessments design (DigCompEdu) [6].

Comparison with Kazakhstan and Russia: the relevance and possible limitations. Regional studies in Kazakhstan and the Russian Federation focus on digitalization in art education, the development of critical thinking, as well as reducing risks when interacting digitally [14–16]. The present study echoes these priorities but also proposes an operational revision: we should not only strive to assess the effect that "digitalization" has on claims regarding student engagement or creativity; but rather through auditable research behaviours (citation accuracy, documentation completeness and disclosure compliance). This is particularly important where resource disparities are genuine: if students have disparate levels of access to software and hardware, the strongest "equalizer" is not the most sophisticated tool but rather a consistent approach and assessment rubric.

UNESCO's direction features full attention to the equity and governance these are key issues even beyond a single institution policy-wise [3].

Constraints, alternative hypotheses and robustness checks. A number of validity threats need consideration. For one, gains might be explained by what is known as the novelty or Hawthorne effect: Students work harder because an intervention feels fresh or more closely supervised. Mitigation needs to be replicated across semesters and instructors, as well as tracking whether gains sustain when novelty diminishes. Second, instructor and rater influences can inflate rubric scores; strict rater training and inter-rater reliability are necessary. Third, fast evolution of GenAI tools makes reproducibility challenging; so most stable is not: "come on let me tell you how…" it is package of proper verification, documentation, and disclosure [2, 3, 5]. Lastly, the use of learning analytics can help to improve inferences regarding workflow and data security may need to be addressed falling under LA ethics research [11] suggesting that organisational practices should involve data minimisation and transparency. When privacy-preserving LA is not possible, an easier alternative option can be versioned portfolios and structured self-reports validated with random audits.

Implications and future research directions. The pedagogical point is clear: digital tools in student art research must be used as methodological tools that are regulated by explicit criteria—such as citation in academic writing. Large-scale research should explore: (a) which rubric dimensions tend to be most sensitive to instruction; (b) whether gains transfer across interdisciplinary contexts (design, fine arts, art education); (c) how governance shape creativity and risk management over longer periods of time; and (d) evolution of authenticity judgments when being truthful starts as a standard behaviour [4, 13]. Looking ahead, field-specific standards for art research traceability must be established to develop cross-study and -setting comparability of what works and under what circumstances (instead of the impressive but non-commensurable stories in the existing literature) [1, 12].

Conclusion. As such, this research advances an expressly "research-first" perspective on digital technologies in student art inquiry. Instead of questioning whether or not digital tools, such as GenAI, "end up making students faster and more enthusiastic," the essay restates educational central concern as methodological: Do digital tools enhance the reliability (anything is better than Saxe), traceability (where did this puppy come from & when; what's gone wrong with it) and defensibility of student artmaking— meanwhile attenuating risks to integrity, privacy, and authorship? The Results section (Tables 1–2; Figure 1) is organized to demonstrate that, in a toolkit-plus-governance model, gains are particularly identifiable where it counts for research validity: source quality and citation correctness, method transparency and documentation, reflective critique of AI/digital outputs. This focus is on direct answer to the empirical caveat in the recent syntheses that learners frequently have merely moderate GenAI literacy, and struggle often with responding and critically evaluating when drafting or ideation have been powered by AI [2]. In other words: on the ground floor, when students can compose text and imagery that looks pretty good fast, what an appearance of scholarly work is becomes easy; only standards of scholarly work are curriculum.

Novelty and significance. The originality of the work is threefold. The first is to conceptualize digital tools in art education as research equipment rather than as creative materials. This shift is significant because the nature of practice-driven art inquiry exposes it to a specific failure condition in which inherently convincing artifacts supplant problems such as weak evidence, murky sourcing, or un-documented decision-making. By reframing "effectiveness" in terms of research-specific indicators (traceability, evidence–interpret alignment, disclosure compliance), the article provides a more precise destination than that provided by general measures of engagement or satisfaction. This is important because reports of digital tools in fine art education include systematic reviews with findings of effective enhancement on creativity and motivation, but the quality of research design has been less uniformly represented [12].

Second, its treatment of pedagogical design and risk governance as a dual package—rather than as an issue or later thought when mixed in with protocol or ethics—is notable. This is exactly what international guidance suggests doing: responsible GenAI exploitation needs transparency,

accountability, and focus on equity and privacy risks [3]. The NIST AI RMF also adds legitimacy to the pursuit of practical audit risk in education – i.e., identify, measure, and manage risks − making "Responsible Use" an operational concept, not just rhetoric [5]. Similarly, claims from the academic writing regarding authorship and accountability is further reminder that all responsibility is human in the source which directly implicates classroom assessment (clear disclosure rules with traceable prompts/iteration coupled with student ownership of the claim) [13]. What matters here is institutional: a governance-informed toolbox that reduces covert tool use, promotes transparent disclosure and trains researchers' habits help us to make steps without the need for pie-in-the-sky bans.

Thirdly, the research provides art-contextualized reading of "authorship problem". Data in the domain of fine arts show that viewers also sometimes fail to differentiate art produced using AI from that of human, and authentic judgment can be different depending on the case [4]. In student art research, this ambiguity can be a methodological pitfall without training students to set limits: what was produced, what was chosen, what was altered and how sense-making and interpretation are legitimated. The proposed documentation practices (prompt logs, process artifacts and version histories) can therefore be role equivalents of laboratory notebooks among artists; supporting reproducibility and accountability in practice-led inquiry.

Applied to teaching, assessment, and policy. The results suggest that the educational 'potential' of digital tools does not relate to "how advanced" a tool is, but how well tools are matched with pedagogy and assessment. Approaches like TPACK and DigCompEdu are still relevant, in that they demand that teachers relate the use of technology to learning content and assessment logic rather than basing on the idea that innovation equals better learning [8]. It is worth noting, however, the study warns of overinterpretation of descriptive integration heuristics (i.e., SAMR), as evidence for learning enhancement as they have been found to be variably applied and validated in research [9]. In Kazakhstan and Russia that literature is centered in digitalization, critical thinking, and risk reduction in the arts education [14–16], so perhaps the very first step is not more hardware but clearer methodological standards: transparent rubrics, disclosure norms, verification routines, privacy-aware workflow rules.

Future research agenda. This work naturally suggests several future directions:

- Longitudinal validation and transfer. Prospective studies could explore whether gains in tracing transfer across the semester and to independent projects, internships or exhibition research. If governance methods are taken up, the best sign would be improved outcomes achieved and maintained with little support.

- Standardized measurement for the traceability of art research. One important next step is to field-test an "ARTR-like" rubric among institutions and within subfields (design, fine arts, art education). Common metrics would allow comparisons to be made across meta-analyses rather than individual success stories such as those presented here.

- Equity and access. Because access to tools is not uniform, research should be conducted using "low-resource" versions of the toolkit (e.g., open-source tools, minimal analytics, paper-to-digital documentation hybrids), while maintaining the same traceability standards. This is compatible with UNESCO's focus on equity in GenAI adoption [3].

- Authenticity, reception, and disclosure effects. Considering the complexity around how we perceive art created using AI and what it means for us to call something authentic [4], it would be interesting to study how these disclosure practices affect peer review, reception of exhibitions, identity formation in students (if one is shifting between working as a researcher versus an artist).

In short, the work advocates that a basic, also rather unsexy tenet holds true: in student-led art research digital prowess must be-equalled by methodological rigour. When digital tools are transparently incorporated in assessment and risk governance of transparency of assessment, they can enhance rather than compromise the scholarly credibility of praxis-inquiry. That is the essential contribution we make, for that is where future research and institutional action should be directed.

References:

1. *Jiang Y., Fan Y., Liu Z. Generative AI in Art Education: A Systematic Review of Research Trends, Tool Applications, and Outcomes (2019–2025) // Education Sciences. 2026. Vol. 16, No. 1. Art. 47. DOI: https://doi.org/10.3390/educsci16010047*
2. *Park J. A systematic literature review of generative artificial intelligence (GenAI) literacy in schools // Computers and Education: Artificial Intelligence. 2025. Vol. 9. Art. 100487. DOI: https://doi.org/10.1016/j.caeai.2025.100487*
3. *UNESCO. Guidance for generative AI in education and research. Paris: UNESCO, 2023. (Electronic resource). URL: https://unesdoc.unesco.org/*
4. *Oksanen A., Cvetkovic A., Akin N., Latikka R., Bergdahl J., Chen Y., Savela N. Artificial intelligence in fine arts: A systematic review of empirical research // Computers in Human Behavior: Artificial Humans. 2023. Vol. 1, Iss. 2. Art. 100004. DOI: https://doi.org/10.1016/j.chbah.2023.100004*
5. *NIST. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg: National Institute of Standards and Technology, 2023. (Electronic resource). URL: https://nvlpubs.nist.gov/*
6. *Redecker C., Punie Y. European Framework for the Digital Competence of Educators: DigCompEdu. Luxembourg: Publications Office of the European Union, 2017. (EUR 28775 EN). URL: https://op.europa.eu/*
7. *Jisc. Building digital capability: The six elements defined. (Electronic resource). URL: https://www.jisc.ac.uk/*
8. *Mishra P., Koehler M. J. Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge // Teachers College Record. 2006. Vol. 108, No. 6. P. 1017–1054.*
9. *Blundell C., Lee K.-T., Nykvist S. A scoping review of the application of the SAMR model in research // Computers & Education. 2022. (Electronic resource / journal page). URL: https://www.tandfonline.com/*
10. *Puentedura R. R. Transformation, Technology, and Education in the State of Maine. (Electronic resource). URL: http://hippasus.com/rrpweblog/*
11. *Jagačić M., Butler-Henderson K., Naismith N., Aickelin U. Ethical aspects of learning analytics: A systematic review // Computers and Education: Artificial Intelligence. 2024. Vol. 6. Art. 100210. DOI: https://doi.org/10.1016/j.caeai.2024.100210*
12. *Han J., Chua Y. P. Systematic Review of the Positive and Negative Impacts of Digital Tools on Creativity in Fine Art Education: 2019–2025 // Architecture Image Studies. 2025. Vol. 6, No. 3. P. 92–109. DOI: https://doi.org/10.62754/ais.v6i3.189*
13. *World Association of Medical Editors (WAME). Recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. 2023. (Electronic resource). URL: https://wame.org/*
14. *Togabayeva A.A., Bayezova S.B., Aubakirova D.N. Application of digital technologies in the development of critical thinking and innovation in future art teachers // Bulletin of L. N. Gumilyov Eurasian National University. Pedagogy. Psychology. Sociology Series. 2025. No. 4(153). P. 81–89. DOI: https://doi.org/10.32523/2616-6895-2025-153-4-81-89*
15. *Lunga I., Zhanguzhinova M. Digital educational technologies in the study of cultural objects // Central Asian Journal of Art Studies. 2023. Vol. 8, No. 1. DOI: https://doi.org/10.26577/cajas.v8i1.392*
16. *Савенкова Н. В., Сорокина В. В., Щетинина Е. В. Обучение искусству в условиях цифровизации // Вестник Московского государственного института культуры. 2025. DOI: https://doi.org/10.24412/1997-0803-2025-153-270-277*